



**HAL**  
open science

# Semeru volcano, Indonesia: measuring hazard, exposure and response of densely populated neighbourhoods facing persistent volcanic threats

Jean-Claude Thouret, Marie Taillandier, Emeline Wavelet, Nourddine Azzaoui, Olivier Santoni, Boedi Tjahjono

## ► To cite this version:

Jean-Claude Thouret, Marie Taillandier, Emeline Wavelet, Nourddine Azzaoui, Olivier Santoni, et al.. Semeru volcano, Indonesia: measuring hazard, exposure and response of densely populated neighbourhoods facing persistent volcanic threats. *Natural Hazards*, 2023, 117 (2), pp.1405-1453. 10.1007/s11069-023-05910-5 . hal-04261504

**HAL Id: hal-04261504**

**<https://uca.hal.science/hal-04261504>**

Submitted on 10 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Natural Hazards

## Semeru volcano, Indonesia: Measuring hazard, exposure and response of densely populated neighbourhoods facing persistent volcanic threats --Manuscript Draft--

<b>Manuscript Number:</b>	NHAZ-D-22-01507R1
<b>Full Title:</b>	Semeru volcano, Indonesia: Measuring hazard, exposure and response of densely populated neighbourhoods facing persistent volcanic threats
<b>Article Type:</b>	Manuscript
<b>Keywords:</b>	human exposure; statistics; neighbourhood, volcano; Semeru; Indonesia
<b>Abstract:</b>	<p>We studied Semeru, East Java to show the population exposure to volcanic threats from its persistent, daily eruptive activity which endangers at least 50,000 of the 950,000 inhabitants living on the East and SE slopes and ring plain. Surveys, mapping and statistical investigation enabled us to assess the extent of exposure of 145 neighbourhoods (termed blocks) and characterize hazards and response to eruptions in 15 rural villages and small towns. Statistical analyses of datasets of 23 variables (11 of exposure, 7 of hazards, and 5 of response) and their attributes involved three operations: 1. Univariate and bivariate analyses enabled us to explore data and characterize the relationships between 11 variables to compute a multi-component exposure index. 2. Polytomous Logistic Regression (PLR) models selected six optimal exposure variables, suggesting that logistic regression can predict the exposure index for blocks outside the survey area and potentially on any active volcano. 3. Multivariate analyses and Hierarchical Agglomerative Clustering (HAC) distinguished four groups of blocks based on attributes of all variables correlated to the exposure index score. To contribute to disaster risk reduction, the distance/time criterion was applied to access ways and response facilities to highlight remote or blocked blocks in danger of imminent eruption including evacuation. Statistical analysis of optimal variables from local scale surveys can help identify neighbourhoods where disaster risk mitigation requires improvement.</p>
<b>Response to Reviewers:</b>	<p>Dr. James Goff, Editor in Chief, Natural Hazards Journal</p> <p>We would like to submit the revised article entitled: "Semeru volcano, Indonesia: Measuring hazard, exposure and response of densely populated neighbourhoods facing persistent volcanic threats" to Natural Hazards.</p> <p>In response to your request, we apologize and we thank you for having drawn our attention to the short paragraphs that may be considered as plagiarism and needed to be reworked before peer review. We have revised the third paragraph in sections 1.1, the first and third paragraphs in section 1.2, the second paragraph in section 1.3, and finally the first paragraph in section 1.5. We have shortened and reworked the mentioned lines, indicated their sources and added citations where appropriate. In the case of sections 1.3 and 1.5, we have referred to the publication from which the lines have been borrowed. In fact, we borrowed the content of these lines from our publication "Thouret et al., 2022" (International Journal of Disaster Risk Reduction). We hope that the revision would meet your expectation before peer review.</p> <p>Thank you for your consideration on our manuscript.</p>

1 **Semeru volcano, Indonesia: Measuring hazard, exposure and response of densely**  
2 **populated neighbourhoods facing persistent volcanic threats**

3

4 Jean-Claude Thouret <sup>1</sup>, Marie Taillandier <sup>2</sup>, Emeline Wavelet <sup>1</sup>, Nourddine Azzaoui <sup>2</sup>, Olivier  
5 Santoni <sup>3</sup>, Boedi Tjahjono <sup>4</sup>

6

7 <sup>1</sup> Université Clermont-Auvergne, Laboratoire Magmas et Volcans, UMR 6524 CNRS, OPGC  
8 et IRD, F 63000 Clermont-Ferrand (j-claude.thouret@uca.fr)

9 <sup>2</sup> Université Clermont-Auvergne UCA, Laboratoire de Mathématiques UMR 6566, CNRS,  
10 Campus les Cézeaux, 63178 Aubière, France (nourddine.azzaoui@uca.fr)

11 <sup>3</sup> FERDI & Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000 Clermont-Ferrand,  
12 France (olivier.santoni@uca.fr)

13 <sup>4</sup> IPB University, Faculty of Agriculture, Bogor, Indonesia (boetjah@apps.ipb.ac.id)

14

15

16

*Submitted to Natural Hazards*

17

*20 September 2022*

18

19

20

21

22

23

24

## 25 **ABSTRACT**

26 We studied Semeru, **East Java** to show the population exposure to volcanic threats from its  
27 persistent, daily eruptive activity which endangers at least 50,000 of the 950,000 inhabitants  
28 living on the **East and SE slopes and ring plain**. Surveys, mapping and statistical investigation  
29 enabled us to assess the extent of exposure of 145 neighbourhoods (termed blocks) and  
30 characterize hazards and response to eruptions in 15 rural villages and small towns. Statistical  
31 analyses of datasets of 23 variables (11 of exposure, 7 of hazards, and 5 of response) and their  
32 attributes involved three operations: 1. Univariate and bivariate analyses enabled us to explore  
33 data and characterize the relationships between 11 variables to compute a multi-component  
34 exposure index. 2. Polytomous Logistic Regression (PLR) models selected six optimal  
35 exposure variables, suggesting that logistic regression can predict the exposure index for  
36 blocks outside the survey area and potentially on any active volcano. 3. Multivariate analyses  
37 and Hierarchical Agglomerative Clustering (HAC) distinguished four groups of blocks based  
38 on attributes of all variables correlated to the exposure index score. To contribute to disaster  
39 risk reduction, the distance/time criterion was applied to access ways and response facilities to  
40 highlight remote or blocked blocks in danger of imminent eruption including evacuation.  
41 Statistical analysis of optimal variables from local scale surveys can help identify  
42 neighbourhoods where disaster risk mitigation requires improvement.

43 **Key-words:** human exposure; statistics; neighbourhood, volcano; Semeru; Indonesia.

44

## 45 **1. Introduction**

### 46 **1.1. Study background and terminology**

47 The present study is part of the ‘Local Adaptation to Volcanic Risk’ research project, and  
48 seeks to understand how and why dense rural communities continuously exposed to persistent  
49 volcanic threats can thrive on Semeru’s slopes and ring plain. Another paper explains how  
50 Semeru’s communities adjust to, compensate for, and tolerate continuous exposure to  
51 persistent volcanic threats (Thouret et al., 2022). Here, based on field surveys, mapping and  
52 statistical analyses, we seek to determine the population exposure to persistent volcanic  
53 threats by means of a composite index computed at the scale of neighbourhoods (termed  
54 blocks) in rural villages and small towns (Fig. 1, ESD Table 1). Using a second group of

55 variables on volcanic hazards and a third group on access and response, enabled us to rank  
56 and map a range of remote blocks in the case of imminent eruptions.

57 In this study, we use the international terminology on Disaster Risk Reduction ([UNDRR,](#)  
58 [2017](#); [UNISDR, 2017](#)) to which the reader is referred. Risk is the product of probability  $\times$   
59 losses, where the probability is a function of hazard and the losses depend on both exposure  
60 and external and internal vulnerability ([Aspinall and Blong, 2015](#)). Volcanic risk analysis has  
61 focused on cost-benefit evaluation (e.g., [Woo, 2015](#)) in particular the case of evacuations  
62 preceding imminent eruptions ([Jumadi et al., 2018](#), [Lechner and Rouleau, 2019](#)). The concept  
63 of risk also includes hazard knowledge, risk perception and resilience among communities  
64 living on active volcanoes ([Gaillard, 2008](#); [Gaillard and Dibben, 2008](#); [Paton et al., 2008](#);  
65 [Lavigne et al. 2008](#); [Donovan, 2010](#); [Donovan et al., 2018](#)). A broader risk concept stems  
66 from the appraisal of value systems and beliefs, governance systems and decisions, and  
67 political economies ([Bakkour et al., 2015](#)).

68 Exposure and vulnerability are two principal components of risk. Exposure is the location and  
69 spatial distribution of people, buildings, property, infrastructure, networks, lifelines,  
70 production capacities and other tangible human assets within reach of a hazard event  
71 ([UNDRR, 2017](#)). In the context of disaster risk reduction, [Wisner et al. \(2004\)](#) defined  
72 vulnerability as “the characteristics of a person or group to anticipate, cope with, resist or  
73 recover from the impact of a natural hazard”. Coping capacity is part of the widely used term  
74 of resilience, which is the ability of a community exposed to hazards to resist, absorb,  
75 accommodate to and recover from the effects of a hazard in a timely and efficient manner  
76 ([UNISDR, 2008, 2017](#)). Here we focus on exposure, a concept that has often been included in  
77 the vulnerability field; but exposure means being potentially affected by volcanic activity in a  
78 peculiar place, whereas vulnerability is the inability to withstand the effects of a harmful  
79 process. As stated by the [First IAVCEI-GVM workshop \(2018\)](#), “[exposure and vulnerability](#)  
80 [may be related: exposure quantifies number of people and/or buildings in the area, while](#)  
81 [vulnerability is one of the characteristics of the exposed elements that may suffer hazard](#)  
82 [impacts](#)”.

83 Studies measuring population exposure to volcanic hazards ([Ewert, 2007](#); [Kinvig et al., 2010](#);  
84 [Brown et al., 2015a](#); [Nieto-Torres et al., 2021](#)) are relatively few, compared with works on  
85 physical exposure of buildings and critical infrastructure (e.g., [Lerner-Lam, 2007](#); [Jenkins et](#)  
86 [al., 2014, 2015](#); [Wilson et al. 2014](#); [Jimenez et al., 2019](#)) and profuse investigations on

87 vulnerability and risk perception (e.g., Jóhannesdóttir and Gísladóttir, 2010; Zuccaro et al.,  
88 2015; Thouret et al., 2014a; Michellier et al., 2020). Yet, exposure, having both spatial and  
89 temporal patterns, is a fundamental component of risk analysis, and together with  
90 vulnerability, underpins loss. Exposure compels threatened people to adjust to risk through  
91 resilience to avoid the dangers of persistent volcanic activity.

## 92 **1.2. Global and other approaches to exposure**

93 We use local-scale surveys and statistical methods for defining exposure around Semeru. The  
94 main alternative is the Global approach. Exposure of people to volcanic risk has been the  
95 focus of the ‘priority countries of the Global Facility for Disaster Reduction and Recovery’  
96 (GFDRR) report by Aspinall et al. (2011), This pilot study on volcanic risk in 31 countries  
97 presented a method for “measuring the Volcano Population Index (the number of people  
98 threatened by each volcano: Ewert and Harpel, 2004) combined with the hazard level of each  
99 volcano to quantify population risk”. The Population Exposure Index (PEI: Aspinall et al.,  
100 2011, later summarized by Brown et al. (2015a) is one of the main indices used in assessing  
101 volcanic risk in highly populated areas together with Volcanic Hazard Index (VHI, Brown et  
102 al., 2015b), Human Development Index (HDI, UNDP, 2020) and Social Vulnerability Index  
103 (SoVI, The Hazards and Vulnerability Research Institute, University of Southern Carolina).  
104 VHI is an index rating the hazard level of a volcano based on the recurrence of past eruptions,  
105 the average and maximum Volcanic Explosivity Index ratings, and the extent of pyroclastic,  
106 mud, or lava flows in the eruptions. The Human Development Index (UNDP, 2020) measures  
107 levels of social and economic development based on “four criteria: mean and expected years  
108 of schooling, life expectancy at birth, and gross national income per capita”. The Social  
109 Vulnerability Index to hazards is based on “many socioeconomic, demographic, and built  
110 environment variables at the country or district level” (HVRI, University of South Carolina).

111 The GFDRR report (2011) estimated the numbers of people living within 10 km, 30 km and  
112 100 km of each volcano, which was weighted according to evidence on historical distributions  
113 of fatalities within a given distance from the vent. Population numbers were weighted,  
114 summed, and assigned to one of seven PEI scores in populated areas (Brown et al., 2015b).  
115 The authors estimated population risk for each volcano by “taking the product of the Hazard  
116 Level and PEI, and the numerical product was assigned to one of three Population Risk  
117 Levels”. Following this approach, Semeru volcano, one of the prominent Indonesia’s active  
118 volcanoes within the GFDRR category A, was assigned a high PEI level 3 with hazard level 3

119 and a low uncertainty level 1. The method relies on a number of parameters having high  
120 uncertainty or that are unknown, which restricts its application to volcanoes with good  
121 historical records.

122 A physical exposure component was introduced in the USGS re-assessment of US dangerous  
123 volcanoes. The updated Volcanic Threat Assessment (e.g., Ewert et al., 2018; Mangan et al.,  
124 2018) combines 15 hazard factors and 9 exposure factors that describe an individual volcano's  
125 hazard potential and the exposure of people and property to hazards. The exposure factors  
126 include population within 30 kilometers of the volcano; visitor numbers if it is located in a  
127 national park or monument; population beyond 30 kilometers if a far-traveling lahar is a  
128 primary hazard; prior eruption fatalities; prior evacuations; aviation impacts; impacts on  
129 power and transportation infrastructure; and major developments such as parks. A physical  
130 exposure component is also part of Del Negro et al (2020) quantification of lava flow risk on  
131 the flanks of Mt. Etna volcano using a GIS-based approach that integrates exposure of  
132 elements at stake within the hazard. The total exposure results from a weighted linear  
133 combination of four thematic layers, population, buildings, service networks, and land use, the  
134 weights of which were calculated using the Analytic Hierarchy Process. Population exposure  
135 relied on the density of population for each community normalized by the value obtained for  
136 the more populated municipality. Wood and Soulard (2009) proposed a composite exposure  
137 index for communities living in lahar-prone hazard zones around Mt. Rainier. They estimate  
138 the exposure index on the amount and percentage of six variables: developed land, residents,  
139 employees, public venues, dependent-population facilities, and parcel value, to estimate the  
140 spatial variations in community exposure.

141 More recently, Nieto-Torres et al (2021) quantified a multi-component volcanic risk index,  
142 which encompasses as many as 41 parameters: 9 parameters for hazard, 9 for exposure, 10 for  
143 vulnerability and 13 for resilience. However, the exposure criteria contain only one single  
144 human parameter, i.e., the density of population within 5, 10, 30 and 100 km radii from the  
145 main crater.

### 146 **1.3. Semeru volcano, setting and rationale for targeting continuous exposure**

147 Semeru, East Java, is Indonesia's highest and the southernmost volcano of the Tengger  
148 massif, which includes the Tengger caldera with the active Bromo cone (Fig. 2). The proximal  
149 (10 km), medial (30 km) and distal (100 km) areas exposed to the effects of volcanic hazards

150 rank Semeru as one of the world's most dangerous volcanoes according to the GFDRR report.  
151 It hosts one of the most exposed populations worldwide, as the 100-km circle distance  
152 encompasses Java's second densest populated area (see Freire et al., 2019, their Fig. 9),  
153 including 10 million people of the metropolitan area of Surabaya, Indonesia's second  
154 economic centre and second international airport. The province of East Java with an area of c.  
155 48,000 km<sup>2</sup> is home to about 40 million people, making it one of the most densely populated,  
156 largely rural areas on Earth with 830 people per km<sup>2</sup> on average. East Java presents typical  
157 low- to middle-income population at the lower range of the country's HDI (0.59-0.71; BPS  
158 2017) and mean SoVI score (Siagan et al., 2013). Java hosts 58% of Indonesia's population  
159 over < 7% of the country size, i.e., 141 million (2015) among which 10 % are exposed to  
160 disaster risk within 30 km from 22 historically active volcanoes.

161 Semeru is a real concern to civil authorities owing to the combination of daily explosive  
162 activity and a dense population: at least 950,000 people live within a radius of 35 km from the  
163 volcano summit (Thouret et al., 2022). Lahar-related disasters caused >10,000 casualties  
164 during the 20th century alone. The 1909 catastrophic event ranks fourth among the ten  
165 deadliest eruptions in the world between 1900 and 2010, as the 1909 PDCs and lahars killed  
166 at least 5500 people, i.e., 5.5% of all fatalities of the ten world's deadliest eruptions since  
167 1900 (Doocy, 2013). These events created havoc in thriving urban centres located in the  
168 eastern ring plain, in particular Pasirian, Tempeh, and Lumajang, harboring 123,000 people  
169 (as of 2015). Semeru's summit separates the Regencies of Lumajang to the East (1791 km<sup>2</sup>,  
170 1,036,000 people in 2015) and Malang to the West (3531 km<sup>2</sup>, 2,547,000 people) (Fig. 2;  
171 BPS, 2017). Here we focus on the regency of Lumajang where 621,000 people live within 35  
172 km of the summit, and the SW flank where 317,000 people represent a quarter of Malang's  
173 regency.

174 The study area hosts a range of areas exposed to deadly or disastrous pyroclastic density  
175 currents (PDCs), tephra fallout on proximal slopes, lahars along valleys as far as 35 km, and a  
176 variety of exposed assets, such as crops, road network, lifelines, factories, trade centres,  
177 religious monuments and touristic facilities in the ring plain. We defined and computed  
178 exposure of population living in 15 *dusun* of 6 districts (*Kecamatan*) with c. 800  
179 inhabitants/km<sup>2</sup> on average, although the size of rural settlements is highly variable between  
180 hamlets (350 people) and small, mixed rural/urban towns c. 7,500 people (ESD Table 1). The  
181 outlook of many of the towns is still rural as the growth of urban centres (Pronojiwo, Senduro,  
182 Candipuro) has not been accompanied by a parallel growth of industry. Urban centres benefit



183 from diversity, as they offer a variety of jobs and business such as agricultural trade, wood  
184 industries, traditional crafts, and local tourism. Towns around Semeru mirror other Javan  
185 centres in displaying great socio-economic diversity, which underlies a three-tiered social  
186 hierarchy. Villagers are ethnically Javanese with a significant Madurese minority. Agriculture  
187 and forestry are the main activities occupying 50% of the land and 40% of the workforce.  
188 Trade routes and small manufacturing are growing in towns, but roads and railways remain  
189 poorly developed.

#### 190 **1.4. Semeru's persistent eruptive activity and extensive hazard-prone areas**

191 Eruptive activity, which has been recorded since 1884, is usually mild (VEI 1-2), but  
192 increases every 8 to 11 years (Thouret et al., 2007; Solikhin et al., 2012). The constant activity  
193 includes four eruptive styles:

194 1. The persistent vulcanian and sometimes strombolian regime consists of short-lived, ash-  
195 laden < 0.5 km-high columns several times a day. Small columns usually disperse ash about 4  
196 km around the summit, but all villages can be affected by ash dispersed from 4-6 km-high  
197 columns as far as 8 to 12 km mostly to East, SE and NE, while exceptional seasonal winds  
198 can blow fine ash as far as 20 km ESE and WSW (Fig. 3).

199 2. Increased vulcanian/pelean activity every 8 to 11 years produces several kilometer-high  
200 eruption columns, ballistic bombs and thick tephra fall around the vent, and dispersed ashfall  
201 40 km downwind. Collapses from crater- and dome-fed, steep lava flows produce block-and-  
202 ash flows that travel toward the SE as far as 12 km from the summit, e.g., 2002 and 2020-21.

203 3. Flank 'aa' lava flows 5 km long erupted on the lower SE and E flanks in 1895 and in 1941–  
204 1942, while the crater termed Jenggring-Seloko (Fig. 3) has regularly produced stubby lava  
205 flows reaching 1 to 3.3 km along the steep scar open to the SSE over the past 40 years.

206 4. Highly explosive eruptions are not unknown. Nakada et al. (2019) refer to deposits from  
207 relatively large ( $\geq$  VEI 3) eruptions from the 3rd to the 11th Centuries as well as thick scoria  
208 falls and PDCs that destroyed the temple of the Majapahit Kingdom (13-16th Century) at  
209 Candi Jawar (5.5 km SW of the present summit), around the 16th century. The East and South  
210 flanks of Semeru together with the East and SE ring plain are the most prone to volcanic  
211 hazards where 50,000 to 100,000 people are exposed to the effects of eruptive activity.

212 **The hazard-zone map** (Fig. 3) displays three areas prone to volcanic threats as follows  
213 (Thouret et al., 2007, 2014b):

214 1) The high and steep summit cone within a circle of 5 km radius is affected by tephra-fall,  
215 lava flows and PDCs on the **SE** flank on daily to annual basis. No one lives there, but scores  
216 of tourists frequently climb and visit the summit area (5 victims were reported in 2000). New  
217 tephra fallout covers the summit cone on a daily basis, while dome-collapse rock avalanches  
218 and PDCs propagate annually as far as 4 km through the scar open to the **SE**, but they do not  
219 reach villages during the usual volcano “unrest” level 2 on a scale of 4.

220 2) The extensive South, **SE** and **ESE** flanks affected by PDCs as far as 12 km and lahars as far  
221 as 9 to 18 km from the vent, along which more than 50,000 people now live within 0.5 km of  
222 the active rivers (e.g., K. Koboan, K. Bang and K. Kembar, Fig. 3). On 4 December 2021,  
223 both margins of the Koboan valley were affected by exceptionally long runout PDCs and hot  
224 lahars flowing 16 km down valley and causing more than 50 fatalities (GVP, 2021).

225 3) The principal valleys draining the South (K. Glidik, K. Bang and K. Kembar), **SE** (K.  
226 Koboan), and East slopes (K. Tengah or Besuk Sat) convey many annual lahars across the  
227 ring plain at least 35 km down valley to the Indian Ocean (Thouret et al., 2014b). Post-  
228 eruptive lahars are the most frequent hazardous flows that propagate every rainy season along  
229 the principal valleys to the South, **SE** and East. Voluminous lahars, exceeding 5 million m<sup>3</sup>,  
230 killed hundreds of people at least five times since 1884; they swept the **SE** and **E** ring plain in  
231 1976, 1981, and devastated the city of Lumajang 35 km East in 1909 (Fig. 1).

232 In sum, the most exposed population live along the valleys within 8 to 25 km from the  
233 summit. Recorded fatalities due to lahars and PDCs are higher along the principal valley  
234 reaches between 9 and 12 km distance (75%). Fatalities due to lahars occurred between 12  
235 and 25 km (22%), the remainder (3%) being located along the > 25 km distal valleys in 1909,  
236 1976 and 1981 (Thouret et al., 2007).

### 237 **1.5. Location choice and local scale of observation: *Dusun* and neighbourhood**

238 We studied dense, mostly rural communities located in the exposed areas between 8 and 20  
239 km from the summit of Semeru noting past eruptions and reported fatalities (Fig. 3; Thouret et  
240 al., 2007, their Table 1). The primary unit of study was neighbourhood (block) in sub-villages  
241 (termed *dusun*) located on the valley margins, terraces and interfluves within 0.5 km from the  
242 active channel that used to be affected by lahars and PDCs in the recent past. Locations

243 included the South (K. Kali= river) Bang, K. Kembar, e.g., 2002-2003), SSE (K. Koboan and  
244 C. Lengkong, 1994-1995, 2020-2021), East (K. Tengah/ Besuk Smut, 1981), and the SE and  
245 East ring plain (K. Rejali, Mujur) towards the city of Lumajang, affected in 1909 (Fig. 3). We  
246 also studied villages farther away from active rivers and the town of Senduro outside the  
247 affected valleys to avoid any bias that would stem from the exposed settlements in affected  
248 hazard zones (Thouret et al., 2022). The remaining *dusun* within the reach of light ashfall  
249 were selected for the purpose of comparison as they extend beyond active valleys (e.g.,  
250 Pasrujambe, Fig. 2) and farther away from the volcano. We examined small rural  
251 communities at higher altitude on the west flank that are exposed to frequent tephra-fall  
252 within 9 km of the crater (Fig. 2). We also considered mixed rural/urban communities in two  
253 towns: Pronojiwo in the vicinity of the active K. Bang and Kembar valleys, and Senduro  
254 farther away (17 km) from the volcano and any active river.

255 We selected 15 *dusun* that belong to six *Desa* on the SW, S and E slopes of Semeru inside  
256 four *Kecamatan* in the Regency of Lumajang and one in Malang Regency (Table 1, Figs. 2-3;  
257 ESD Fig. 1) based on the following rationale:

258 1. Both *dusun* Blubuk and Karanguko (*Desa* Tamansatryan), located between 900 and 1250  
259 m asl on the WSW flank 8-9 km from the vent, are exposed to frequent, light tephra-fall from  
260 the volcano. Both mid-altitude hamlets exemplify chronic exposure of a population living  
261 close to the crater.

262 2. The small town of Pronojiwo (*Desa* and *Kecamatan* Pronojiwo) is the largest settlement (c.  
263 7,600 people) on the South flank of Semeru between 600 and 700 m asl and at a 11-12 km  
264 distance of the summit. The rural suburb to the East of the town near two large rivers (K.  
265 Bang and K. Kembar) has regularly been affected by PDCs and annual lahars. On the ENE  
266 side of Pronojiwo, the *dusun* Supit and Supit Timur 9 to 11 km from the vent and between  
267 720 to 820 m asl are most exposed to volcanic threats (PDC, lahars, and tephra-fall) along the  
268 West margin of the K. Bang, with a few casualties due to overbank PDCs reported in Supit  
269 Timur in 2002. *Dusun* Rowobaung is located on a volcanoclastic fan between two active rivers  
270 at 11-12 km from the summit. These *dusun* host exposed dwellers living in relatively high  
271 magnitude and frequency PDC- and lahar-prone zones.

272 3. On the SE flank of Semeru, a group of four *dusun* belong to *Desa* Supit Urang. Oro-oro  
273 Ombo, away from active valleys, is less exposed than Summersari and Gumuk Mas, sitting on  
274 a high terrace to the South of the K. Koboan 9 to 11 km from the summit. Curah Koboan is

275 situated 9-10 km from the summit on a high terrace of Curah Lengkong, a tributary to K.  
276 Koboan. The population of these *dusun* live in high magnitude / frequency PDC-prone areas,  
277 with fatalities in 1994, 1995, 2021, and paid a heavy toll in 2020-2021.

278 4. Kajar Kuning (*Desa Sumberwuluh; Kecamatan Candipuro*) is located 0.5 km North of  
279 Curah Koboan at the foot of hills on which the Volcano Observatory sits at Gunung Sawur.  
280 We chose this hamlet to observe how the location away from the Lengkong River would  
281 influence the response of less exposed blocks.

282 5. Both *dusun* Jabon and Tulungrejo (*Desa and Kecamatan Pasrujambe*) are located 10-13 km  
283 away on the Semeru East slope on the North side of the K. Tengah valley. The *dusun*  
284 Tulungrejo spreads out on the low and middle terraces of the active river, while Jabon is less  
285 exposed on ridges 0.8 km north of K. Tengah. We targeted these *dusun* because the Tengah  
286 valley was the site of many PDC- and lahar-related fatalities of the 1981 VEI >3 eruption.

287 6. Farther East on the ring plain, three *dusun* Sumbermulyo, Juranglangak and Rekesan (*Desa*  
288 *and Kecamatan Senduro*) are located in the SE area of the town of Senduro (population c.  
289 7,500) 18-19 km from the summit. Senduro, only 20 km away from *Kabupaten Lumajang*, is  
290 also a touristic hub, hosting historic Hindu temples, and a gateway to the Semeru-Tengger  
291 massif. These *dusun* host a mixed urban/rural population well away from the volcano and  
292 active rivers, hence much less exposed, in a low-frequency fallout-prone zone.

## 293 **2. Methods**

### 294 **2.1. Data acquisition and survey procedure**

295 The SE, East and NE slopes of Semeru were the target of two field campaigns in September  
296 2018 and 2019. Four researchers from the IPB University, Bogor, the University Gadjah  
297 Mada, Yogyakarta (Indonesia), and the Laboratoire Magmas et Volcans, Université Clermont-  
298 Auvergne (France) conducted field observations on the population and home exposure to the  
299 impacts from PDCs, lahars and tephra fall. A Trimble TC 1000 GIS mapper device, Google  
300 Earth maps and topographic maps (scale 1:25,000 as of 2000) have been used in QGIS to  
301 locate the surveyed sites and collate structural observations on edifices (homes, offices,  
302 schools, health centres, mosques, and markets: Fig. 4; ESD Tables 1, 2). Using a satellite  
303 image SPOT5 (pixel: 2.5 m) as of 2014 and Google Earth maps, we outlined the boundaries  
304 of *dusun* and built or un-built blocks in the vicinity of principal valleys that convey lahars or  
305 PDCs.

306 The campaigns provided a representative number of observations, as we collected data from  
307 279 households and we mapped 145 blocks in 13 rural *dusun* and 2 small towns. A *dusun*  
308 usually entails 4 to 5 *RukunWarga* (RW, a set of 50-75 buildings, mostly homes with a few  
309 offices, mosques or schools), which in turn includes 3 to 5 *RukunTetanga* (RT, 20-25 homes).  
310 We collected data from 2 to 4 households per RT, the smallest administrative unit at which  
311 respondents can be identified. A *dusun* has an average area of 2.07 km<sup>2</sup> and density of 806  
312 inhabitants/km<sup>2</sup>, while a built block area is 0.28 km<sup>2</sup> with a density of 1,397 inhabitants/km<sup>2</sup>  
313 on average (ESD Table 1). We categorized the quality of construction and roof, and building  
314 orientation, as these parameters play a role on dwellers' exposure to tephra fall and volcanic  
315 flows (ESD Table 2). To complete the 2017 BPS census at the smallest scale possible, we also  
316 collected data on population, economic activity (mostly agriculture, husbandry and  
317 agroforestry) and emergency facilities from eight governmental offices in: *Desa* (sub-district)  
318 Tamansatryan (WSW flank), *Desa* and *Kecamatan* (district) Pronojiwo (South), *Desa* Oro-oro  
319 Ombo and Supit Urang (SSE), *Kecamatan* Candipuro and *Desa* Sumberwuluh (SE), *Desa*  
320 Pasrujambe, and *Desa* Senduro (East) (Figs. 2, 3; ESD Table 1).

321 Few limitations encumbered the field survey on multi-component exposure of people and  
322 dwellings, which was conducted in Indonesian and/or Javanese thanks to our partners. We  
323 interviewed respondents available in the household, either male or female, through a random  
324 door-to-door approach and with the interpreter's help. One limitation may be  
325 representativeness. Among 950,000 people around Semeru, 50,000 live along active rivers,  
326 and up to 100,000 people in areas where eruptive activity exceeds VEI 3. Surveys involving  
327 about 279 households and 300 buildings represent approximately 0.6% to 2.6% of the  
328 respective cohorts. However, mapping the exposure index involved 145 blocks, hence 6% of  
329 the exposed population.

## 330 2.2. Statistical analyses

331 Statistical analyses explored a set of 23 variables collected through surveys and organized in  
332 three fields: exposure, hazard, and access/response (Fig. 1, Tables 2, 3). The first group of 11  
333 variables and their 43 attributes helped compute the population exposure index of 145 blocks  
334 (see 2.2.1 below). The second group of seven variables determined the hazard level of blocks  
335 both inside and outside the hazard-prone zones around Semeru (Fig. 3; Thouret et al., 2007).  
336 The third group of five variables defined the access to blocks and response facilities for  
337 imminent evacuation.

338 [Table 1](#) summarizes the statistical methods and techniques together with their purposes for  
339 coding, quantifying and validating three groups of variables and their attributes. Statistical  
340 investigation conducted with the R software has involved four operations.

341 First, univariate and bivariate analyses were used to extract relevant attributes from variable  
342 observations and detect their relationships.

343 Second, Polytomous Logistic Regression models helped select optimal exposure variables and  
344 predict the exposure index of populated neighbourhoods (EIPN). Logistic regression is a  
345 technique commonly applied to highlight dependences between one variable that must be  
346 explained (dependent, endogenous, here the Exposure Index of a block) and several variables  
347 that explain relationships (independent, exogenous, quantitative and qualitative, ordinal or  
348 nominal). Logistic regression allows a model to be developed while the selection between  
349 models is done through a set of criteria such as reduction of the Akaike Information Criterion  
350 (AIC; Akaike, 1987) and confusion matrix. AIC was designed as the divergence between the  
351 true model (that actually generated the data) and a proposed statistic approximation of this  
352 model. Polytomous logistical regression (PLR) was performed using Cumulative Links  
353 Models to quantify the strength of association between each active variable and the response  
354 variable (EIPN score) to be explained.

355 Third, multiple correspondence analyses (MCA) of all variables cross-referenced with the  
356 EIPN score were used to characterize the blocks. Hierarchical Ascending Classification  
357 (HAC) was used to discriminate clusters amongst blocks.

358 Fourth, we used the distance/timing criterion combined with the third group of access and  
359 response variables to rank blocks with respect to relief operations for evacuation.

### 360 **2.2.1. Methods to compute the population exposure index in neighbourhoods**

361 To compute the Exposure Index of Populated Neighbourhoods, we used five techniques ([Fig.](#)  
362 [1, Table 1](#)):

363 1. A normalization technique was used to avoid biases arising from the fact that all qualitative  
364 variables do not have the same number of attributes. For each attribute,  $x$  being the initial  
365 value, we subtract the minimum value and we divide by the maximum interval, as follows:

366  $x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \in [0,1]$  Thus, all values of qualitative variables are equidistributed

367 between 0 and 1.

368 2. A Chi<sup>2</sup> test was performed to determine whether the variables are dependent or independent  
369 (ESD Table 3). In other words, to find out whether the difference between observed and  
370 expected data is random or due to a relationship between the variables under study.

371 3. The non-parametric Wilcoxon test (Wilcoxon, 1945) was used because some pairs of  
372 variables were identified as independent according to the Chi<sup>2</sup> test. This test allowed us to test  
373 whether two variables show similar distributions.

374 Based on results of the Chi<sup>2</sup> and Wilcoxon tests, we determined that all variables were  
375 dependent and followed similar distributions. The sum of the normalized values of 11  
376 variables and their 43 attributes for each of the 145 blocks provided the EIPN score.

377 4. The Shapiro-Wilks test (Shapiro and Wilks, 1965) can be used to determine whether a  
378 particular dataset follows a normal distribution, which is a common assumption used in many  
379 statistical tests. The Shapiro-Wilks test is based on the null hypothesis according to which the  
380 EIPN score follows a normal distribution. We discarded the null hypothesis when the p-value  
381 was lower than the  $\alpha$  risk fixed to 0.5 (5%).

382 5. Discretization was based on Jenks optimization method or natural breaks classification  
383 method (Jenks, 1967) because the distribution of values does not follow a normal distribution  
384 and the number of attributes varies between variables. The EIPN score parametric distance  
385 between classes is calculated using the Jenks technique. As it minimizes the intraclass  
386 variance while maximizing the interclass variance, we fix the number of breaks, i.e., number  
387 of discretized classes. The Jenks-type discretization, as opposed to other methods, allowed us  
388 to fit data to the shape of the statistical distribution, while it provided four homogenous  
389 classes.

### 390 **2.2.2. Statistical analyses to characterize block exposure, hazards, and access/response**

391 1. Univariate analysis (UA, Table 1) is the first step for exploring and preparing a dataset for  
392 further analysis. UA summarizes descriptive statistics and provides graphical representations  
393 for their univariate distribution (e.g., Chambers et al., 2018).

394 2. Bivariate analysis (BA, Table 1) involves the analysis of two variables with the aim to test  
395 simple hypotheses of association or any relationship between two variables and attributes. The  
396 Chi-square test aims to compare observed results with the expected results (see the test  
397 purpose in section 2.2.1 above; ESD Table 3). Factorial Correspondence Analysis (FCA) is

398 used to study links between two qualitative variables and explore correlations and oppositions  
399 of categorical variables in a table of frequencies or contingency (ESD Table 4). From these  
400 correlations and oppositions, FCA thus allows us to state hypotheses to identify typical blocks  
401 in the Semeru *dusun*.

402 3. Multiple Correspondence Analysis (MCA) is run when a set of observations includes  
403 multiple qualitative variables. MCA is a data analysis technique used to detect and represent  
404 underlying structures in a large categorical dataset (Abdi and Valentin, 2007). This technique  
405 represents data as points, illustrated by biplot graphs. MCA consists of the following steps  
406 (Table 1; ESD Fig. 3):

407 (1) Eigenvalues helped define the number of retained dimensions having the highest inertia  
408 rate, hence the most relevant information from 81 attributes of 23 variables. Benzécri  
409 correction (Benzécri, 1979) applied to the number of dimensions (inertia) helps estimate how  
410 much information is included in each dimension and select the smallest number of dimensions  
411 which contain the maximum information available (ESD Fig. 2). We retained the first three  
412 dimensions, which contained almost 91% of the information.

413 (2) Biplots show the relationships between attributes of all variables, while taking into  
414 account the quality of representation quantified by square cosines, contribution and  
415 coordinates of each attribute.

416 (3) Scatter plots establish relationships between the attributes among the same qualitative  
417 variable. Confidence ellipses highlight proximities (closeness) and oppositions between the  
418 attributes of one variable in the projection of blocks along a given axis in scatterplots.  
419 MCA procedure includes two tests: Fisher test to highlight the correlation between a given  
420 variable and the factorial axis and Student test to verify which attributes have singular  
421 coordinates along the axis (dimension). ESD Table 5 displays the coordinates, squared  
422 cosines and contributions of the attributes with respect to dimensions. These contributions  
423 help select the contributory variables retaining more information and their attributes. Squared  
424 cosines show the quality of representation and help select variables with sufficient quality  
425 within the most contributory ones. Coordinates of this table indicate the position of attributes,  
426 the most distant from the barycentre of each biplot contributing most.

427 4. Hierarchical Agglomerative Clustering (HAC; see Table 1) was performed on the MCA  
428 outputs to confirm previous MCA groups and/or construct groups of blocks sharing similar  
429 statistical characteristics. HAC leads to a factor map identifying four groups of blocks (ESD  
430 Fig. 4) based on results from frequencies of attributes.



431

### 432 3. Statistical results

433 The statistical analysis using UA, BA and MCA techniques of 11 exposure variables and their  
434 attributes (Table 5, ESD Table 3) aimed to predict the variations in the index of population  
435 exposure between blocks.

#### 436 3.1. Defining variables for the multi-component exposure index

437 To compute the EIPN at the block scale, we used 11 variables and their attributes: 9 variables  
438 increase the exposure to volcanic threats, to which we added two defining the exposure of  
439 dwellings (Fig. 1; Table 2; ESD Table 3). Each variable shows attributes divided by  
440 thresholds, amounting to 43 attributes (Table 2). A number from 0 to 5 is attributed to each  
441 attribute, although variable V8 exceptionally has 10 attributes (Table 2). One variable V4 only  
442 was attributed a weighting coefficient (see the method involving four steps: Table 3). The  
443 rationale for each variable and its attributes relies on risk-related exposure works (e.g., Ewert,  
444 2007; Ewert et al., 2018; Loughlin et al., 2015; Nieto-Torres et al., 2021), as follows:

445 **V1:** Three circle radii around the vent (Fig. 2) are: a maximum 9 km distance at which  
446 ballistics and heavy (lapilli-sized) tephra-fall occur and may impact roofs commonly made of  
447 tiles supported by light timber; a maximum 12 km distance at which the longest PDCs  
448 reached settlements in recent history (e.g., in 2002 with VEI 3), and an average 18 km  
449 distance for lahars, exceptionally 35 km to the sea for the most voluminous ones (e.g., in  
450 2021). The energy line concept (Malin and Sheridan, 1982) yields maximum runout distances  
451 between 12 and 18 km for Semeru's PDCs while the 30-35 km range is commonly reported  
452 for voluminous lahars (e.g., Ewert et al., 2005). Each of the circle radii were assigned the  
453 following values based on the distribution of *dusun*: 9-12 km= 3, as 71% of *dusun* are located  
454 and most of prior fatalities occurred within this range; 12 to 18 km= 2 (22% of *dusun*); and <  
455 9 km= 1 (7% of *dusun*).

456 **V2:** We retrieved population density at the *dusun* scale from the 2017 census (BPS, 2017).

457 **V3:** Population density within the built blocks was estimated from house counting (4 persons  
458 on average per household) on Google maps and field observations with an error range of c.  
459 10%. We ranked densities in increasing order from the lowest to the highest threshold (using  
460 quartile interval) to reflect the increasing effect on people exposure in both *dusun* and blocks  
461 (Table 2).

462 **V4:** Table 3 shows the four steps of the methods used to compute the coefficients for V4. The  
463 inferred number of fatalities from the recorded eruption dataset (Table 2 in Thouret et al.,  
464 2007), total 1,115 victims from 18 events reported since 1884. As about 950,000 people live  
465 in the 35 km-distance circle with an area of ca. 1380 km<sup>2</sup> around the volcano, the ratio yields  
466 c. 1175 victims per million people, c. 8.5 victims per year, and c. 62 fatalities per event on  
467 average. These figures are likely under-estimated, while numbers of injured or displaced  
468 people are possibly large.

469 Using the hazard zones affected by the historical eruptive activity and lahars (Fig. 3) and the  
470 number and approximate location of fatalities, we chose the circle distances at 9, 12, 18 and  
471 35 km. Population numbers in villages per circle area are 2,864 with radius distances of 9 km,  
472 59,236 with 12 km; 85,303 with 18 km, and 842,597 people with 35 km radius, based on our  
473 calculation of the number of houses in the circles 1 to 3, and on the BPS 2017 census beyond  
474 18 km. The weighted population (fatalities and distances) according to Brown et al.'s (2015a)  
475 method (Table 3) is 103 in circle of 9 km radius; 4,858 in circle of 12 km; 15,739 in circle of  
476 18 km; and 587,800 in circle of 35 km radius. As 1,115 fatalities are reported for a population  
477 of 26,432 living in the 15 surveyed *dusun* based on 9 km, 12 and 18 km circle distances (Fig.  
478 2), the final weighting coefficient for V4 is 0.036 for the 9-km circle, 0.082 for the 12-km  
479 circle, 0.185 for the 18-km circle, and 0.698 for the 35-km circle (Table 3).

480 **V5:** Location of fatalities with respect to valleys that transmit flows across three sites with  
481 values ranked from 3 to 0 according to hazard-type effects: 3 is given to valley margins  
482 reached by pyroclastic surges and overbank lahar flows; 2 is assigned to valley channels  
483 conveying confined pyroclastic flows and lahars, and terraces or interfluves swept by ash-  
484 cloud surges up to 500 m from the river; 0 indicates 'no reported death' along valleys (Table  
485 2).

486 **V6:** Lahar fatalities is a binary variable (Yes 1/No 0) and both figures 1,0 have been  
487 distributed along the valley channels, terraces and margins.

488 **V7:** PDC fatalities is also binary (Yes 1/No 0), and spatially distributed, valley-confined flows  
489 being distinct from un-confined surges on valley margins and adjacent interfluves.

490 **V8:** Location of mapped blocks on terraces is based on a range of distances from, and  
491 elevations above the river channel. Ten cases based on both criteria involved distance  
492 thresholds from 15-50 m to > 120 m, and elevation thresholds from < 2 m to > 30 m. We  
493 attributed marks 10 to 1 based on decreasing exposure to lahars and PDCs, from the  
494 maximum 10 for channels and the minimum 1 for high and distant terraces.

495 **V9:** Timely location of people living and/or working on Semeru's flanks encompass three  
496 situations in decreasing exposure duration order: permanent, daily work and home at night;  
497 temporary, home at night only, and; temporary, daily work only and night elsewhere.  
498 To the 9 variables that define human exposure we added two variables that usually are criteria  
499 defining the physical vulnerability of buildings. Here we consider them within the exposure  
500 field, as they determine the extent to which residents are exposed to volcanic threats owing to  
501 tephra fallout load on roofs and pressure impacts from PDCs and lahars on homes.

502 **V10:** Orientation of house, i.e., perpendicular, oblique or parallel to PDC and lahars.

503 **V11:** The quality and type of roof and construction are ranked as poor (wood), regular (zinc)  
504 with light timber, and fair when the roof is concrete and/or covered by tiles supported by  
505 sturdy timber. Almost all roofs, being pitched and many being covered by tiles or zinc, could  
506 probably withstand about 20 cm-thick tephra-fall, but concomitant, frequent rainfall would  
507 diminish the resistance threshold, in particular for roofs of light timber.

### 508 **3.2. Computing the exposure index of populated neighbourhoods (EIPN)**

509 The table in **Figure 5** and **ESD Table 6** display the EIPN score (from 1.17 to 8.56, median  
510 4.17) obtained for each block after discretization. **ESD Table 6** shows columns as variables  
511 and their thresholded attributes, while rows are the 145 surveyed *dusun* blocks. As a result,  
512 four colour-coded EIPN scores levels, from 'residual' (beige) to 'very high' (red), have been  
513 attributed to each of the 145 blocks (**Fig. 5 A-F**). The relatively narrow range of EIPN  
514 (standard deviation c. 1.05) reflects a relatively low exposure of all *dusun* to the potential  
515 effects of the Semeru activity except those which are located inside the proximal hazard zones  
516 and along valley channels depicted in **Figure 3**. The EIPN score displays three groups of  
517 blocks (**Figs. 4**): c. 39% of blocks show a moderate to high EIPN, and almost 25%, located in  
518 the vicinity of valley channels, show a very high EIPN. On the other end, almost 37% of the  
519 blocks show a very low or residual EIPN score, which correspond to older neighbourhoods  
520 that used to settle away from the active valleys. At this stage, we need to explore the  
521 relationships between the attributes of exposure variables to highlight the most contributory  
522 ones.

#### 523 **3.2.1. Relationships between attributes of exposure variables**

524 MCA scatter plots represent attributes using elliptic envelopes of confidence around the  
525 barycentre of groups (**Fig. 6**). Ellipses of 36 attributes that belong to the 11 exposure variables

526 were plotted in the dimensions 1 and 2 scatter plots (Fig. 6). A number of opposed attributes  
527 (i.e. away from the barycentre) lead to the following inferences:

528 1) The 9-12 km circle distance stands out among other distances from the vent, coinciding  
529 with the majority of exposed *dusun* within this circle. 2) In V2 and V3 scatter plots as well as  
530 in the V9 plot (timely location), the ‘0 density’ stands out, which identifies temporary people  
531 location and un-built blocks. 3) The attributes valley channel and valley margins, although  
532 they are less distinct from other ‘fatalities locations’, coincide with river channels and low-  
533 altitude terraces that are more exposed to lahars and PDCs. 4) In V6 and V7 scatter plots, the  
534 presence/absence of prior fatalities due to PDCs and lahars also suggest the extent to which  
535 blocks are exposed. 5) The V8 plot shows that the relationship distance/elevation to the  
536 channel is not discriminant enough, except for the river channel. 6) The V10 and V11 scatter  
537 plots also suggest that building parameters (quality of roof, construction, and house  
538 orientation) may be significant to dwellers’ exposure.

539 As a result, the following eight contributory attributes of nine exposure variables may  
540 characterize the block exposure: the 9-12 km circle distance (V1), the *dusun* and block density  
541 (V2, V3) and temporary location (V9), the channel and low terrace location for prior fatalities  
542 (V8), the presence of PDCs and lahars (V6, V7), and both building parameters (V10, V11).  
543 For the sake of research efficiency, we reduced the number of variables to be collected in the  
544 field for predicting the EIPN.

### 545 **3.3. Predictive capacity of Polytomous Logistic Regression to obtain EIPN scores**

546 Polytomous Logistic Regression (PLR) is one of the logistic regressions adapted to the study  
547 of categorical and/or ordinal variables with more than two attributes (Kleinbaum and Klein,  
548 2010). As the EIPN score, variable of interest, is qualitative, PLR is the most adequate type of  
549 regression, involving a succession of techniques (Table 1):

#### 550 **3.3.1. Selection of a small number of active variables, and model significance**

551 The initial PLR model included 11 variables and their 43 attributes (Table 4A). In fact, the  
552 selection procedure reduced 11 variables down to 8 to avoid collinearity (a linear relationship  
553 between two explanatory variables); for example, V1 distance to vent and V4 ‘Brown’s PEI’  
554 parameter are dependent and collinear (i.e., retaining the same information and same  
555 distribution) with V2 and V3 (population density), hence V4 was not retained (Table 4A). We  
556 performed backward and forward model selection procedures, which consisted in sequentially  
557 adding or removing variables and investigating the resulting model performances. During  
558 backward or forward selection processes, AIC offers a compromise between parsimony and

559 error reduction by penalizing a high number of parameters. The model having the smallest  
560 AIC possible was chosen to select optimal variables.

561 [Table 4A](#) shows the initial PLR model with estimated coefficients of the selected attributes.  
562 ‘Significativity’, indicating that the underlying model coefficients are significantly different  
563 from 0, is the most important criterion for prediction. Based on the significativity, ‘3 stars’ are  
564 attributed to attributes with < 0.1% error in the model, ‘2 stars’ to attributes with < 1% error  
565 and ‘1 star’ to those with < 5% error ([Table 4B](#)). According to the significativity results, a  
566 rigorous model would only include eleven 3-stars attributes with high probability. The selected  
567 PLR model ([Table 4B](#)) then reduced the non-collinear 8 variables to 6 only: V1 distance to  
568 vent; V2 density of *dusun*, V3 density of blocks, V5 Location of prior fatalities, V6 Lahar  
569 fatalities, and V7 PDC fatalities. The model also reduces the 43 attributes down to 15, as shown  
570 in [Table 4B](#).

571 The selected model comprises both positive and negative coefficients ([Table 4B](#)). Positive  
572 coefficients mean that when attributes increase, the EIPN score increases as well, while  
573 negative coefficients diminish the EIPN score. Thus, the attributes of the following variables:  
574 V1 distance to vent; V2 density of *dusun*, V3 density of blocks and V5 Location of prior  
575 fatalities will increase the chance that the blocks score a high EIPN, i.e., are more exposed.  
576 Conversely, blocks that possess low thresholds of both variables, V6 Lahar fatalities, and V7  
577 PDC fatalities, will have a weak EIPN score: both hazards threatening these blocks are frequent,  
578 but they do not imply a large number of fatalities.

### 579 **3.3.2. Probability of obtaining a block exposure index and quality of the model**

580 PLR allows us to compute the probability of getting EIPN for any given block. We first  
581 assessed the quality of the PLR model ([Table 4](#)): the calculated MSE is 0.179, while the  
582 efficient ranking rate is as high as 82.07%. We used a Table termed confusion matrix ([ESD](#)  
583 [Table 7](#)), which crosses ‘real’ observed EIPN scores with predicted ones, to test the quality of  
584 the estimated model. For a ‘perfect’ model, the confusion matrix would show a diagonal  
585 matrix in which non-diagonal entries are all equal to zero. The model is efficient if the  
586 number of non-diagonal, predicted values is below 40% of the total number of observed  
587 values (Kleinbaum and Klein, 2010). Among 145 observed blocks, we cannot predict 26,  
588 yielding a relatively low error rate of c. 18%, which ensures relative efficiency to the model.

589 Two blocks located in the town of Pronojiwo and in the *dusun* Rowobaung ([Fig. 5A](#)) are given  
590 in the table below as examples based on specific attributes of six variables: R3, away from both

591 active rivers (K. Kembar, K. Bang), has been attributed a residual EIPN score, whereas P1,  
 592 located near the K. Bang valley channel, has been attributed a very high EIPN score in maps of  
 593 exposed blocks (Fig. 5A). We compute the EIPN probability of both blocks in order to test how  
 594 well PLR predicts the score:

Block no.P1 Pronojiwo Fig. 5A	V1 distance to vent	V2 density of <i>dusun</i>	V3 density of block	V5 Fatality location	V6 Lahar fatalities	V7 PDC fatalities
P1	9 - 12 km	1,091-1,690	891-1,340	Valley channel	No	Yes
R3	9 - 12 km	349 - 823		Un-reported fatality	No	No

595

596 Using the function ‘predict’, the probabilistic results for P1 block are as follows:

EIPN score prediction for block P1			
Very high	Moderate - High	Very Low - Low	Residual
0.9964144	0.003583707	1.850412e-06	2.701923e-09

597

598 Thus, the P1 block has 99.64% probability of obtaining a very high EIPN score. The result  
 599 shows that the selected model is working well, as we calculated a very high EIPN score for P1  
 600 (Fig. 5A). Using the same function, the probabilistic results for the R3 block yields:

EIPN score prediction for block R3			
Very high	Moderate - High	Very Low - Low	Residual
1.602824e-07	0.0003109862	0.1756036	0.8240853

601

602 Thus, the R3 block has 82.41% probability of obtaining a residual EIPN score. The result  
 603 shows again that the selected model is valid, as we calculated a residual EIPN score for R3  
 604 (Fig. 5A).

#### 604 4. Mapping the composite exposure index at the scale of *dusun* neighbourhoods

605 Maps display levels of the colour-coded EIPN score of 145 blocks (Fig. 5 A-F, ESD Table 6).  
 606 The most exposed, active valleys convey lahars and PDCs, in particular the large valley  
 607 channels (K. Koboan to the SE, K. Bang and Kembar to the S, and K. Tengah to the E). These  
 608 wide channels host no permanent homes, but only temporary shelters and small makeshift  
 609 shops, located on low terraces close to the river. Workers extracting material are highly  
 610 exposed to lahar and PDC impacts.

611 1. Most exposed *dusun* blocks to the current eruptive activity of Semeru are located along the  
 612 principal valleys that convey lahars and/or that have been affected by PDCs over the past 140

613 years. These are Supit Timur near K. Bang, North and West blocks of Rowobaung between K.  
614 Bang and K. Kembar (Fig. 5A), affected in 2002. Highly exposed are the North blocks of  
615 Summersari and Curah Koboan on both margins of K. Koboan (heavily damaged in December  
616 2020 and January 2021), and on the North bank of Curah Lengkong (Fig. 5B). The blocks of  
617 Tulungrejo (Pasrujambe) on the low and middle banks of K. Tengah are also highly exposed  
618 (Fig. 5C), while dykes built in the years 1970-1980 along the North bank of K. Tengah may  
619 provide a false sense of security.

620 2. Increasing eruptive activity producing the VEI 3 events every 8 to 11 years. At such times  
621 additional blocks would suffer heavy tephra-fall, lahars, and ash-cloud surges from PDCs  
622 along river valleys. Such *dusun* blocks encompass Supit along K. Bang, the East blocks of  
623 Pronojiwo near the bridge across K. Bang, most of Rowobaung between K. Bang and K.  
624 Kembar (Fig. 5A), most of Summersari and part of Gumuk Mas and Curah Koboan near K.  
625 Koboan (Fig. 5B), and most of Tulungrejo in the Desa Pasrujambe (Fig. 5C).

626 3. Other *dusun* blocks located down valley on low terraces can be affected by large lahars,  
627 e.g., K. Koboan down valley as far as Sumberwuluh (ESD Fig. 1D), as happened in January  
628 2021, or by ash-cloud surges as far as Tawonsongo upstream of K. Tengah (Fig. 5C). Heavy  
629 tephra-fall can affect the *dusun* of Blubuk and Karangusko in large  $VEI \geq 3$  eruptions (Fig.  
630 5D).

631 4. Blocks with very low to residual EIPN characterize the town of Senduro 17-19 km away  
632 from Semeru vent and away from active rivers (Fig. 5E). Senduro's dwellers are the least  
633 exposed, except for tephra-fall dispersed from Semeru in large ( $VEI \geq 4$ ) but un-frequent  
634 eruptions, and from Bromo (Tengger caldera) in 2009-2011, and 2020.

635 5. Contrasting EIPN scores of blocks situated in the same *dusun* are due to spatial changes  
636 and specific attributes at a local scale, such as density, hazard frequency, home quality and  
637 orientation, and dirt roads instead of paved roads (e.g., Supit and Pronojiwo, Fig. 5A).

638 Decreasing EIPN to moderate level across *dusun* is due to decreasing densities and increasing  
639 elevation and/or distance from the channel (e.g., North and East blocks of *dusun* Curah  
640 Koboan and Summersari down valley, Fig. 5B, upper area of Tulungrejo and Tawonsongo,  
641 Fig. 5C). Different EIPN scores are also due to the distance to health centres and locations  
642 with limited lahar-related fatalities (e.g., Blubuk and Karangusko, Fig. 5D).

643 EIPN maps reflect the 2018-2019 situation when Semeru eruptive activity was mild (VEI 2)  
644 and the alert status was 2 on a scale of 4. As Semeru activity escalated to VEI 3 in December  
645 2020 and January 2021, triggering dome-collapse PDCs and tephra-fall from > 6 km-high

646 columns, together with hot lahars (GVP, 2022), every EIPN score would increase to the next  
647 higher colour-coded level (see EIPN Table in Fig. 5A-C).

## 648 **5. Characterizing blocks (hazard, access/response) based on statistical analyses**

649 We used UA, BA and multivariate analyses to define the characteristics of blocks based on  
650 hazard and access/response variables, compared with the EIPN score. We defined the  
651 variables of the second and third groups as follows:

### 652 **5.1. Second group of variables: volcanic hazards**

653 The second group of variables (Table 2) determines the hazard level of neighbourhoods inside  
654 and outside the hazard-prone zones outlined around Semeru (Fig. 3; see Thouret et al., 2007  
655 for the type, occurrence and frequency of lahars, PDCs, and tephra fallout.)

656 **V12:** Volcanic Explosive Index is ranked according to current, daily (since 1967) and chronic,  
657 episodes of increased eruptive activity: VEI 1-2 unrest and mild activity, VEI 3 every 8 to 11  
658 years on average, and  $VEI \geq 4$  every 25 years on average.

659 **V13:** Two types of lahars are debris flows (DF) and hyperconcentrated flows (HCF), ranked  
660 in three attributes: DF and/or HCF, HCF alone, and absent. Frequent lahars occur every rain  
661 season, and shortly after eruption.

662 **V14:** Lahar frequency is in decreasing interval order from high, week or months, to low: > 25  
663 years.

664 **V15:** Two PDC types are valley-confined pyroclastic flows affecting valley channels and  
665 banks, and un-confined, pyroclastic surges affecting valley terraces and sometimes adjacent  
666 interfluves. Because valley margins are populated, we ranked surge first, then confined PFs,  
667 and last, interfluves exceptionally affected > 0.5 km from valley channels.

668 **V16:** The PDC frequency range follows the same time interval order from high (1-8 years) to  
669 low > 25 years.

670 **V17:** Tephra fallout types are twofold: ballistics and lapilli within 8-9 km from the vent and 1  
671 for tephra-fall beyond the 8-9 km distance.

672 **V18:** Tephra-fall frequency is ranked in decreasing interval order from < 1 year to 5-25 years.

### 673 **5.2. Third group of variables: access and response**

674 The third group defines the access to blocks and the response to imminent eruption including  
675 evacuation (Table 2), ranked on distance and quality criteria. We attributed the maximum  
676 mark to the poorest quality or to the absent attribute of five variables, as the ultimate goal was  
677 to rank a range of blocks based on exposure and remoteness in case of crisis.



678 **V19:** Three types of access ways according to the decreasing quality of the network, from  
679 paved road, dirt road to trail.

680 **V20:** Evacuation roads, shelters and storage facilities include distance (5 km-threshold, i.e.,  
681 20 minutes driving a 4x4 vehicle or c. 1 hour walking) and absence from the *dusun*.

682 **V21:** Civil Protection works, such as dykes and check dams, are present or absent.

683 **V22:** Early warning system and local offices of the Indonesian board for natural disaster  
684 management (*Badan Nasional Penanggulangan Bencana* BNPB) are ranked according to  
685 distance (2.5 km-threshold, i.e., within walking distance) or absence.

686 **V23:** Health centres: four hospitals around Semeru are located in Dampit, Pronojiwo, Pasirian  
687 and Lumajang, but at least one small hospital or dispensary is present in all *Desa*. Distance  
688 (10 km- and 5 km-threshold) and presence or absence helped rank the health centres.

### 689 **5.3. Attributes of representative blocks based on univariate analysis**

690 Univariate analysis (UA, Fig. 7) was used to compute the frequency of attributes of all  
691 variables from the dataset including 145 *dusun* blocks.

692 First, the most frequent attributes allow us to elaborate on representative blocks (Fig. 7):

693 (1) As many as 84% of blocks are built and permanently occupied, 75% are located between 9  
694 and 12 km from the summit, and most of them are located on high (> 10 m) terraces at mid-  
695 distance (> 120 m) from the valley channel. (2) About 62% have recorded 0 fatalities, but  
696 most of them can be affected by light ashfall within 1 to 3 years interval; almost 27% of  
697 blocks can be damaged by lahars in valleys nearby, and 10% to 19% recorded PDC- and  
698 lahar-related fatalities, respectively. (3) The construction quality of 75% of homes is regular.  
699 (4) About 61% of blocks are close to a shelter or storage facility and c. 77% of blocks lie  
700 within 10 km of hospitals. However, this means that c. 23% of blocks lie  $\geq$  10 km from  
701 hospitals, while 51% remain poorly connected to other response facilities.

702 Second, UA results highlight the most relevant variable attributes among the blocks (Fig. 7):

703 (a) the population density of *dusun* and built blocks is discriminant, but higher densities of  
704 built blocks increase human exposure to volcanic threats; (b) the location of c. 24% of blocks  
705 on low to middle terraces and within short to middle distance to river channel may be more  
706 discriminant than the location of prior fatalities (as c. 62% of blocks record no death) and  
707 timely location of people (as homes in 83.50% of blocks are permanent).

708 Third, strong contrasts in exposure of inhabitants are due to the block location with respect to  
709 rivers due to lahar and PDC type, frequency, extent, and impacts. Blocks without reported  
710 fatalities prevail (c. 62%) in *dusun*, which long settled away from the active valley channels.  
711 Among PDCs, the predominant, un-confined surges reflect the correlation between fatalities  
712 and valley terraces on which few *dusun* have recently spread out. The impact of the daily  
713 tephra-fall hazard on blocks is binary: very few *dusun* lie within the 8-9 km radius from the  
714 vent affected by ballistics and lapilli from the recurrent, low-altitude columns at the summit,  
715 whereas decreasing tephra-fall frequency beyond 9 km has little impacts on *dusun* blocks if  
716 VEI remains low (< 3).

717 Fourth, hazard impacts depend on perpendicular or oblique orientation of buildings, which  
718 renders almost 63% of homes exposed to flows, while the poor (5.5%) and regular (c. 76%)  
719 quality of roofs would not withstand thick ( $\geq 25$  cm) and wet tephra-fall deposit (Fig. 4).  
720 About 51% of the blocks show poor access ways and most of them (c. 85%) lack civil  
721 protection works (dykes, dams) against volcanic flows, while check dams across active  
722 valleys remain filled up or damaged. In contrast, three variables of the third group may  
723 counterbalance eruption impacts, therefore reducing block exposure (Fig. 7): (a) Almost half  
724 of the ways are paved roads, favouring access to *dusun*; (b) evacuation shelters, roads signs  
725 and storage facilities do exist in c. 61% of blocks, and; (c) early warning system and BNPB  
726 offices are close enough (< 2.5 km) to c. 81% of blocks, while c. 77% of *dusun* lie within 10  
727 km of any hospital.

#### 728 **5.4. Block definition based on BA and MCA of attributes**

729 MCA was used to find statistical relationships between the exposure index score and other  
730 attributes that belong to hazards, access and response.

731 First, contributory variables of MCA were tested by means of Chi-square test (Table 5) to find  
732 out how dependent they are with other contributory variables to whom they are correlated.  
733 The Chi<sup>2</sup> test shows a strong statistical link (95% confidence) between variables indicated as  
734 dependent in Table 5 (bold figures) and ESD Table 3. There are six exposure variables: V2  
735 density of *dusun* and V3 density of built block population, V8 Terrace elevation/distance  
736 relationship, V9 Timely location of people, V10 House orientation, and V11 Roof and  
737 construction quality. Table 5 also include variables outside exposure that are dependent with  
738 the majority of variables: V13 Lahars, V15 PDCs, V16 PDC frequency, V18 tephra-fall  
739 frequency, V19 access roads, and V22 Early warning system and BNPB office.

740 Second, colour-coded MCA biplots show how the attributes contribute to the definition of  
741 blocks according to the Dimensions 1 and 2 (Fig. 8A) and Dimensions 2 and 3 (Fig. 8B).  
742 These graphs display proximities and oppositions between all attributes: the darker the colour,  
743 the more contributory the attribute is; conversely, light colour points to less contribution.  
744 MCA biplots (Figs. 8A, B) distinguish four groups of attributes based on contributions and  
745 squared cosines (ESD Table 6).

#### 746 **5.4.1. Four groups of attributes in the Dimensions 1 & 2 and 2 &3 biplots**

747 1. The right-hand side group along Dimension 1 include the following attributes: channel and  
748 low terrace/short distance (V8), both attributes DF or HCF (V13), 0 inhabitant density (V2,  
749 V3) together with the 'temporary location' attribute of V9. Such attributes point to un-built  
750 blocks located along the valley channels where shelters are temporary and people used to  
751 work daily and occasionally. These un-built blocks or temporary settlements, being not  
752 constantly exposed, show a residual to moderate EIPN. They recorded lahars and PDCs  
753 fatalities as they can be hit by both confined pyroclastic flows and unconfined surges, while  
754 their access by trail is challenging.

755 2. Two groups are positively and negatively correlated along Dimension 2 (Fig. 8A). The  
756 second group of attributes, positively correlated, define blocks with residual EIPN due to long  
757 distance and high elevation from the rivers (V8), absent lahars and PDCs (V13, V15),  
758 together with low frequency of tephra fall (V18). The relatively densely populated blocks are  
759 located within 12 and 20 km from vent, but they are devoid of evacuation shelter, early  
760 warning system, and civil protection office.

761 The third group is observed below the graph barycentre (Fig. 8A), but the quality of  
762 contribution is weaker than that of the previous group. Several attributes include: 9 to 12 km  
763 (V1), high block density (V3), PDC surge (V15) every 8 to 25 years (V16), ashfall every 1 to  
764 3 years (V18), and existing civil protection works (V21). These blocks, located in the most  
765 affected 9-12 km circle distance from the volcano summit, may be damaged by tephra fall and  
766 sometimes by lahars and PDCs, which characterize their moderate to very high EIPN.

767 Figure 8B shows the contribution of variable attributes in the Dimensions 2 and 3 biplot,  
768 although the contribution is relatively weak. The small fourth group of correlated,  
769 contributory attributes appears in the upper right corner of the biplot: distance < 9 km (V1),  
770 Lapilli and ash within 9 km from vent (V15), < 1 year (V16), and no health centre (V23).

771 Such attributes point to blocks located near the volcano summit (> 1,000 masl) liable to be  
772 affected by tephra fall and far from available emergency facilities.

773

## 774 **6. Statistical grouping of blocks; Application to relief operation in case of evacuation**

### 775 **6.1. Four groups of blocks result from MCA and HAC**

776 Attributes extracted from MCA (Fig. 8) and HAC (Table 6, ESD Fig. 4) point to four groups  
777 of blocks, ranked according to the most discriminant variables (see the Chi<sup>2</sup> test conducted on  
778 HAC variables, ESD Table 8) and the EIPN score. These groups are similar to those defined  
779 using MCA biplots.

780 Group 1 looks like the first group described in subsection 5.4.1: un-built blocks located along  
781 active valley channels and their margins are temporarily occupied by laborers or workers. The  
782 location explains why they are affected by all types of and frequent lahars and PDCs,  
783 inducing fatalities or injuries. Trails and dirt roads hinder access to these blocks, which  
784 remain blocked from any health centre. Examples are located on the margins of K. Bang and  
785 Kembar (*Dusun* Supit Timor), K. Koboan (*Dusun* Summersari), and K. Tengah (*Dusun*  
786 Tulungrejo) (Fig. 5A-C).

787 Group 2 is similar to the second group (see 5.4.1). Built blocks with residual EIPN are located  
788 farther away (12 to 20 km) from the volcano, on low slopes and in the ring plain. These  
789 blocks are occasionally affected by distal lahars along the active valleys and light, relatively  
790 un-frequent tephra-fall associated with long-reach PDCs from VEI<sub>≥</sub> 3 eruptions. Most of  
791 them are located away from the valley channels, but low-frequency surges may reach them  
792 during large, un-frequent eruptions. What makes the exposure residual is the average density  
793 of both blocks and *dusun*, and the lack of response facilities. Examples are Kajar Kuning,  
794 Gumuk Mas, Oro-Oro Ombo, and Jabon (Fig. 5B, and C).

795 Group 3 looks like the third group (see 5.4.1). Densely populated built blocks within highly  
796 populated *dusun*, located between 9 and 12 km distance, have long settled on high terraces  
797 away from the valley channels. Un-affected by lahars and valley-confined PDCs, these blocks  
798 can be impacted by relatively frequent tephra-fall and low-frequency surges. Such blocks  
799 exhibit moderate EIPN, when response facilities are close and they are protected by civil  
800 protection works (e.g., Oro-Oro Ombo and West of Summersari, Fig. 5B). Other blocks

801 exhibit high EIPN if they remain un-protected and remote from response facilities (Supit  
802 Timor, Rowobaung, [Fig. 5A](#); Curah Koboan, [Fig. 5B](#)).

803 Group 4 resembles the fourth group (see 5.4.1). Blocks within 9 km from Semeru's vent,  
804 frequently affected by light tephra-fall, are located higher up on the volcano's flanks, far from  
805 the main roads, response facilities such as hospitals. These blocks with poor house  
806 construction quality and small dispensaries exhibit very low or low EIPN, such as Blubuk and  
807 Karangsono ([Fig. 5D](#)), because the people density is low and they are affected by light ashfall  
808 only.

## 809 **6.2. Application to relief operations based on access to blocks and response facilities**

810 Our method to characterize blocks around Semeru was applied to rank them to highlight  
811 potential challenges in evacuation procedure. First, the distance/time criterion is computed to  
812 evaluate evacuation based on access to the blocks and means of transport. This criterion was  
813 related to distances to the shelters and/or storage facilities and health centres, as shown by  
814 [Table 7](#) and [ESD Table 9](#), [ESD Figure 5](#). Second, we used the results of analysis of attribute  
815 frequencies ([Table 7](#)) to distinguish a range of blocks based on access to response facilities.  
816 Factor map ([ESD Figure 5](#)) helps distinguish the number of clusters. Maps in [Figure 9](#) display  
817 a range of four block clusters based on access versus remoteness in case of evacuation. Blocks  
818 accessible only by dirt roads (e.g., Summersari, Curah Koboan along K. Koboan, [Fig. 9A](#);  
819 Tulungrejo along K. Tengah, [Fig. 9B](#); Supit and Supit Timor along K. Bang, [Fig. 9C](#)) are  
820 challenging for relief operations in case of evacuation. This was reflected by the cumbersome  
821 evacuation of injured and affected people following the 4 December 2021 eruption on both  
822 mid reaches of the Koboan valley. *Dusun* like Rowobaung ([Fig. 9C](#)) linked to Pronojiwo by  
823 one bridge would remain cut off by lahars propagating along two rivers surrounding the  
824 blocks. *Dusun* located high on the Semeru' slopes with narrow dirt roads (e.g., Blubuk and  
825 Karangsono, [Fig. 9D](#)) include upstream blocks at high elevation (> 1,000 m asl) connected  
826 only by trails that may become blocked by heavy tephra fall. The town of Senduro, far from  
827 active rivers and well connected by paved roads to Lumajang, can be evacuated, but the  
828 narrow street network in densely populated *dusun* may be challenging ([Fig. 9E](#)).

## 829 **7. DISCUSSION**

830 Table 8 compares the number of exposure parameters from the literature with our local-scale,  
831 multi-component EIPN method. Specific characteristics of exposure arise from this  
832 comparison.

833 Globally, two criteria measure human exposure index using the distribution of population  
834 potentially affected by a volcanic eruption: either the (log) number or the density per area  
835 (e.g., Wild et al., 2021; Nieto-Torres et al., 2021). The majority of publications, following  
836 Yokoyama et al. (1984) and Ewert (2007), consider the number or density of population  
837 within circle distance or radii from the main crater. Recent studies correlated the circle-  
838 distance thresholds with fatalities (Auker et al., 2013; Brown et al., 2017) and population  
839 density (Freire et al., 2019). A limited number of studies encompass several parameters of  
840 human exposure (e.g., five among 10 parameters describing exposure on the Nisyros Island;  
841 Kinvig et al., 2016). The most complete study to date in ranking volcanic risk (Nieto-Torres et  
842 al., 2021) includes nine exposure parameters within a total of 41 covering hazard, exposure,  
843 vulnerability and resilience. However, their single human exposure parameter is the density of  
844 population within 5, 10, 30 and 100 km radii.

845 In contrast, the multi-component EIPN is based on six optimal variables using PLR models  
846 from a set of 11 initial exposure variables. Instead of a simple statistical technique to weigh  
847 and sum exposure parameters, statistical operations enabled us to convert the qualitative  
848 variables in semi-quantitative criteria and to elaborate on normalization of attribute values and  
849 discretization of EIPN scores. Then polytomous logistic regression models quantified the link  
850 between the probability of a EIPN score and the variables of exposure. For research  
851 efficiency, future surveys should collect as few optimal variables as possible to obtain the  
852 EIPN score for blocks potentially around any populated, active volcano.

853 In the literature, the global-scale 'PEI' underestimates local factors that induce spatial and  
854 temporal exposure patterns. In contrast, our work highlights local factors that combine on a  
855 multi-component human exposure. Local factors that govern the extent to which settlements  
856 are exposed to specific threats involve: (1) topography and geomorphological location near  
857 valleys that convey most of the flows; (2) fatalities location related to hazard type and  
858 frequency, which affect human habitat and livelihoods, and; (3) home and roof quality that  
859 may increase exposure of residents. We consider exposure related to volcanic flows (PDCs,  
860 lahars) inasmuch these hazards induce impacts on settlements located near active river  
861 channels and adjacent low banks. We thus accounted for shallow and sinuous valley channels,

862 which favour overbank pyroclastic surges, overbank lahars and associated floods along the  
863 most active rivers (Fig. 3). We do not discard outside hazard-prone areas affected in the recent  
864 past by lahars and associated floods produced by overbank and avulsion into secondary  
865 drainage. On the other hand, we also considered local factors that may contribute to decrease  
866 exposure, e.g., easy access on paved roads to the *dusun*, short distance/time ( $\leq 5$  km and  $\leq 25$   
867 minutes by car depending on the quality of roads) to health centres, emergency facilities, and  
868 finally the existence of civil protection works.

869 Measurements of exposure depends on the time frame at which assessments are conducted  
870 (Auker et al., 2015). We examined exposure not only in primary residence (Wild et al., 2021),  
871 but also in working areas: schools, shops, farmland and valleys, as many men extract  
872 construction material from lahar and PDC deposits in active river channels. Here, the  
873 ‘counting record’ of events and impacts held by the Dutch and Indonesian Volcanological  
874 Survey since 1884 ensures completeness of the Semeru daily explosive activity and its  
875 multiple chronic spurts. Monitoring from the Semeru Observatory at Gunung Sawur, 12 km  
876 SE of the summit (Fig. 3) contributed to the eruption record as early as 1953.

877 Few studies address exposure of highly populated communities on a persistently active  
878 volcano. Instead of sporadic (e.g., Mayon) or chronic eruptive activity (e.g., Merapi),  
879 Semeru’s acute and constant, daily explosive activity since at least 1967—the longest daily  
880 explosive unrest worldwide with the exception of Sakurajima—exerts a heavy, constant toll on  
881 human life, thus on perception and adaptation of communities exposed to volcanic risk.

## 882 8. CONCLUSION

883 The local-scale method, including field data collection, mapping and a range of statistical  
884 techniques, helped compute four levels of a multi-component index of exposure applied to  
885 145 neighbourhoods mapped in 13 *dusun* and two small towns.

886 Polytomous logistic regression models allowed us to select six optimal variables and predict  
887 the EIPN score of blocks. These optimal variables are reproducible parameters to assess  
888 human exposure on active, populated volcanoes.

889 Computing and mapping human exposure at the block scale may be more adapted to: 1) the  
890 characteristics of the population, (2) the diversity of hazards and timely change in exposure to  
891 a persistently explosive activity, and (3) the mixed rural-small urban communities with a  
892 variety of resources, which support livelihoods and sustain the community coping capacity.

893 Ranking blocks using distance-timing and access to response facilities is a useful tool to point  
894 blocks that need relief operations to be implemented. Civil authorities may provide advice and  
895 funds to retrofit home construction, relocate exposed homes that encroached on low terraces,  
896 clean up check dams, pave access ways to, and implement dispensaries in remote villages.  
897 Results should help disaster risk management staff to improve their participation at the scale  
898 of neighbourhoods on active volcanoes.

## 899 **Acknowledgments**

900 Fieldwork and laboratory analyses were funded by the ANR 'RiskAdapt' research project.  
901 This research was also financed by the French government IDEX-ISITE initiative 16-IDEX-  
902 0001 (CAP 20-25). The authors are grateful to DIKTI (Directorate General of Higher  
903 Education, Ministry of National Education of Indonesia), who bestowed two research permits  
904 to the first author. We acknowledge the technical and scientific support from Dr. A.-F. Yao  
905 Lafourcade (Laboratory of Mathematics, UCA), University Gadjah Mada, Yogyakarta (Isna  
906 Pujiastuti) and University IPB, Bogor (Muhammed Syaif Habi, F. Muhammed A.W. Hasan).  
907 We thank Mr. Mahjum and Pak Sam for their logistical support in field.

## 908 **Declarations**

909 *Authors' contribution statements.* All authors contributed to the study conception and design.  
910 Material preparation, data collection and analysis were performed by JC Thouret, M  
911 Taillandier, E Wavelet, N Azzaoui, and B Tjajhono. The first draft of the manuscript was  
912 written by JC Thouret and M Taillandier and all authors commented on previous versions of  
913 the manuscript. All authors read and approved the final manuscript. Artwork was performed  
914 by M Taillandier, E. Wavelet, JC Thouret and O Santoni.

915 *Ethics approval.* The research did not involve Human participants and/or Animals

916 *Funding and/or Conflict of interests/Competing interests.* Funding (no specific grant received)  
917 has been cited in the Acknowledgments. The authors have no competing interests to declare  
918 that are relevant to the content of this article. All authors certify that they have no affiliations  
919 with or involvement in any organization or entity with any financial interest or non-financial  
920 interest in the subject matter or materials discussed in this manuscript.

## 921 **REFERENCES**

- 922 Abdi, H., Valentin, D., 2007. Multiple Correspondence Analysis, *in*: Salkind, N. (Ed.)  
923 Encyclopedia of Measurement and Statistics. Thousand Oaks (CA), Sage.
- 924 Akaike, H., 1987. Factor analysis and AIC. *Psychometrika*, 52, 317-332.



925 Aspinall, W., Blong, J., 2015. Volcanic risk assessment. Chapter 70, pp. 1215-1231, *in*:  
926 Sigurdsson, H. et al., Encyclopedia of Volcanoes, 2nd edition, Academic Press.

927 Aspinall, W., Auker, M., Hincks, T., Mahony, S., Nadim, F., Pooley, J., Syre, E., 2011.  
928 Volcano hazard and exposure in GFDRR priority countries and risk mitigation measures-  
929 GFDRR Volcano Risk Study. Bristol University Cabot Institute and NGI Norway for the  
930 World Bank, NGI Report, 20100806, 3.

931 Auker, M.R., Sparks, R.S.J, Siebert, L., Crossweller, H.S., Hewert, J., 2013. A statistical  
932 analysis of the global historical volcanic fatalities record. *J. Appl. Volcanol.* 2, 2, 1-24.

933 Auker, M.R., Sparks, R.S.J., Jenkins, S.F., Aspinall, W., Brown, S.K., Deligne, N.I., Jolly, G.,  
934 Loughlin, S.C., Marzocchi, W., Newhall, C.G., Palma, J.L., 2015. Development of a new  
935 global Volcanic Hazard Index (VHI), pp. 349-357, *in*: Loughlin, S. C. et al. (Eds.), *Global*  
936 *Volcanic Hazards and Risk*. Cambridge: Cambridge Univ. Press.

937 Bakkour, D., Kast, R., Enjolras, G., Thouret, J.-C., 2015. The adaptive governance of natural  
938 disasters: Insights from the 2010 Mount Merapi Eruption in Indonesia. *Int. J. Dis. Risk Red.*  
939 13, 167-188, doi:10.1016/j.ijdr.2015.05.006

940 Benzécri, J.-P., 1979. Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. *Les*  
941 *cahiers de l'analyse des données* 4, 3, 377-378

942 BPS Badan Pusat Statistik, Indonesia, 2017. *Tinjauan Regional Berdasarkan PDRB*  
943 *Kabupaten/Kota 2015-2019; Buku 2: Jawa and Bali*. Jakarta, 169 pp.

944 Bronto, S, Hamidi, S, Martono A., 1996. Disaster-prone zone map of Semeru volcano, East  
945 Java (1:50,000 scale, colour). Direktorat Vulkanologi, Volc Survey Indonesia, Bandung.

946 Brown, S. K., Auker, M. R., Sparks, R. S. J., 2015a. Populations around Holocene volcanoes  
947 and development of a Population Exposure Index, pp. 223-232. *In*: Loughlin, S. et al. (Eds.),  
948 *Global Volcanic Hazards and Risk*. Cambridge: Cambridge Univ. Press.

949 Brown, S.K., Loughlin, S.C., Sparks, R.S.J., Vye-Brown, C., Barclay, J., Calder, E., Cottrell,  
950 E., Jolly, G., Komorowski, J.-C., Mandeville, C., Newhall, C., Palma, J., Potter, S., Valentine,  
951 G., 2015b. Global volcanic hazard and risk, pp. 81-172, *in*: Loughlin, S. C. et al. (Eds.),  
952 *Global Volcanic Hazards and Risk*. Cambridge: Cambridge Univ. Press.

953 Brown, S.K., Jenkins, S.F., Sparks, R.S.J., Odbert, H., Auker, M.R., 2017. Volcanic fatalities  
954 database: Analysis of volcanic threat with distance and victim classification. *J. Appl.*  
955 *Volcanol.* 6, 15.

956 Chambers, J. M., Cleveland, W. S., Kleiner, B., Tukey, P. A., 2018. Graphical methods for  
957 data analysis. Chapman and Hall/CRC.

958 Del Negro, C., Cappello, A., Bilotta, G., Ganci, G., Hérault, A., Zago, V., 2019. Living at the  
959 edge of an active volcano: Risk from lava flows on Mt. Etna. *Geol. Soc. Amer. Bull.* 132, 7-8,  
960 1615–1625. <https://doi-org.insu.bib.cnrs.fr/10.1130/B35290.1>

961 Donovan, A, Ayala, I.A, Eiser, J, Sparks, R.S.J., 2018. Risk perception at a persistently active  
962 volcano: warnings and trust at Popocatépetl volcano in Mexico, 2012–2014. *Bull. Volcanol.*  
963 80, 5, 47.

964 Donovan, K., 2010. Doing social volcanology: exploring volcanic culture in Indonesia.  
965 *Area* 42, 1, 117-126.

966 Doocy S, Daniels A, Dooling S, Gorokhovich Y., 2013. The Human Impact of Volcanoes: a  
967 Historical Review of Events 1900-2009 and Systematic Literature Review. *PLOS Currents*  
968 *Disasters.* 1. doi: 10.1371/currents.dis.841859091a706efebf8a30f4ed7a1901

969 Ewert, J.W., Harpel, G.J., 2004. In Harm’s Way: Population and Volcanic Risk. *Geotimes* 49,  
970 14-17.

971 Ewert, J.W., 2007. System for ranking relative threats of U.S. volcanoes. *Nat. Haz. Rev.* 8,  
972 112-124.

973 Ewert, J.W., Diefenbach, A.K., Ramsey, D.W., 2018. 2018 update to the U.S. Geological  
974 Survey national volcanic threat assessment, U.S. Geol. Surv. Sci. Invest. Rep. 2018-5140, 40  
975 pp., [doi.org/10.3133/sir20185140](https://doi.org/10.3133/sir20185140).

976 First IAVCEI-GVM Workshop, 2018: “From Volcanic Hazard to Risk Assessment”,  
977 Consensual document, 40 p. by Bonadonna, C., Biass, S., Calder, E., Frischknecht, C., Gregg,  
978 C., Jenkins, S., Loughlin, S., Menoni, S., Takarada, S., and Wilson, T. Geneva, Switzerland,  
979 27-29 June 2018, <https://vhub.org/resources/4498>.

- 980 Freire, S., Florczyk, A.J., Pesaresi, M., Sliuzas, R., 2019. An improved global analysis of  
981 population distribution in proximity to active volcanoes, 1975–2015. *ISPRS Inter. J. Geo-*  
982 *Infor. (MDPI)* 8, 341; doi:10.3390/ijgi8080341
- 983 Gaillard, J.C., 2008. Alternative paradigms of volcanic risk perception: the case of Mt  
984 Pinatubo in the Philippines, *J. Volcanol. Geoth. Res.* 172 (2008) 315–328.
- 985 Gaillard, J.C., Dibben, C.J.L., 2008. Volcanic risk perception and beyond, *J. Volcanol. Geoth.*  
986 *Res.* 172 163–169.
- 987 GFDRR, Global Facility for Disaster Reduction and Recovery of the World Bank (Aspinall et  
988 al.), 2011. *Volcano Risk Study. Volcano hazard and exposure in GFDRR countries and risk*  
989 *mitigation measures. NGI report 20100806, GFDRR, University of Bristol. 309 pp.*
- 990 GVP Global Volcanism Program, 2022. Report on Semeru (Indonesia) by Crafford, A.E.,  
991 Venzke, E, (Eds). “Pyroclastic flows from dome collapse on 4 December 2021 destroyed  
992 multiple communities and caused fatalities”. *Bulletin of the Global Volcanism Network*, 47,  
993 1, Smithsonian Institution.
- 994 Jenkins, S.F., Spence, R.J.S., Fonseca, J., Solidum, R.U., Wilson, T.M., 2014. Volcanic risk  
995 assessment: Quantifying physical vulnerability in the built environment. *J. Volc. Geoth. Res.*  
996 276, 105-120.
- 997 Jenkins, SF, Wilson, TM, Magill, CR, Miller, V, Stewart, C, Marzocchi, W, Boulton, M.,  
998 2015. Volcanic ashfall hazard and risk: technical background paper for the UNISDR 2015  
999 global assessment report on disaster risk reduction. A report by Global Volcano Model and  
1000 IAVCEI, 43 p., Commonwealth of Australia.
- 1001 Jenks, G.F. 1967. The data model concept in statistical mapping. *International Yearbook of*  
1002 *Cartography* 7, 186–190.
- 1003 Jiménez, D., Becerril, L., Carballo, A., Baires, S., Martí, J., 2019. Estimating exposure around  
1004 San Miguel Volcano, El Salvador. *J. Volcanol. Geoth. Res.* 106675, 9.
- 1005 Jóhannesdóttir, G., Gísladóttir, G., 2010. People living under threat of volcanic hazard in  
1006 southern Iceland: vulnerability and risk perception. *Nat. Haz. Earth Syst. Sci.* 10, 407–420.

1007 Jumadi, Heppenstall, A.J., Malleson, N.S., Carver, S.J., Quincey, D.J., Manville, V.R., 2018.  
1008 Modelling individual evacuation decisions during natural disasters: A case study of volcanic  
1009 crisis in Merapi, Indonesia. *Geosciences MDPI* 8, 196, 30.

1010 Kinvig, H.S., Winson, A., Gottsmann, J., 2010. Analysis of volcanic threat from Nisyros  
1011 Island, Greece, with implications for aviation and population exposure. *Nat. Haz. Earth Sys.*  
1012 *Sci.* 10, 1101–1113.

1013 Kleinbaum, D.G., Klein, M., 2010. Polytomous logistic regression, pp. 429-462, *in:*  
1014 Kleinbaum, D.G., Klein, M., *Logistic regression, A self-learning text, Statistics in Biology*  
1015 *and Health*, 3rd edition, Springer

1016 Lavigne, F., De Coster, B., Juvin, N., Flohic, F., Gaillard, J.-C., Texier, P., Morin, J.,  
1017 Sartohadi, J., 2008. People's behavior in face of volcanic hazards: Perspectives from Javanese  
1018 communities, Indonesia. *J. Volc. Geoth. Res.* 172, 273-282.

1019 Lechner, H.N., Rouleau, M.D., 2019. Should we stay or should we go now? Factors affecting  
1020 evacuation decisions at Pacaya volcano, Guatemala. *Int. J. Dis. Risk Red.* 40, 101160.

1021 Lerner-Lam, A., 2007. Assessing global exposure to natural hazards: Progress and future  
1022 trends. *Environ. Hazards* 7, 10–19.

1023 Loughlin, S.C., Sparks, S., Brown, S.K., Jenkins, S.F., Vye-Brown, C. (Eds.), 2015. *Global*  
1024 *Volcanic Hazards and Risk*, Cambridge Univ. Press, Cambridge, 391 pp.

1025 Mangan, M., Ball, J., Wood, N., Jones, J.L., Peters, J., Abdollahian, N., Dinitz, L.,  
1026 Blankenheim, S., Fenton, J., Pridmore, C., USGS, 2018. California's exposure to volcanic  
1027 hazards. *Scient. Investig. Report 2018-5159*, 44 pp + 3 Appendices.

1028 Malin, M.C., Sheridan, M.F., 1982. Computer-assisted mapping of pyroclastic surges. *Science*  
1029 217, 4560, 637-640, doi:10.1126.science.217.4560.637

1030 Michellier, C., Kervyn, M., Barette, F., Muhindo Syavulisembo, A., Kimanuka, C., Kulimushi  
1031 Mataboro, S., Hage, F., Wolff, E., Kervyn, F., 2020. Evaluating population vulnerability to  
1032 volcanic risk in a data scarcity context: The case of Goma city, Virunga volcanic province  
1033 (DR Congo). *Int. J. Dis. Risk Red.* 45, 101460.

1034 Nakada, S., Maeno, F., Yoshimoto, M., Hokanishi, N., Shimano, T., Zaennudin, A., Iguchi,  
1035 M., 2019. Eruption scenarios of active volcanoes in Indonesia. *J. Disas. Res.* 14, 1, 40-50.

1036 Nieto-Torres, A., Freitas Guimarães, L., Bonadonna, C., Frischknecht, C., 2021. A new  
1037 inclusive volcanic risk ranking, Part 1: Methodology. *Frontiers in Earth Science*, doi.org /  
1038 10.3389/feart.2021.697451

1039 Paton, D., Smith, L., Daly, M., Johnston, D., 2008. Risk perception and volcanic hazard  
1040 mitigation: individual and social perspectives, *J. Volcanol. Geoth. Res.* 172, 179–188.

1041 Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete  
1042 samples). *Biometrika*, 52, 3-4, 591–611, Doi:10.1093/biomet/52.3-4.591

1043 Siagian, T.H., Purhadi, P., Suhartono, S., Ritonga, H., 2013. Social vulnerability to natural  
1044 hazards in Indonesia: driving factors and policy implications. *Nat Haz.* 70, 2, 1603-1617. Doi:  
1045 10.1007/s11069-013-0888-3

1046 Solikhin A., Thouret J.-C., Harris A., Liew S.C., Gupta A., 2012. Geology, tectonics, and the  
1047 2002-2003 eruption of Semeru volcano, Indonesia: interpreted from high-spatial resolution  
1048 satellite imagery. *Geomorph.* 138, 364-372. Doi:10.1016/j.geomorph.2011.10.001.

1049 The Hazards and Vulnerability Research Institute HVRI at the University of Southern  
1050 Carolina USC, <https://start.umd.edu/data-tools/social-vulnerability-index-sovi>. Accessed 22  
1051 July 2022.

1052 Thouret, J. C., Lavigne, F., Suwa, H., Sukatja, B., 2007. Volcanic hazards at Mount Semeru,  
1053 East Java (Indonesia), with emphasis on lahars. *Bull. Volcanol.* 70, 2, 221-244.

1054 Thouret, J.-C., Ettinger, S., Guitton, M., Santoni, O., Magill, C., Martelli, K., Zuccaro, G.,  
1055 Revilla, V., Charca, J.A., Arguedas, A., 2014a. Assessing physical vulnerability in large cities  
1056 exposed to flash floods and debris flows: the case of Arequipa (Peru). *Nat. Haz.* 73, 3, 1771-  
1057 1815. Doi: 10.1007/s11069-014-1172-x.

1058 Thouret, J. C., Oehler, J. F., Gupta, A., Solikhin, A., Procter, J. N., 2014b. Erosion and  
1059 aggradation on persistently active volcanoes—a case study from Semeru Volcano,  
1060 Indonesia. *Bull. Volcanol.* 76, 10, 857.

1061 Thouret, J.-C., Wavelet, E., Taillandier, M., Tjahjono, B., Jenkins, S., Azzaoui, N., Santoni,  
1062 O., 2022. Defining population socio-economic characteristics and adaptive capacity of  
1063 communities to persistent volcanic threats from Semeru, Indonesia. *Int. J. Dis. Risk Reduction*  
1064 103064, [https://doi.org/ 10.1016/j.ijdr.2022.103064](https://doi.org/10.1016/j.ijdr.2022.103064).

1065 UNDP, United Nations Development Programme, 2020. Human Development Report 2020  
1066 The Next Frontier: Human Development and the Anthropocene, pp. 343–350. ISBN 978-92-  
1067 1-126442-5.

1068 UNISDR 2 February 2017. Terminology on Disaster Risk Reduction. Basic definitions on  
1069 disaster risk reduction to promote a common understanding on the subject for use by the  
1070 public, authorities and practitioners, [https://www.preventionweb.net/files/50683](https://www.preventionweb.net/files/50683_oiewgreportenglish.pdf)  
1071 [oiewgreportenglish.pdf](https://www.preventionweb.net/files/50683_oiewgreportenglish.pdf)

1072 UNDRR United Nations Office for Disaster Risk Reduction, 2017. Terminology.  
1073 [we/inform/terminology](https://www.preventionweb.net/files/50683_oiewgreportenglish.pdf).

1074 Wilcoxon, F., 1945. Individual comparisons by ranking methods. *Biometrics*  
1075 *Bulletin*, 1, 6, 80–8, doi: 10.2307/3001968.

1076 Wild, A.J., Bebbington, M.S., Lindsay, J.M., Charlton, D.H., 2021. Modelling spatial  
1077 population exposure and evacuation clearance time for the Auckland Volcanic Field, New  
1078 Zealand. *J Volcanol Geoth Res* 416, 107282

1079 Wilson, G., Wilson, T.M., Deligne, N.J., Cole, J.V., 2014. Volcanic hazard impacts to critical  
1080 infrastructure: A review. *J. Volcanol. Geoth. Res.* 286, 148-182.

1081 Wisner, B, Blaikie, P, Cannon, T, Davis, I., 2004. *At Risk: Natural hazards, people's*  
1082 *Vulnerability and Disasters*, 2nd edition, Routledge, London, 284 pp.

1083 Woo, G., 2015. Cost–Benefit Analysis in Volcanic Risk, Chapter 11, pp. 289-300, *In: Papale,*  
1084 *P. (Ed.), Volcanic Hazards, Risks and Disasters*, Elsevier.

1085 Wood, N., Soulard, C., 2009. Variations in population exposure and sensitivity to lahar  
1086 hazards from Mount Rainier, Washington. *J. Volcanol. Geoth. Res.* 188, 4, 367-378.

1087 Yokoyama, I., Tilling, R., Scarpa, R., 1984. International mobile Early-Warning Systems for  
1088 Volcanic Eruptions and Related Seismic Activities. FP/ 2106-82-01 (2286). Paris: UNESCO.

1089 Zuccaro, G., De Gregorio, D., Baxter, P., 2015. Human and structural vulnerability to  
1090 volcanic processes, Chapter 10, pp. 261-288, *In: Papale, P. (Ed.), Volcanic Hazards, Risks*  
1091 *and Disasters*, Elsevier.

1092

1093 **TABLE CAPTIONS**

1094 **Table 1.** Methods, techniques and objectives of the statistical analyses (UA, BA, FCA, MCA  
1095 and HAC) conducted on variables and attributes of the Exposure Index of Populated  
1096 Neighbourhoods (EIPN). Observations are neighbourhoods termed blocks. Polytomous  
1097 Logistic regression (PLR) is a model to extract most pertinent variables for predicting EIPN  
1098 scores of blocks, which can be applied outside the survey area and on any populated, active  
1099 volcano. HAC is a clustering technique to distinguish groups of variable attributes that  
1100 describe a range of exposed blocks.

1101 **Table 2.** Variables and their attributes (with thresholds) used to compute the EIPN at the local  
1102 *dusun* (sub-village) and block (neighbourhood) scale. **A.** Eleven variables define what can  
1103 increase population exposure. **B.** Seven variables linked to hazard types, occurrence and  
1104 frequency. **C.** Five variables describe accessibility to blocks and the existence/distance of  
1105 response facilities in case of imminent eruption and evacuation. Ordinal numbers (without  
1106 weighting) were attributed as thresholds of attributes within each variable, but different values  
1107 for each attribute were normalized for computing the EIPN, while the index score intervals  
1108 were discretized using the Jenks method. We calculated weighting coefficients associated to  
1109 the variable no.4 using the method of intersection of probabilities adapted from [Brown et al.](#)  
1110 (2015), see [Table 3](#).

1111 **Table 3.** Four steps describing the method for computing the exposure index adapted from the  
1112 People Exposure Index (PEI: [Brown et al., 2015a](#)) according to four circle distances (9, 12,  
1113 18, and 35 km) around Semeru, taking into account the recorded volcanic events, village  
1114 population, and reported fatalities since 1884.

1115 **Table 4. A.** PLR initial model with 8 variables and their attributes with coefficients of  
1116 explanatory variables, significativity and thresholds of EIPN coefficients. **B.** PLR selected  
1117 model showing the 13 attributes of 6 variables with their estimated coefficients and  
1118 ‘significativity’. We selected the most significant variables (3 stars), meaning < 0.1% error in  
1119 probabilities, although a less strict selection would also retain 2-stars attributes, i.e., with <  
1120 1% error in probabilities.

1121 **Table 5.** Results of the Chi-square test conducted on the three groups of variables. The dashed  
1122 black line indicates the results for the 11 exposure variables. Bold P-values indicate variables  
1123 that are independent (no statistical relationship) because p-value exceeds 5% based on the fact

1124 that a few sample cohorts do not exceed 5 observations (ESD Table 3). P-value, a probability  
1125 number between 0 and 1, is defined as the probability of getting a result that is either the same  
1126 or more extreme than the actual observations. Almost all p-values are significant for a  
1127 threshold of 5%. Hence, we pose two hypotheses: two variables are independent if p-value  
1128 exceeds 5% (i.e., the null hypothesis  $H_0$ ), against the alternative  $H_1$ , i.e., two variables are  
1129 dependent if p-value is  $< 5\%$ . While taking account of the size of the table of contingency  
1130 (number of freedom degree), the Chi-square test computes the deviation between observed  
1131 and theoretical counts (i.e., those expected if two variables were independent). Upon  
1132 assessing this deviation, the hypothesis of independence is accepted or rejected.

1133 **Table 6.** Results from Hierarchical Analytical Clustering (HAC) showing four clusters or  
1134 groups of blocks based on the frequency of attributes, as shown by the outputs of MCA. See  
1135 Factor map, ESD Figure 4 that helps support the distinction between four clusters of blocks.

1136 **Table 7.** Four clusters of at-risk blocks in *dusun* based on the distance/timing criterion, access  
1137 to blocks and response facilities. See Factor map, ESD Figure 5 and ESD Table 9 that help  
1138 support the distinction between four clusters of blocks in case of evacuation.

1139 **Table 8.** Comparison of parameters used in the Exposure Index of Populated Neighbourhoods  
1140 at Semeru with previous methods and parameters defining exposure within risk studies from  
1141 the literature.

## 1142 **FIGURE CAPTIONS**

1143 **Figure 1.** Flow chart showing four research steps and how we defined, computed and mapped  
1144 an Exposure Index of Populated Neighbourhoods (EIPN) at the local scale of sub-village  
1145 (*dusun*) blocks (neighbourhoods). The fourth research step is an application of the local scale  
1146 method to identify remote and/or blocked blocks for imminent eruption including evacuation.

1147 **Figure 2. A.** Map of Semeru volcano and ring plain in the middle of the regencies of  
1148 Lumajang to the east and Malang to the west. Distance circles 9, 12, 18 and 35 km (dashed  
1149 white lines) with respect to the persistently active vent. Principal cities and towns are  
1150 indicated. Red circle in map **B** points to the Semeru-Tengger massif, East Java. M= Malang,  
1151 S= Surabaya.

1152 **Figure 3.** Hazard-zone map of the Semeru's slopes and ring plain depicting the extent of  
1153 volcanic phenomena in the case of a medium-sized (VEI 3) eruption, based on the 1994, 1995



1154 and 2002 eruptions and post-eruption lahars (Thouret et al., 2007). The map also shows the  
1155 extent of lahars and floods in the case of a catastrophic eruption ( $VEI > 3$ ) along valleys  
1156 through the distal south and east ring plain, based on the 1909 and 1981 events and the  
1157 disaster-prone areas map (Bronto et al., VSI, 1996). Initials indicate the surveyed 13 *dusun*:  
1158 B-K Blubuk, Karanguko, SU Supit, RB Rowobaung, SB Sumbersari, GM Gumuk Mas, CK  
1159 Curah Koboan, KK KajarKuning, TU Tulungrejo, JA Jabon, SM Sumbermulyo, JL  
1160 Jaranglangak, RE Rekesan, and two towns of Pronojiwo and Senduro.

1161 **Figure 4.** Results of univariate analysis (given in percentage) conducted on 43 attributes of 11  
1162 EIPN variables applied to all studied *dusun* blocks. The grayscale block attributes, from the  
1163 lowest exposure index (light grey) to the highest exposure index (dark grey), are similar for  
1164 all plots.

1165 **Figure 5.** Maps showing the EIPN score at the scale of each *dusun* block. **A.** Town of  
1166 Pronojiwo, *Dusun* Supit-Supit Timur, and Rowobaung. **B.** *Dusun* Sumbersari, Gumuk Mas,  
1167 Curah Lengkong, and Kajar Kuning on both sides of K. Koboan. **C.** *Dusun* Tulungrejo,  
1168 Tawonsongo and Jabon in Desa Pasarjumble. **D.** *Dusun* Blubuk and Karanguko. **E.** *Dusun*  
1169 Sumbermulyo, Juranglangak, and Rekesan in *Desa* Senduro. The four colour-coded final  
1170 scores of the EIPN (obtained from ESD Table 6) show how one of the exposure index score  
1171 levels were assigned to every block.

1172 **Figure 6.** Scatter plots derived from MCA results showing the most opposed attributes (i.e.,  
1173 away from the plot barycentre) of 11 contributory variables for exposure along the  
1174 Dimensions 1 and 2 that convey 76% of the information.

1175 **Figure 7.** Results of univariate analysis (given in percentage) conducted on all attributes of  
1176 the second group and third group of variables applied to all studied *dusun* blocks. The seven  
1177 hazard variables are V12 to V18. The five variables V19 to V23 describe access to blocks and  
1178 the existence and distance of response facilities.

1179 **Figure 8.** Results from Multiple Component Analysis (MCA) on variables and their attributes  
1180 correlated to the EIPN score. **A.** Plot showing contribution of variables along Dimensions 1  
1181 and 2 (76% of the information). **B.** Plot showing contribution of variables along Dimensions 2  
1182 and 3 (23% of the information). Percentages of information are shown in ESD Figure 3.

1183 **Figure 9.** Maps of four clusters of blocks in *dusun* shown in Figure 5 based on  
1184 distance/timing criterion (including the quality of access and means of transport) to the

1185 response facilities (shelters, storage facility, early warning system, BNPB offices and health  
1186 centres). The four block clusters stem from the high/low frequency analysis (Table 7) and  
1187 HAC Factor map (ESD Figure 5). **A.** Green: Easy access to blocks (10-25 minutes) and  
1188 short distance response facilities (2.5–5 km). **B.** Yellow: Slow access to blocks (25 to 45  
1189 minutes) and short- and mid-distance response facilities (2.5–10 km). **C.** Orange:  
1190 Challenging access to blocks (45 to 170 minutes) and long-distance (> 10 km) response  
1191 facilities. **D.** Red: Delayed access to blocked blocks (80 to 170 minutes) and remote or absent  
1192 response facilities.

### 1193 **Electronic Supplement Data ESD**

1194 ESD Figure 1. Maps showing the setting of the 15 *dusun* together with the blocks in which we  
1195 conducted field survey and statistical analyses on exposure parameters. A. Karangsono and  
1196 Blubuk (Desa Tamansatryan), west flank. B. Town of Pronojiwo, Supit-Supit Timur and  
1197 Rowobaung (Desa Pronojiwo), South flank. C. Oro-Oro Ombo (Desa and *dusun*), Sumbersari,  
1198 Gumuk Mas, Curah Lengkong (Desa Supit Urang), SSE and SE flank. D. Kajar Kuning (Desa  
1199 Sumberwuluh) and Desa Candipuro, SSE flank. E. Tulungrejo and Jabon (Desa Pasrujambe),  
1200 ESE flank, and F. Sumbermulyo, Juranglangak, and Rekesan (Desa Senduro), East flank.

1201 ESD Figure 2. Scree model with distribution of information according to dimensions.

1202 ESD Figure 3. Factor map obtained from HAC showing four groups of blocks based on  
1203 attribute frequencies: see Table 6 for the list of high and low attribute frequencies.

1204 ESD Figure 4. Factor map obtained from distance/timing criteria and HAC (Table 7) and  
1205 showing four clusters of blocks according to access and response facilities.

1206 ESD Table 1. Setting of surveys carried out in *dusun* (sub-villages): administrative units,  
1207 location, surface area, people density, and number of surveys in each *dusun*. Symbol meaning:  
1208 \* data from BPS reports, *Kecamatan Dalam Angka* 2019, and 2018 for Tamansatryan,  
1209 Sumberwuluh, Candipuro. \*\*A *dusun* usually includes 4 to 5 *RukunWarga* (RW, a  
1210 neighbourhood with 50 to 75 houses). A RW includes usually 3 to 9 *RukunTetanga* (RT, a  
1211 block with 20 to 25 houses). Field survey was carried out at the scale of RWs, including more  
1212 than one observation per RT.

1213 ESD Table 2. Coordinates of buildings, economic status of respondents, and geographical  
1214 exposure with respect to active valleys.

1215 ESD Table 3. Chi-square test on the set of 23 variables to determine whether two variables are  
1216 independent or dependent. In this case, a variable is independent if the p-value exceeds 5%  
1217 (see [Table 5](#)). As a result, a statistical link exists (95% confidence) between variables  
1218 indicated as dependent with corresponding variables listed in the first column. Dark grey  
1219 indicates variables of exposure, grey variables of hazards, and white variables of access and  
1220 response.

1221 ESD Table 4. Burt Table of contingency (all attributes are considered) showing statistical  
1222 links between attributes of two variables at a time.

1223 ESD Table 5. Coordinates, squared cosine, and contribution of attributes used in MCA  
1224 biplots.

1225 ESD Table 6. Master Table of computed EIPN per *dusun* blocks, totalling 145 (horizontal  
1226 rows; two initials indicate the *dusun* name) according to all exposure variables and their  
1227 attributes (vertical rows, see [Table 3](#)). A. The colour-coded final scores of the EIPN are  
1228 displayed at the end of the Table as well as in [Figure 5](#). All blocks delineated in [Figure 5 A-C](#)  
1229 [and D-F](#) were attributed one of the colour-coded Exposure Index score levels.

1230 ESD Table 7. Confusion matrix of the selected PLR model. This Table crosses ‘real’ observed  
1231 EIPN scores with predicted ones when we applied the model to the initial (observed) data  
1232 (145 blocks). Grey boxes show well predicted EIPN values in contrast to yellow boxes  
1233 indicating poorly predicted EIPN values.

1234 ESD Table 8. Chi<sup>2</sup> test on discriminant variables that support HAC clusters.

1235 ESD Table 9. Chi<sup>2</sup> test on variables of timing, access and response that support block clusters  
1236 for relief operation in case of imminent evacuation.

1237

Clermont, September 24, 2022

Dr. James Goff, Editor in Chief,

Natural Hazards Journal

We would like to submit the revised article entitled: “*Semeru volcano, Indonesia: Measuring hazard, exposure and response of densely populated neighbourhoods facing persistent volcanic threats*” to Natural Hazards.

In response to your request, we apologize and we thank you for having drawn our attention to the short paragraphs that may be considered as plagiarism and needed to be reworked before peer review. We have revised the third paragraph in sections 1.1, the first and third paragraphs in section 1.2, the second paragraph in section 1.3, and finally the first paragraph in section 1.5. We have shortened and reworked the mentioned lines, indicated their sources and added citations where appropriate. In the case of sections 1.3 and 1.5, we have referred to the publication from which the lines have been borrowed. In fact, we borrowed the content of these lines from our publication “Thouret et al., 2022” (International Journal of Disaster Risk Reduction). We hope that the revision would meet your expectation before peer review.

The research presented in this article, which was initiated at Laboratoire Magmas et Volcans (LMV), Université Clermont-Auvergne (UCA) in France, has been conducted in collaboration with the Laboratoire de Mathématiques (also UCA), and with Indonesian partners of the IPB University, Bogor in the framework of the “RiskAdapt” research program funded by the French National Agency for Research (ANR).

The article shows the population exposure to Semeru’s volcanic threats from its persistent, daily eruptive activity, which endangers at least 50,000 of the 950,000 inhabitants living on its slopes and ring plain. Surveys, mapping and statistical investigation enabled us to assess the extent of exposure of 145 neighbourhoods (termed blocks) and characterize hazards and response to eruptions in 15 rural villages and small towns. Univariate, bivariate and multivariate analyses were used to explore data and characterize the relationships between 11 variables to compute a multi-component exposure index. Logistic Regression models allowed to select six optimal exposure variables. Multivariate analyses and Hierarchical Agglomerative Clustering distinguished four groups of blocks based on attributes of all variables correlated to the exposure index score.

To contribute to disaster risk reduction, ranking blocks using distance-timing and access to response facilities is a useful tool to point remote or blocked blocks during imminent eruptions including evacuation. Results should help disaster risk management staff to improve their participation at the scale of neighbourhoods on active volcanoes.

We hope this article meets the scope and high standards of Natural Hazards. Thank you very much for your consideration.

Sincerely yours,



Jean-Claude Thouret, First and corresponding author, Professor, Université Clermont Auvergne (UCA), OPGC, CNRS and IRD, Laboratoire Magmas et Volcans, F-63000 Clermont-Ferrand, France. Email: [j-claude.thouret@uca.fr](mailto:j-claude.thouret@uca.fr); Phone +33 4 73 34 67 73;

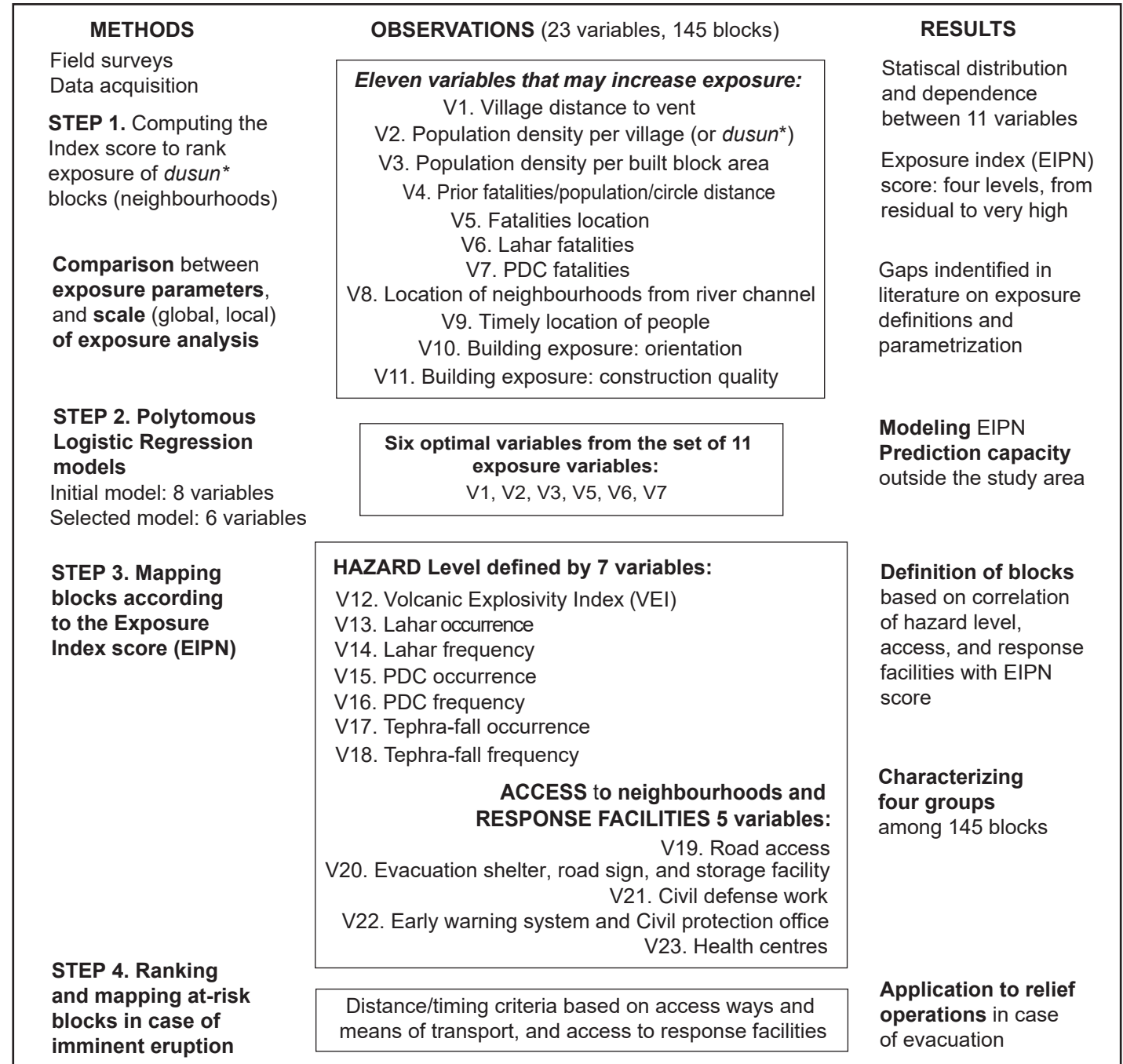
cell phone 33 6 25 19 41 17

Co-authors: Emeline Wavelet, MSc student at LMV, UCA; Marie Taillandier, MSc student Laboratoire de Mathématiques UCA; Boedi Tjahjono, Professor, Faculty of Agriculture, IPB University, Bogor, Indonesia; Nourddine Azzaoui, Professor, Laboratoire de Mathématiques UCA, and; Olivier Santoni (GIS engineer, FERDI also at UCA).

**STATISTICAL METHOD APPLIED TO THE SURVEY DATASET**

Uni-, Bivariate analyses, Factorial and Multiple Correspondence Analyses,  
Hierarchical Agglomerative Clustering, Polytomous Logistic Regression

**Measuring exposure, hazard and response of densely populated neighbourhoods facing Semeru's persistent volcanic threats**



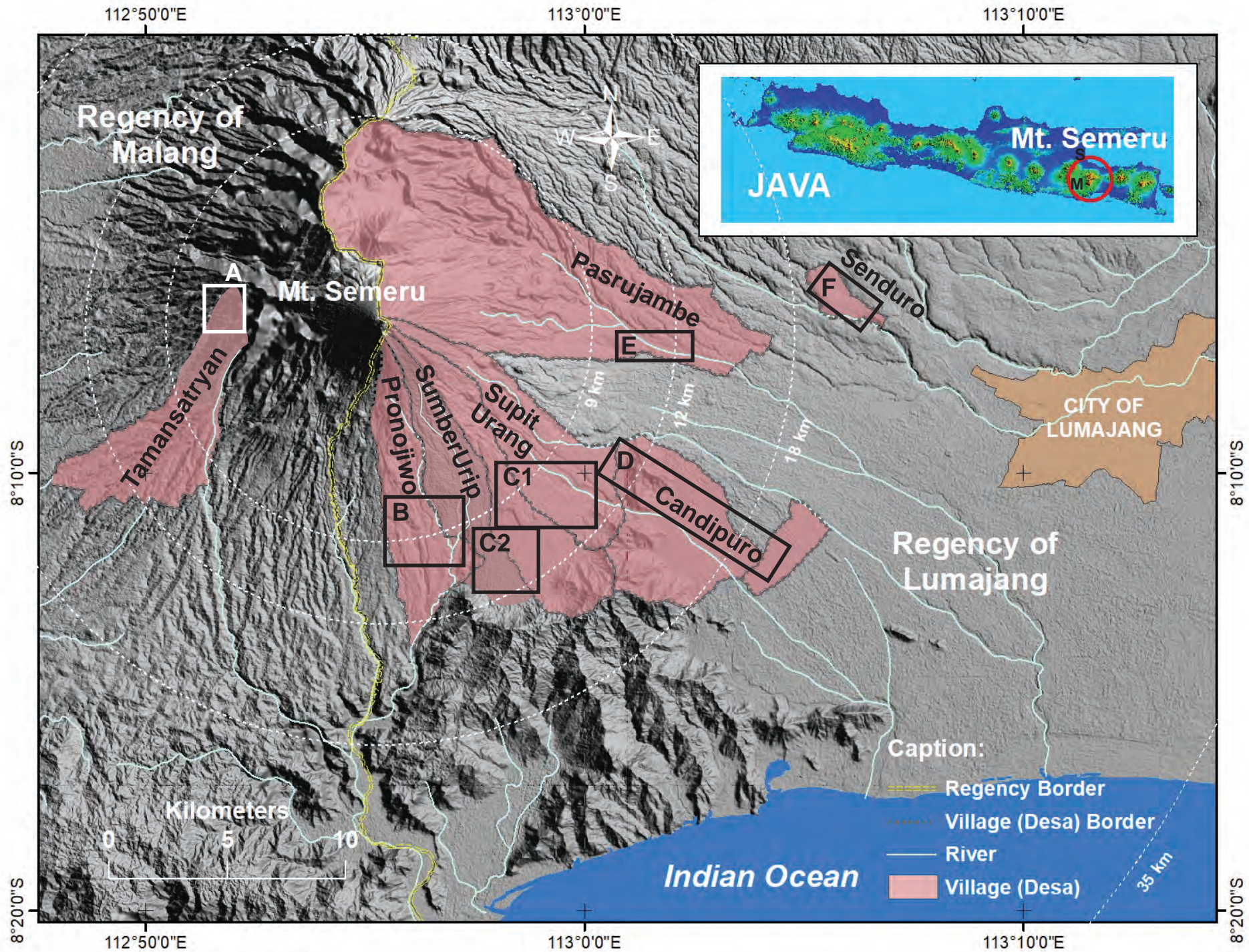
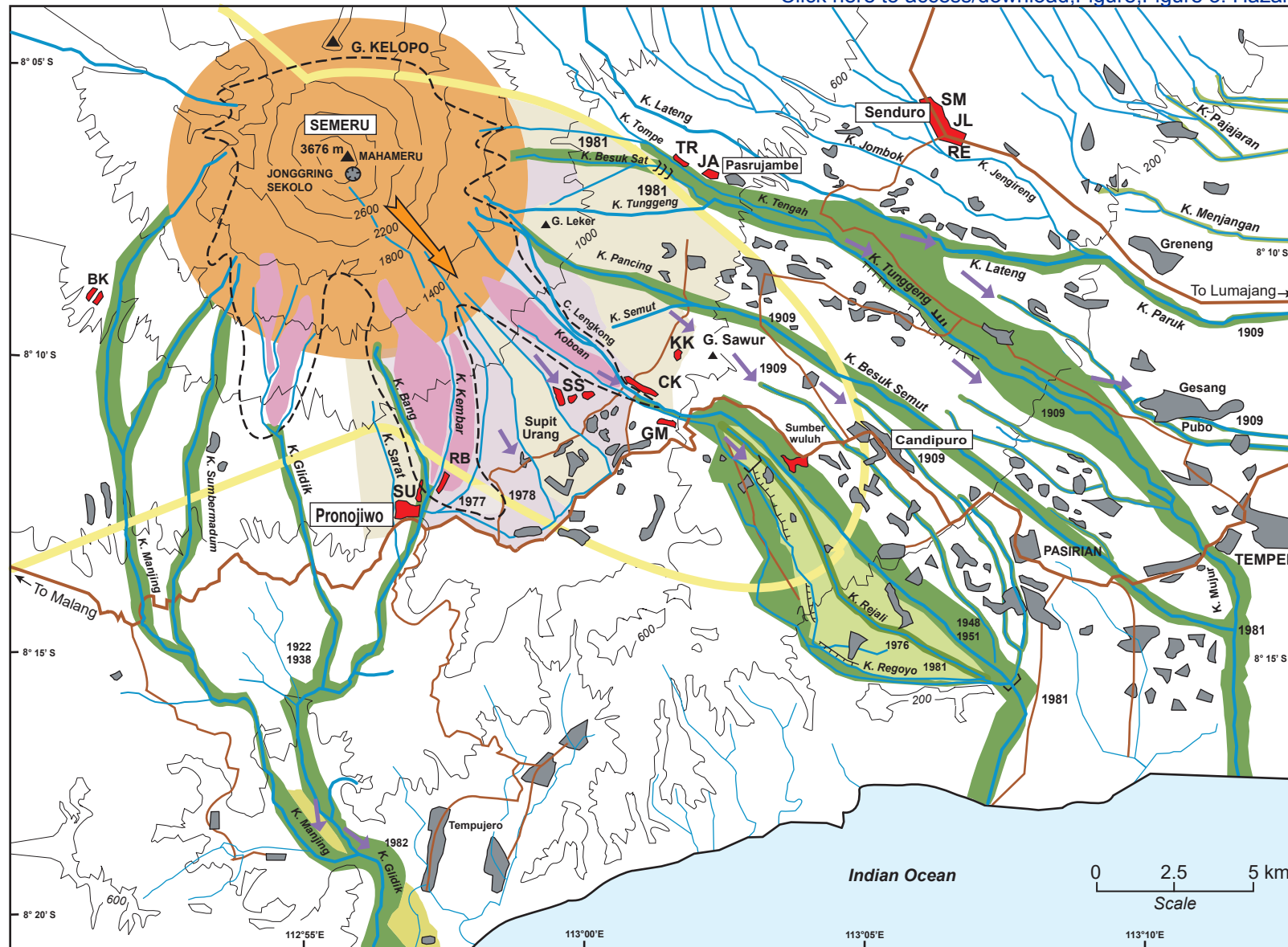


Fig. 2















- |   |   |   |  |
|---|---|---|--|
|  | Areas frequently affected by tephra fallout on annual basis (VEI 2) and ballistics within 5 km                                  |  | Areas likely to be mantled by annual ashfall associated to PDCs and dispersed in case of large eruption: towards East & SE (rain season), West & SW (dry season) |
|  | Boundary of potential pyroclastic density currents (PDCs) in case of VEI $\geq 3$ eruption                                      |  | Valleys swept by large-volume lahars (>1 million m <sup>3</sup> ), e.g., 1909, 1976 and 1981   |
|  | Areas affected by PDCs and companion fallout every 8 to 11 years on average (VEI 3)   |  | Alluvial plains affected by lahars in case of large-scale eruptions (VEI $\geq 3$ ) and/or heavy rainstorm   |
|  | Areas affected by PDCs and companion ashfall due to large VEI $\geq 3$ events every 11 to 25 years on average                   |  | Possible overbank and avulsion in case of large-volume lahars  |
|  | Areas likely to be affected by tephra-fall associated with PDCs or by fallout in case of large eruptions                        |  | Dykes and check dams (Sabo) along active valley channels   |
|  | Preferential path (scar of Jenggring-Seloko) guiding dome-fed rock avalanches, lava flows and pyroclastic flows on annual basis |  | Dusun (sub-village) and small towns under study  |

Fig. 3



Figure 4

[Click here to access/download;Figure;Figure 4. Results of UA exposure \(29.03.22\).pdf](#)

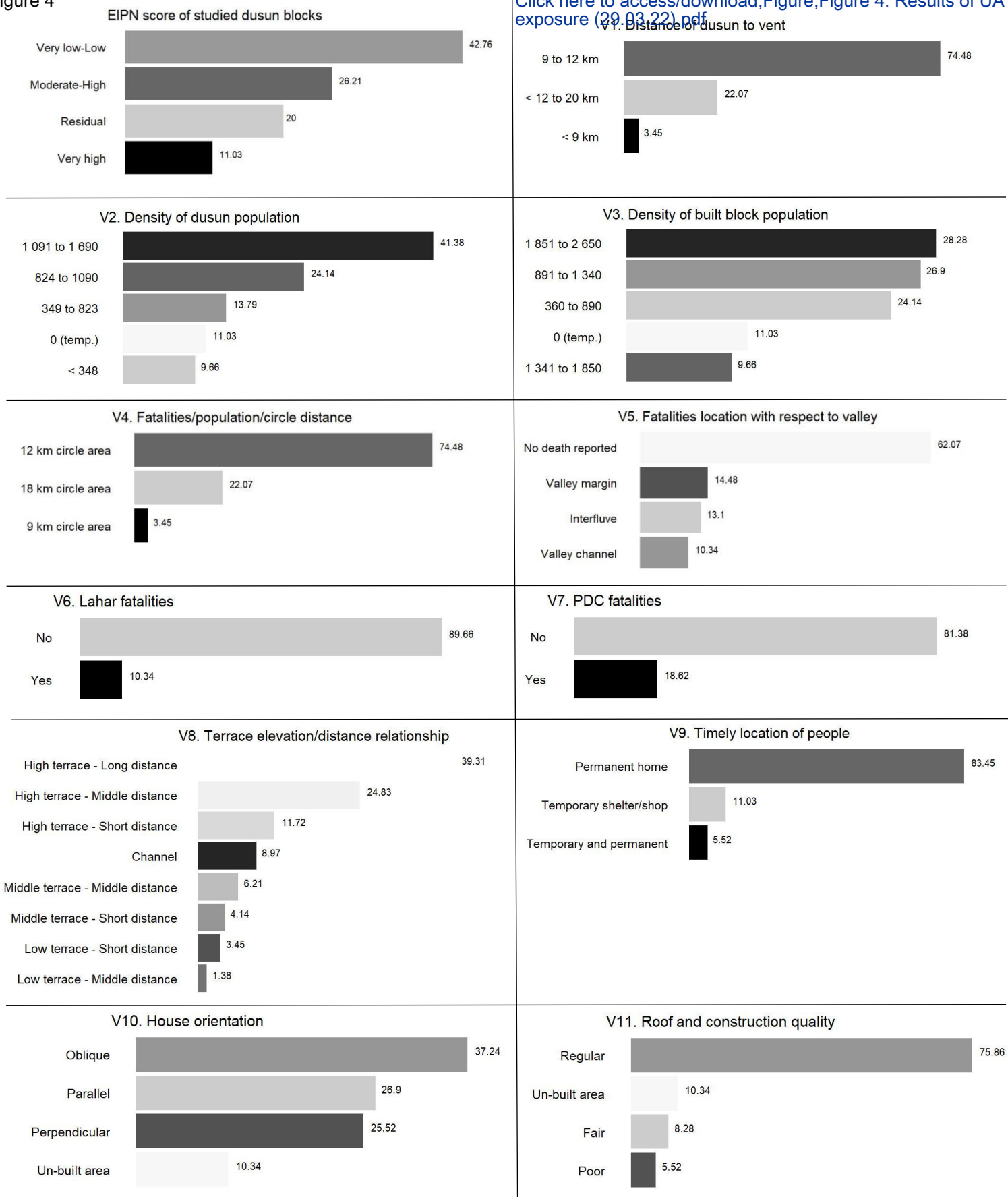
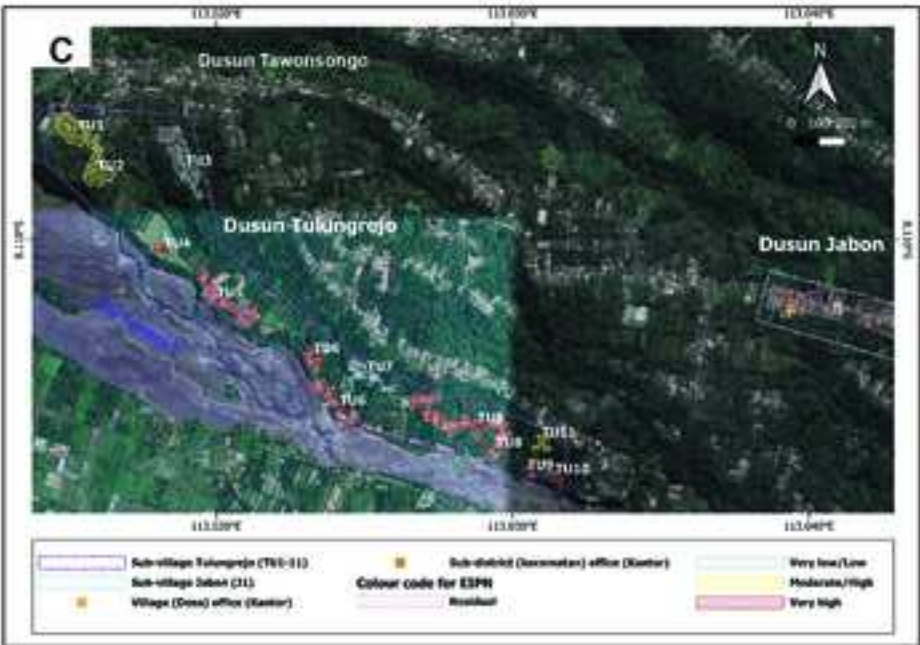
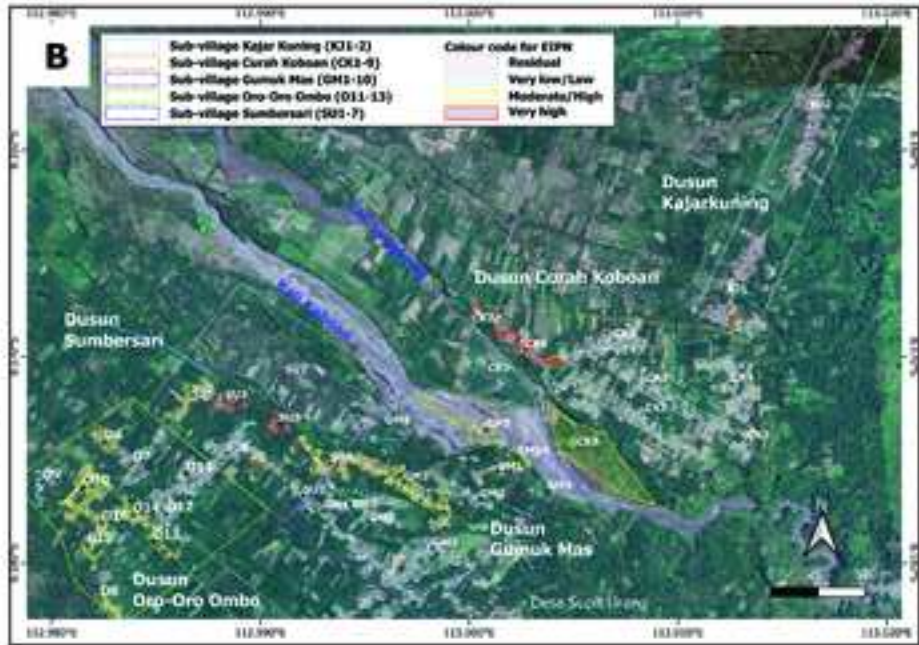


Fig. 4



Exposure Index of Populated Neighbourhoods Level, colour and score range		
EIPN	COLOUR	SCORE
Residual	beige	1.17 - 3.45
Very low/Low	light blue	3.46 - 4.34
Moderate/High	yellow	4.35 - 5.34
Very high	red	5.35 - 6.56

Figure 5 ABC

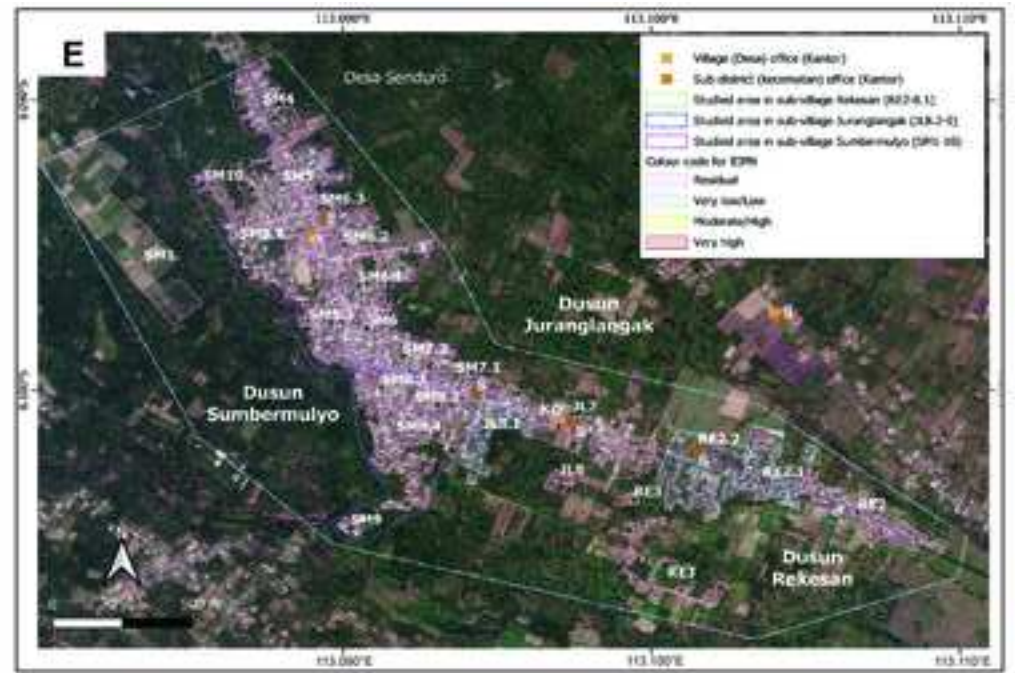
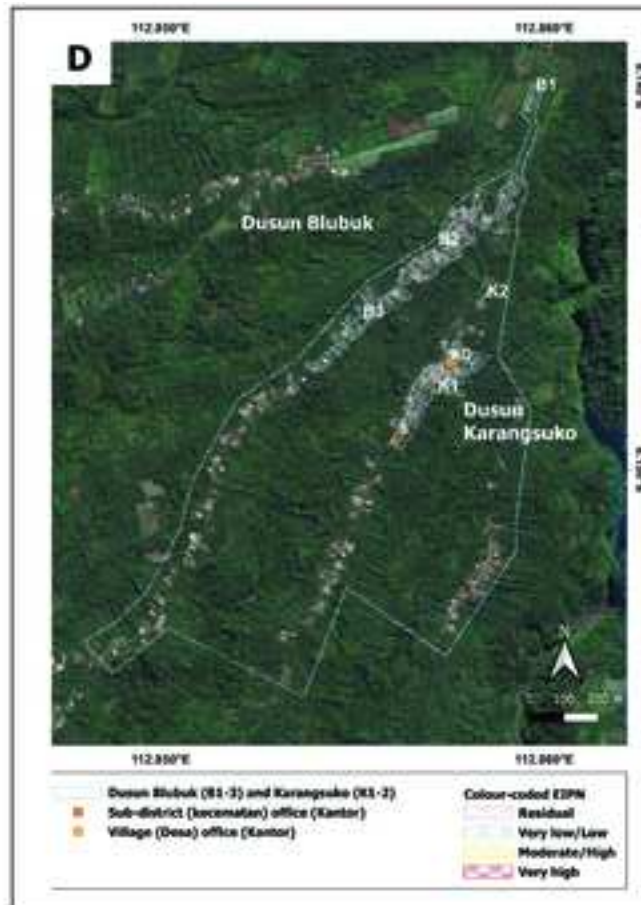


Figure 5 DE

Figure 6

[Click here to access/download;Figure;Figure 6. MCA Scatterplot axes 1-2 MT \(27.04.22\).pdf](#)

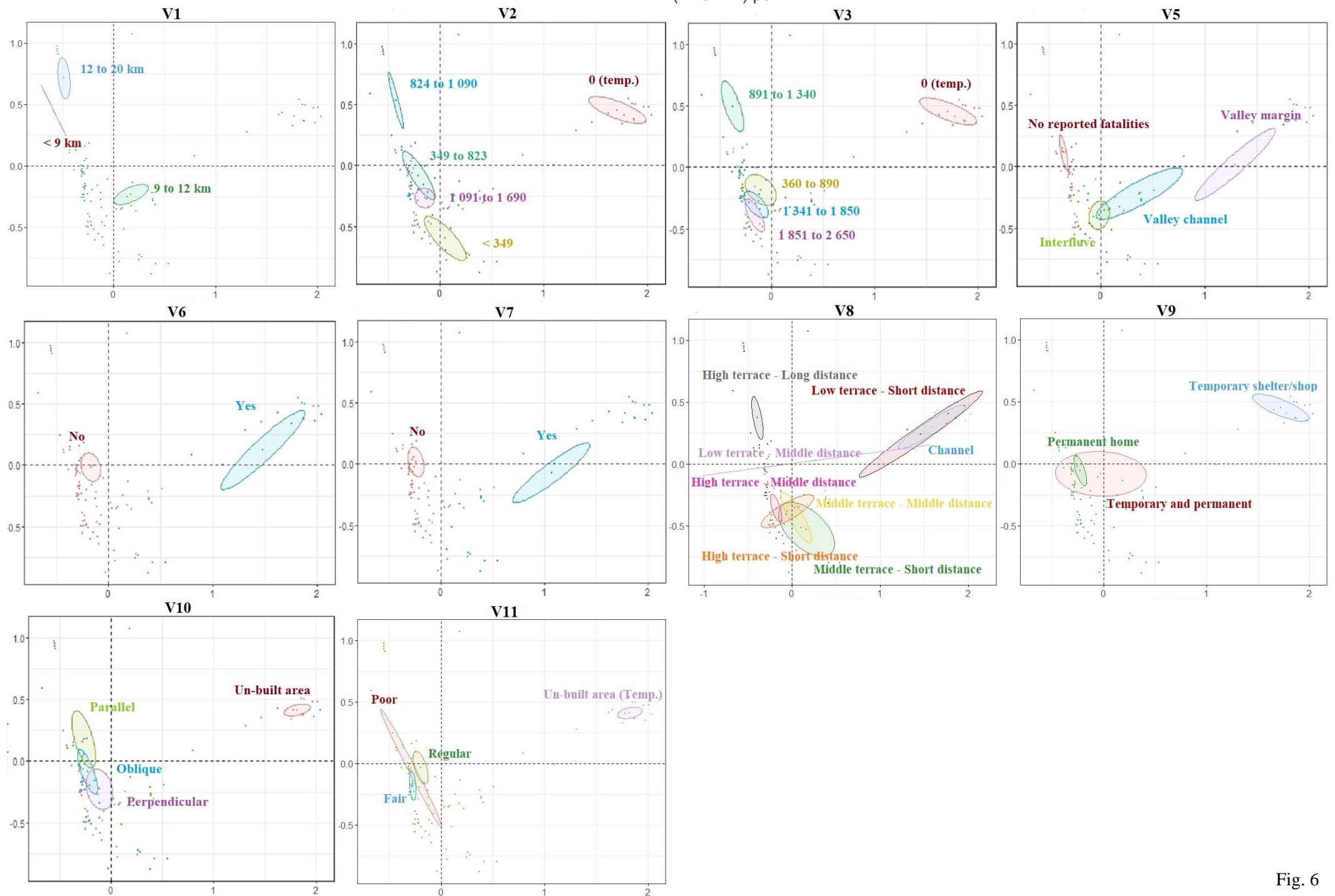


Fig. 6

Figure 7

[Click here to access/download;Figure;Figure 7. Results UA hazards & response MT \(04-08-22\).pdf](#)

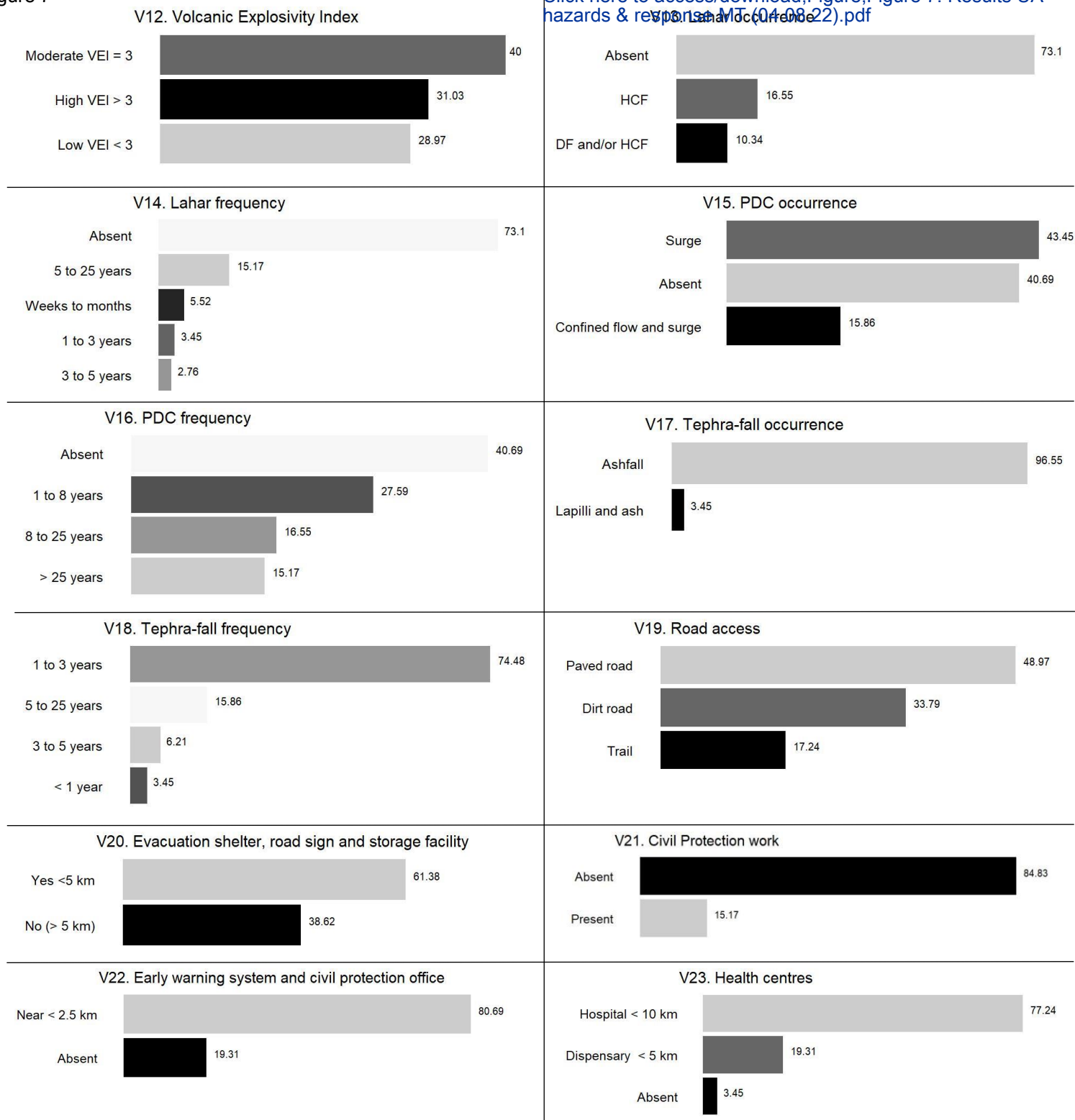


Fig. 7

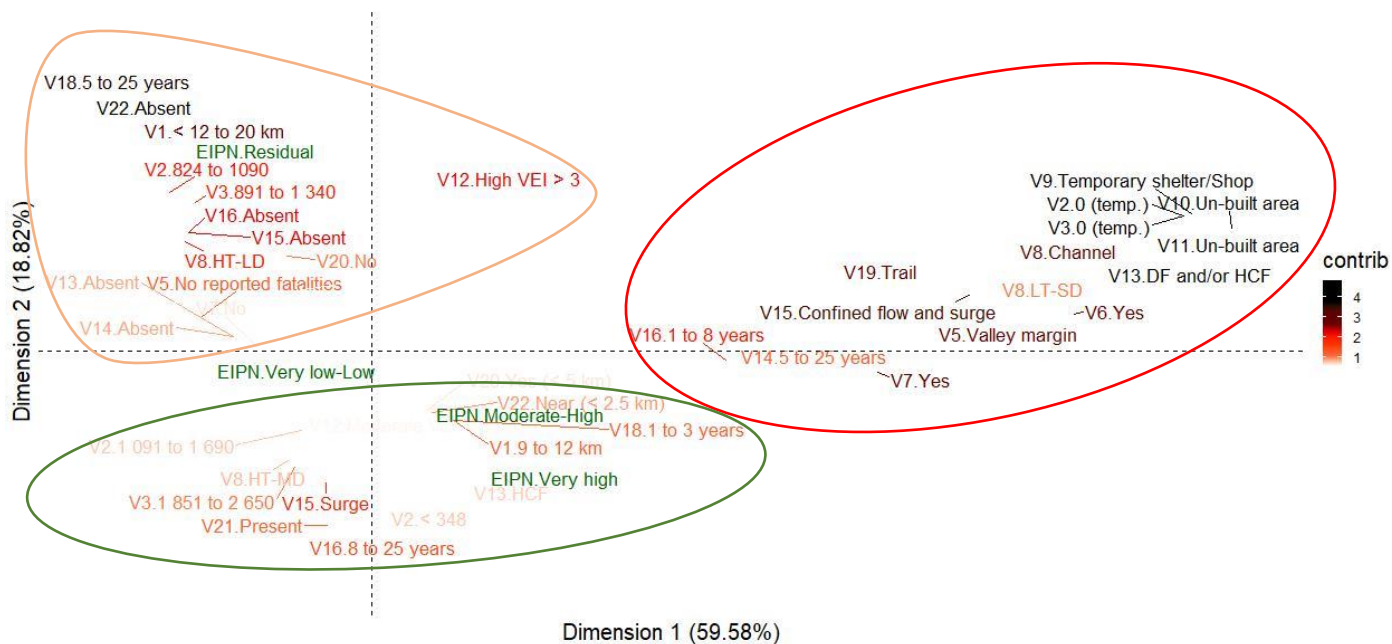


Fig. 8A

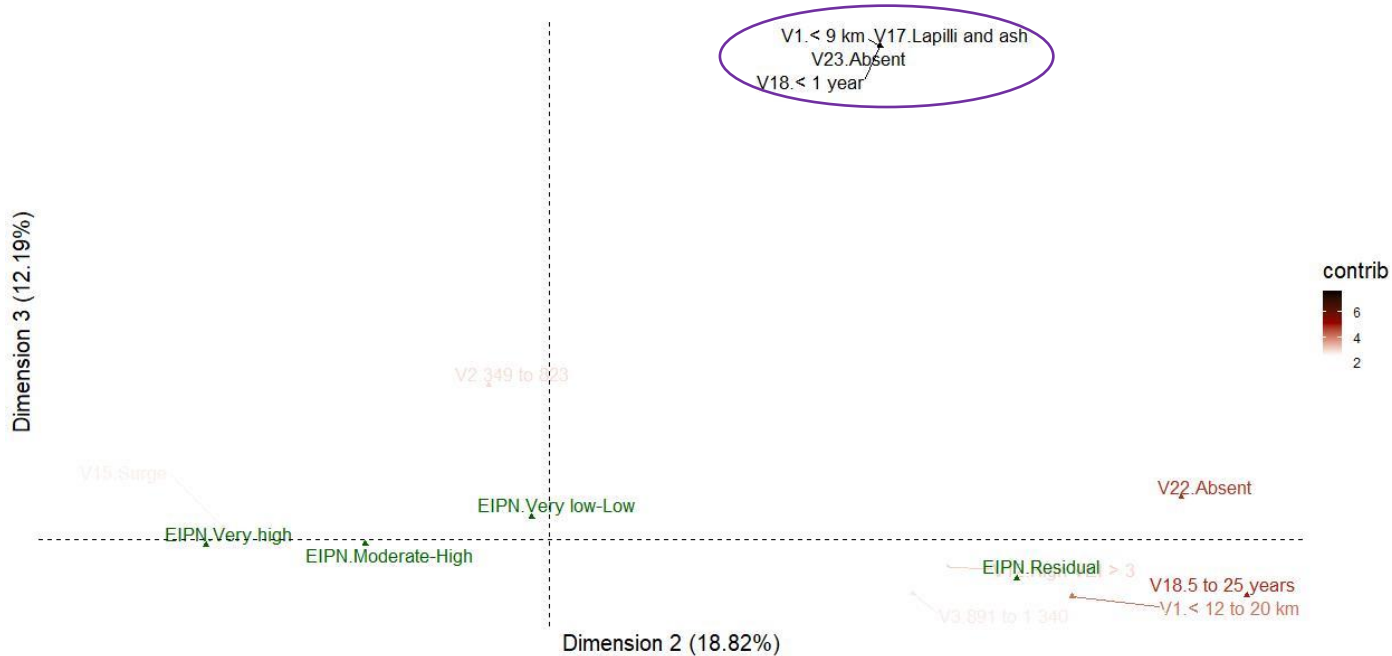
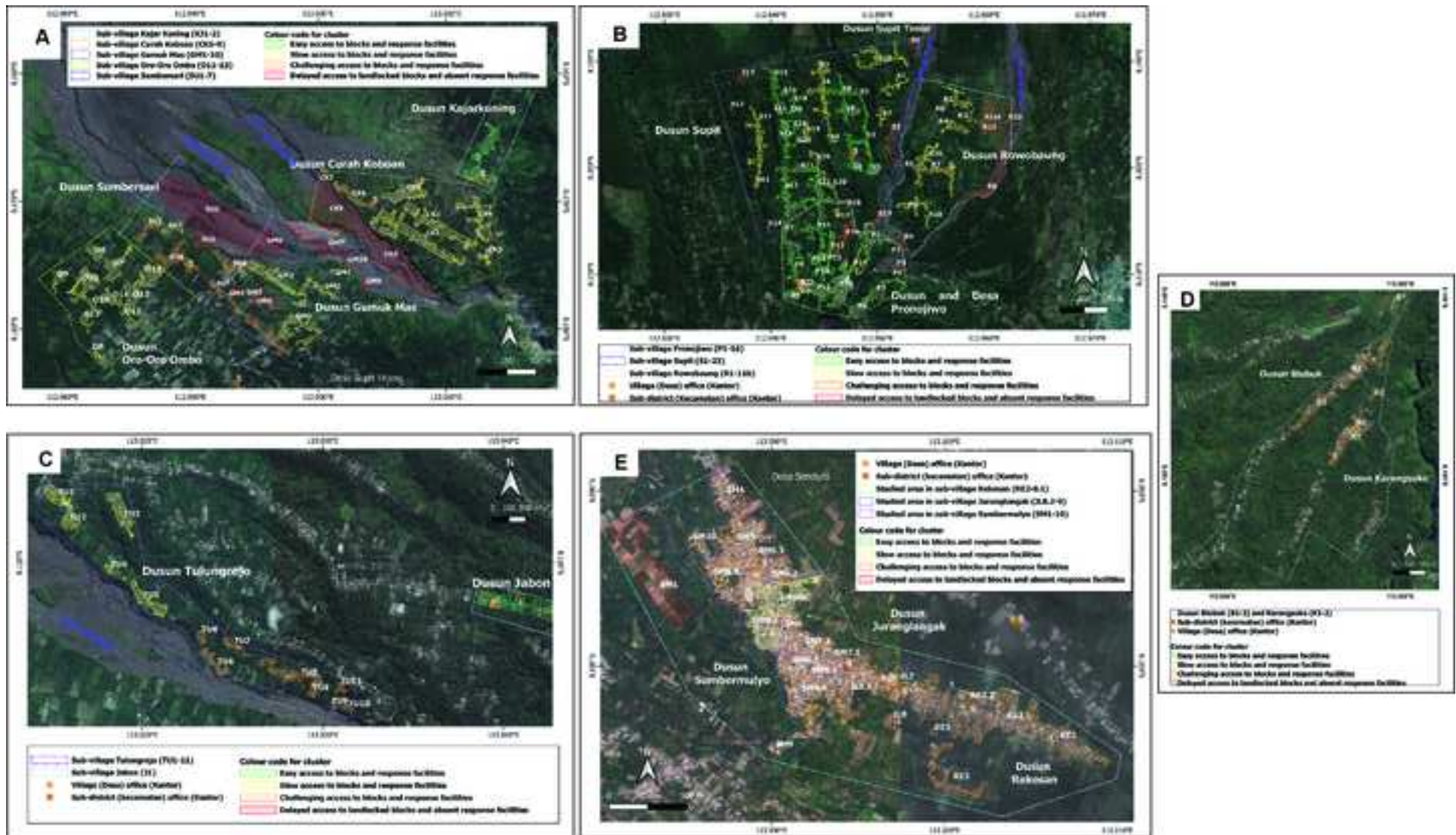


Fig. 8B

Figure 9



<i>Dataset and observations</i>	<i>Methods</i>	<i>Techniques</i>	<i>Purposes</i>
11 EIPN variables 7 Hazard variables 5 Access/Response variables	Univariate Analysis UA	Frequency: bar plots	To obtain the frequency of all categories of every variable
11 EIPN variables Dataset of all variables	Bivariate Analysis BA	Chi-square test Contingence tables: Burt Table Biplot illustrations Factorial Correspondence Analysis FCA	To identify dependence or independence between two variables Allows to analyse the cross frequency between two variables To show links between categories of two variables in biplots
Initial PLR model: Choice of 8 variables among 11 initial EIPN variables	Polytomous Logistic Regression PLR	Backward and forward model selection Comparison of deviance and Akaike Information Criterion (AIC)	To select variables that are more relevant and discard less relevant ones. To reduce the number of optimal variables
Sorting out the 6 most pertinent variables and 43 categories= Final PLR Model for predicting EIPN at the block scale beyond the survey area, potentially applicable to any given active volcano	PLR models to select pertinent variables and predict Exposure Index of Populated Neighbourhoods	Study of probability or significativity of the categories  Explore the confusion matrix Analysis of coefficients of the selected model  Predict function (Application)	To analyse the significativity of categories and select variables that show the lowest probability of error. To determine the predictive quality of the model. To assess the strength of association between each active variable and the variable to be explained (EIPN): positive coefficient (i.e., increasing exposure) or negative coefficient (i.e., decreasing exposure). To compute the probability to obtain EIPN score for a block. To measure the effect of an explanatory variable on EIPN. Ultimate goal: to select optimal variables used around any similarly populated, active volcano.
Set of 23 variables: - Illustrative variables (e.g., EIPN score) - Active variables: V1...V23 (see Table 3)	Multiple Correspondence Analysis MCA	Study of eigenvalues (Benzécri correction) Graphic illustrations: ellipses  Fisher (variables) and Student (categories) tests  Study and projection of observations (e.g., blocks) in biplots	To define and describe the suitable number of clusters To determine proximities and oppositions between categories of one variable in the projection of blocks To determine correlations of variables along each dimension and coordinates of categories To establish groups of observations that provide more contribution and having similar behaviour. To analyse links between different categories.
Outputs of two MCAs conducted on 23 variables	Hierarchical Agglomerative Clustering HAC	Dendrograms and Factor maps Chi-square test Analysis of clusters	To define and describe the suitable number of clusters. To identify variables which discriminate clusters. To identify discriminant categories and their frequencies.
Group of 5 access/response variables and distance/time	Frequency analysis HAC clusters	Distance/time parameter extracted from frequency classification	To identify the remote and landlocked blocks in case of imminent eruption and evacuation.



Table 1

Table 2

Variable V1 Distance of dusun to vent	Population density (inhab/km <sup>2</sup> )			Prior fatalities/dusun population/circle distance, and location with respect to valleys			Prior fatalities		V8 Location of dusun blocks with respect to channel: Terrace elevation / distance relationship					V9 Timely location of people		Edifice exposure				
	V2 Density of dusun	V3 Density of built block in dusun		V4 Fatalities/population/circle distance			V5 Fatalities location		V6 Lahar fatalities		V7 PDC fatalities							V10 House orientation		V11 Roof and construction quality
< 9 km	1 091 to 1 690	1 851 to 2 650	fatalities / pop. / 9 km circle	Valley margin	Yes	Channel	Low terrace - Short distance	Temporary and permanent	Permanent home	Perpendicular	Oblique	3	2	1	0	3	2	1	0	
9 to 12 km	824 to 1 090	1 341 to 1 850	fatalities / pop. / 12 km circle	Valley channel	No	LT - Middle distance	Permanent home	Permanent home	Oblique	Parallel	3	2	1	0	3	2	1	0		
> 12 to 20 km	349 to 823 < 348	891 to 1 340 360 to 890	fatalities / pop. / 18 km circle	Interfluve	Yes	LT - Long distance	Temporary shelter	Temporary shelter	Un-built	Poor	Regular	3	2	1	0	3	2	1	0	
				No death reported	No	Middle terrace - SD			Fair	Un-built										

Table 2.1. Eleven variables (V) increase exposure.

V12 Volcanic Explosivity Index	Hazard type and frequency														
	V13 Lahar occurrence			V14 Lahar frequency			V15 PDC occurrence		V16 PDC frequency		V17 Tephra-fall occurrence		V18 Tephra-fall frequency		
High VEI ≥ 4	Debris flow and/or banjir (hyperconcentrated flow)			Weeks to months			Surge		Confined flow and surge		Lapilli and ash, distance <8 km		Ashfall beyond 8 km		
Moderate VEI 3	HCF (banjir)			1 to 3 years			None		1 to 8 years		None		< 1 year		
Low VEI ≤ 2	None			3 to 5 years			Surge		8 to 25 years		Lapilli and ash, distance <8 km		1 to 3 years		
	None			5 to 25 years			Confined flow and surge		< 25 years		Ashfall beyond 8 km		3 to 5 years		
	None			None			Surge		None		Ashfall beyond 8 km		5 to 25 years		
	None			None			Surge		None		Ashfall beyond 8 km		Perpendicular		

Table 2.2. Second group of variables: volcanic hazards posed to the surveyed neighbourhoods.

V19 Ways of access			V20 Evacuation shelter, road sign, storage facility		V21 Civil Protection work		V22 Early warning system and BNPB office			V23 Health Centres			
Paved road	Yes < 5 km	Present	Near < 2.5 km	Hospital < 10 km	Present	Near < 2.5 km	Hospital < 10 km	0	1	2	0	1	2
Dirt road	Yes > 5 km	None	Far > 2.5 km	Dispensary < 5 km	None	None	Dispensary < 5 km	1	2	0	1	2	0
Trail	No	None	None	None	None	None	None	0	1	2	0	1	2

Table 2.3. Third group of variables: accessibility of neighbourhoods and response facilities.

Computing steps	Variables		Thresholds				Comments
			< 9	12	18	35	
	Distance, km		< 9	12	18	35	Events / fatalities for 1884 - 2014
<b>First step</b>	Volcanic EVENTS		7	9	9	9	Total events with 25 fatalities
	Weighting		0.206	0.265	0.265	0.265	
	Reported FATALITIES		10	595	255	255	1884-2014: 1115 victims (7/year), Table 2, Thouret et al. (2007)
	Weighting		0.009	0.534	0.229	0.229	
	DISTANCE CIRCLE		9-km circle	12-km circle	18-km circle	35-km circle	Circle radius around the Semeru crater
	No. of villages		2 Dusun	Dusun except Supit	other	None in surveys	
	POPULATION		2,864	59,236	85,303	842,597	Total 35-km half circle E, S and SW slopes: c. 843,000 people
	Weighting		0.003	0.060	0.086	0.851	
<b>Second step</b>	SURFACE AREA		9 km radius	12	18	35	Circle surface area
	2/3 of the SA (East)		169.646	301.593	678.584	2565.634	Total 4 areas
	The 9 km area is 1.778 times smaller than the 12 km area: $12/9 \text{ km} = 1.778$						
	The 9 km area 4.00 times smaller than the 18 km area; $18/9 \text{ km} = 4.000$						
	The 9 km area 15.123 times smaller than the 35 km area; $35/9 \text{ km} = 15.123$						
	Total circles = $9 \text{ km}/9 \text{ km} + 12 \text{ km}/9 \text{ km} + 18 \text{ km}/9 \text{ km} + 35 \text{ km}/9 \text{ km} = 21.901$						
	In the 9 km-distance circle we obtain $1/21,901 = 0.046$						
	In the 12 km-distance circle we obtain $1.778/21,901 = 0.081$						
	In the 18 km-distance circle we obtain $4.00/21,901 = 0.183$						
	In the 35 km-distance circle we obtain $15.123/21,901 = 0.691$						
<b>Third step</b>	Two sets of weights are combined and scaled (scaled= with respect to the total circle)						
	Weighting						
	< 9-km circle		$0.28 * 0.046 / (\text{Total}) = 0.206 * 0.046 / 0.263 = 0.036$				
	12- km circle		$(0.36 * 0.081) / 0.356 = 0.082$				
	18- km circle		$(0.36 * 0.183) / 0.356 = 0.185$				
	35- km circle		$(0.36 * 0.691) / 0.356 = 0.698$				
			Total = $0.206 * 0.046 + 0.265 * 0.081 + 0.265 * 0.183 + 0.265 * 0.691 = 0.263$				Coefficients used in Table 3
<b>Fourth step</b>	Population within each circle distance category is multiplied by weighting and the four figures are summed						
	<9 km-distance circle		$2,864 \times 0.036 = 103$				
	12 km-distance circle		$59,236 \times 0.082 = 4,858$				
	18 km-distance circle		$85,303 \times 0.185 = 15,739$				
	35 km-distance circle		$842,597 \times 0.698 = 587,803$				
			Sum = 608,502				
	<b>PEI values attributed to population categories per circles:</b>						
	Population	Rounded population	PEI			Circle radius distance, km	Smaller than
	103	105	1			9	x 1
	4,858	5,000	2			12	x 1.778
	15,739	16,000	3			18	x 4.000
	587,803	590,000	4			35	x 15.123
	> 608,502	> 608,502	5			> 35	

Table 3

Coefficients of the selected categories						
Variable	Category	Estimate	Std. Error	Z value	Pr(> z )	Significat.
V1 dist	9 to 12 km	5.90	2.53	2.33	0.02	*
V1	12 to 20 km	9.70	2.94	3.30	0.00	***
V2 densD	824 to 1090	2.28	1.25	1.82	0.07	.
V2	349 to 823	1.76	1.29	1.37	0.17	
V2	< 348	4.99	1.48	3.37	0.00	***
V2	0 (temp.)	34.71	7.37	4.71	2.47e-06	***
V3 densB	1 341 to 1 850	0.43	1.38	0.31	0.76	
V3	891 to 1 340	3.05	1.01	3.01	0.00	**
V3	360 to 890	7.71	1.53	4.99	6.15e-07	***
V3	0 (temp.)	NA	NA	NA	NA	
V5 locat	Valley channel	7.25	3.27	2.21	0.03	*
V5	Interfluve	11.88	4.00	2.97	0.00	**
V5	No death reported	14.77	4.26	3.46	0.00	***
V6 LH	Yes	-9.47	4.47	-2.12	0.03	*
V7 PDC	Yes	-8.27	2.19	-3.78	0.00	***
V8	Low terrace – Short distance	-0.16	4.12	-0.04	0.97	
V8	Low terrace – Middle distance	-1.78	9.97	-0.18	0.86	
V8	Middle terrace – Short distance	-0.61	2.77	-0.22	0.83	
V8	Middle terrace – Middle distance	-0.68	3.30	-0.21	0.84	
V8	High terrace – Short distance	4.23	2.40	1.76	0.08	.
V8	High terrace – Middle distance	5.57	2.53	2.20	0.03	*
V8	High terrace – Long distance	6.73	2.72	2.47	0.01	*
V9	Temporary shelter/shop	1.14	1.36	0.84	0.40	
V9	Permanent home	NA	NA	NA	NA	

## Threshold coefficients

	Estimate	Std.	Z value
Very high   Moderate - High	17.87	4.73	3.78
Moderate – High   Very low - Low	28.28	6.03	4.69
Very low – Low   Residual	35.73	6.80	5.25

Table 4A. Initial model for polytomous logistic regression.

Coefficients of the selected categories						
Variable	Category	Estimate	Std. Error	Z value	Pr(> z )	Significat.
V1 dist	9 to 12 km	4.81	2.06	2.34	0.02	*
V1	12 to 20 km	9.06	2.43	3.73	0.00	***
V2 densD	824 to 1090	3.50	0.92	3.81	0.00	***
V2	349 to 823	3.85	1.07	3.60	0.00	***
V2	< 348	6.19	1.34	4.61	4.13e-06	***
V2	0 (temp.)	25.83	4.94	5.23	1.68e-07	***
V3 densB	1 341 to 1 850	-0.90	1.06	-0.85	0.39	
V3	891 to 1 340	1.88	0.69	2.72	0.01	**
V3	360 to 890	5.39	1.13	4.77	1.87e-06	***
V3	0 (temp.)	NA	NA	NA	NA	
V5 locat	Valley channel	7.12	2.79	2.55	0.01	*
V5	Interfluve	12.25	3.48	3.52	0.00	***
V5	No death reported	15.45	3.72	4.16	3.22e-05	***
V6 Lahar	Yes	-7.20	2.52	-2.86	0.00	**
V7 PDC	Yes	-5.58	1.42	-3.94	8.05e-05	***

## Threshold coefficients

	Estimate	Std. Error	Z value
Very high   Moderate - High	13.86	3.86	3.59
Moderate – High   Very low - Low	21.43	4.62	4.64
Very low – Low   Residual	27.96	4.18	5.40

Table 4B. Selected model for polytomous logistic regression

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	
V1	0.000	0.000	0.000	0.000	0.000	<b>0.062</b>	0.005	0.000	<b>0.301</b>	0.000	0.000	0.000	0.004	0.040	0.000	0.000	0.000	0.000	0.001	0.000	0.014	0.000	0.000	
V2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
V3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.006	0.000	0.000	0.000	0.000
V4	0.000	0.000	0.000	0.000	0.000	<b>0.061</b>	0.005	0.000	<b>0.303</b>	0.000	0.000	0.000	0.004	0.041	0.000	0.000	0.000	0.000	0.001	0.000	0.016	0.000	0.000	0.000
V5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.227</b>	0.000	0.000	0.000	0.000	<b>0.431</b>	0.001	0.000	0.002	<b>0.647</b>	0.000	<b>0.127</b>	
V6	<b>0.061</b>	0.000	0.000	<b>0.060</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	<b>0.658</b>	<b>0.141</b>	0.000	0.045	<b>1</b>	<b>0.078</b>	<b>0.671</b>	
V7	0.005	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.102</b>	0.000	0.000	0.000	0.000	<b>0.582</b>	0.011	0.000	<b>0.187</b>	<b>0.766</b>	0.005	<b>0.161</b>	
V8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.279</b>	0.000	0.000	0.000	0.000	0.000	0.019	
V9	<b>0.304</b>	0.000	0.000	<b>0.302</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	<b>0.736</b>	<b>0.391</b>	0.000	<b>0.237</b>	<b>0.162</b>	<b>0.138</b>	<b>0.724</b>	
V10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.025	0.000	0.000	<b>0.723</b>	<b>0.357</b>	0.011	<b>0.140</b>	
V11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	<b>0.205</b>	<b>0.090</b>	0.028	0.001	
V12	0.000	0.000	0.000	0.000	<b>0.231</b>	0.005	<b>0.107</b>	0.000	0.002	0.002	0.000	0.000	0.015	0.003	0.000	0.000	0.013	0.000	0.000	0.003	0.000	0.000	0.000	
V13	0.004	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.000	<b>0.265</b>	0.015	0.000	<b>0.509</b>	0.022	0.026	<b>0.309</b>	
V14	0.043	0.000	0.000	0.041	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	<b>0.514</b>	<b>0.148</b>	0.000	<b>0.174</b>	<b>0.219</b>	<b>0.106</b>	0.002	
V15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.052</b>	0.000	0.000	<b>0.225</b>	
V16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	<b>0.127</b>	0.000	0.000	0.000	0.000	0.000	0.017	
V17	0.000	0.000	0.004	0.000	<b>0.435</b>	<b>0.656</b>	<b>0.580</b>	<b>0.275</b>	<b>0.746</b>	0.024	0.001	0.013	<b>0.271</b>	<b>0.504</b>	<b>0.521</b>	<b>0.122</b>	0.000	0.000	0.006	0.007	<b>0.600</b>	0.000	0.000	
V18	0.000	0.000	0.000	0.000	0.001	<b>0.140</b>	0.012	0.000	<b>0.383</b>	0.000	0.000	0.000	0.014	<b>0.147</b>	0.000	0.000	0.000	0.000	0.003	0.000	0.038	0.000	0.000	
V19	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.002	0.000	0.036	0.017	<b>0.063</b>	0.002	
V20	0.000	0.000	0.006	0.000	0.001	0.045	<b>0.190</b>	0.000	<b>0.242</b>	<b>0.722</b>	<b>0.201</b>	0.003	<b>0.505</b>	<b>0.168</b>	<b>0.050</b>	0.000	0.007	0.000	0.035	0.000	0.033	0.000	0.000	
V21	0.013	0.001	0.000	0.013	<b>0.648</b>	<b>1</b>	<b>0.767</b>	0.000	<b>0.162</b>	<b>0.355</b>	<b>0.083</b>	0.000	0.020	<b>0.225</b>	0.000	0.000	<b>0.603</b>	0.035	0.018	0.035	0.000	0.014	<b>0.513</b>	
V22	0.000	0.000	0.000	0.000	0.000	<b>0.075</b>	0.005	0.000	<b>0.136</b>	0.011	0.030	0.000	0.027	<b>0.113</b>	0.000	0.000	0.000	0.000	<b>0.060</b>	0.000	0.015	0.000	0.000	
V23	0.000	0.000	0.000	0.000	<b>0.131</b>	<b>0.675</b>	<b>0.167</b>	0.019	<b>0.724</b>	<b>0.143</b>	0.001	0.000	<b>0.308</b>	0.001	<b>0.224</b>	0.017	0.000	0.000	0.002	0.000	<b>0.515</b>	0.000	0.000	

V1: Distance of village to vent

V9: Timely location of people

V16: PDC frequency

V2: Density of sub-village

V10: House orientation with respect to flow

V17: Tephra-fall

V3: Density of built block area

V11: Roof and construction quality

V18: Tephra-fall frequency

V4: Fatalities/village population/circle distance

V12: Volcanic Explosivity Index

V19: Access ways

V5: Location of fatalities

V13: Lahars

V20: Evacuation shelter, evacuation sign and storage facility

V6: Lahar fatalities

V14: Lahar frequency

V21: Civil Defense work

V7: PDC fatalities

V15: PDCs

V22: Early warning system and Civil protection office

Table 5

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
The most frequent attributes				
EXPOSURE	V2. Density of sub-village: 0 (temp.) V3. Density of built block: 0 (temp.) V6. Lahar fatalities: Yes V7. PDC fatalities: Yes V9. Timely location of people: Temporary shelter/shop V10. House orientation: Un-built area V11. Roof and construction quality: Un-built area	<b>EIPN score: Residual</b> V1. Distance to vent: [12, 20] km V2. Density of sub-village: [824, 1090] V3. Density of built block: [891, 1340]	<b>EIPN score: Moderate - High</b> V1. Distance to vent: [9, 12] km V2. Density of sub-village: [1091, 1690] V3. Density of built block: [1851, 2650] V9. Timely location of people: Permanent home	<b>EIPN score: Very low – Low</b> V1. Distance to vent: < 9km V2. Density of sub-village: [349, 823] V3. Density of built block: [360, 890] V11. Roof and construction quality: Poor
HAZARDS	V13. Lahar occurrence: DF and/or HCF V15. PDC occurrence: Confined flow and surge	V12 VEI: High VEI > 3 V18. Tephra fall frequency: 5 to 25 years	V15. PDC occurrence: Surge V18. Tephra fall frequency: 1 to 3 years	V17. Tephra fall: Lapilli & ash (dist < 9 km) V18. Tephra fall frequency: < 1 year
RESPONSE	V19 Access way: Trail	V20. Evacuation shelter, evacuation sign and storage facility: No V22. Early warning system and BPBD office: None	V20. Evacuation shelter, evacuation sign and storage facility: Yes < 5 km V22. Early warning system and BPBD office: Near (< 2.5km)	V22. Early warning system and BPBD office: None V23. Health Centres: None
The least frequent attributes				
EXPOSURE	V5. Fatalities location with respect to valley: No reported V6. Lahar fatalities: No V7. PDC fatalities: No V9. Timely location of people: Permanent V11. Roof and construction quality: Regular	V1. Distance to vent: [9, 12] km	<b>EIPN score: Residual</b> V2. Density of sub-village: 0 (temp.) V3. Density of built block: 0 (temp.) V10. House orientation: Un-built area V11. Roof and construction quality: Un-built area	V1. Distance to vent: [9, 12] km
HAZARDS	V13. Lahar occurrence: None V14. Frequency: None V15. PDC occurrence: Surge	V18. Tephra fall frequency: 1 to 3 years	V12. VEI: High VEI > 3 V18. Tephra fall frequency: 5 to 25 years	V17. Tephra fall: Ashfall > 8 km V18. Tephra fall frequency: 1 to 3 years
RESPONSE	V19 Access way: Paved road	V20. Evacuation shelter, evacuation sign and storage facility: Yes < 5 km V22. Early warning system and BPBD office: Near (< 2.5 km)	V20. Evacuation shelter, evacuation sign and storage facility: No V22. Early warning system and BPBD office: None	V20. Evacuation shelter, evacuation sign and storage facility: Yes < 5 km V22. Early warning system and BPBD office: Near (< 2.5 km) V23. Health Centres: Hospital (< 10 km)

		Cluster 1	Cluster 2	Cluster 3	Cluster 4
The most frequent attributes					
<b>TIMING</b>		Timing to access shelter: 10 minutes Timing to access Health centres: 10 to 25 min.	Timing to access Health centres: 45 min. Timing to access shelter: 25 min.	Timing to access shelter: 170 minutes Timing to access Health centres: 45 to 170 min.	Timing to access Health centres: 80 to 170 min. Timing to access shelter: 170 min.
	<b>RESPONSE</b>	V19 Access road: Paved road V20. Evacuation shelter, evacuation sign and storage facility: Yes $\leq$ 5 km V22. Early warning system and BPBD office: Near $\leq$ 2.5 km V23. Health Centres: Dispensary $\leq$ 5 km	V19. Access road: Dirt road V21. Civil protection work: Present V23. Health Centres: Hospital $\leq$ 10 km V22. Early warning system and BPBD office: Near $\leq$ 2.5 km	V22. Early warning system and BPBD office: None V20. Evacuation shelter, evacuation sign and storage facility: None V19 Access road: Paved road	V19 Access road: Trail V21. Civil protection work: Absent V22. Early warning system and BPBD office: Near $\leq$ 2.5 km
The least frequent attributes					
<b>TIMING</b>		Timing to access shelter: $\geq$ 170 minutes Timing to access Health centres: 45 min.	Timing to access shelter: 10 min. Timing to access Health centres: 25 min.	Timing to access shelter: 10 min. Timing to access Health centres: 45 min.	Timing to access Health centres: 25 to 45 min. Timing to access shelter: 10 to 25 min.
	<b>RESPONSE</b>	V20. Evacuation shelter, evacuation sign and storage facility: None V19. Access road: Dirt road V22. Early warning system and BPBD office: None	V19 Access road: Paved road or Trail	V22. Early warning system and BPBD office: Near $\leq$ 2.5 km V20. Evacuation shelter, evacuation sign and storage facility: Yes $\leq$ 5 km V23. Health Centres: Dispensary $\leq$ 5 km V19 Access road: Trail	V19 Access road: Paved or Dirt road

Distance/timing criterion according to access ways and means of transport. The distance/time criterion displays three situations: the darker grey is, the longer it takes to access any given remote block.

Distance, in km from block to shelter or health facilities	Timing, in minutes according to access way type and means of transport		
	Paved road (common car)	Dirt Road (4x4 vehicle or truck)	Trail (walk)
2.5	5	15	40
5	10	25	<b>80</b>
10	25	<b>45</b>	<b>170</b>

Reference	Yokoyama et al., 1984	Ewert et al., 2005; Ewert, 2007	Wood & Soulard, 2009	Kinvig et al., 2010	Aspinall et al., 2011	Aucker et al., 2015	Brown et al., 2015	Scandone et al., 2016	Mangan et al., 2018; Ewert et al., 2018	Del Negro et al., 2020	Nieto-Torres et al., 2021	<i>This work</i>
Objective	Threat	Threat	Exposure of communities facing lahars	Threat	Risk	Risk	Threat (country level)	VolcanoRisk Coefficient VRC	VTA Volcanic Threat Assessment	Risk Lava flow AHP	VRR Volcano Risk Ranking	Human exposure at local scale
Method and indicator	$\Sigma(H+E)$	H*E	Composite index for lahar-prone zones	NVEWS Score system Nisyros Island and USA	H*PEI	VHI* PEI	Mean VHI* # volcanoes* pop30	$K_T$ eruption between 0 & 1 year ago; VEI; Log popul.	Surveys	GIS, AHP Analytic Hierarchy Process	VRR(1)=H*E*V VRR(2)=(H*E*V) / (res+1)	EIPN: field survey, mapping & statistics
Scale	Global		Mt Rainier		Global risk for volcanic countries	Global	Global			Etna volcano	Global	Local
Exposure	<b>Number of parameters evaluated for computing the Exposure factor</b>											
	7	9	6	10	1= PEI	1= PEI	1= Total popul. $\leq$ 30 km from active volcanoes (pop30) by country	1= Log number of popul.	9	1= density of population community	9	Initial: 11 Tested:8 Final: 6
Population exposure	<b>Number and/or nature of parameters specifically evaluated for computing the Population Exposure Index</b>											
	2: Number of popul.; Density	3= popul. in hazard zones $\leq$ 30 km radius; Past fatalities and evacuations; aircrafts	3= Residents; Employees; Dependent-population facilities	4= Log10 volcano popul. $\leq$ 30 km and > 30 km; prior fatalities, prior evacuations	2= Number of popul. per distance radii (10, 30 and 100 km); Number of fatal events	1 = same as Aspinall et al.	1= Number of popul. per distance radii	1 = Log number of popul.	4= Popul. <30 and >30 km; visitors number; prior fatalities; prior evacuations	1= Density normalized by the value of the more populated municipality	1= Population density within 5, 10, 30 and 100 km distance radius from the main crater	8 tested= V1, V2, V3, V5, V6, V7, V8, V9 Final= 6 based on selected PLR model: V1, V2, V3, V5, V6, V7

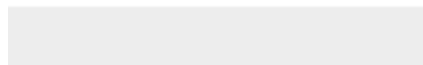
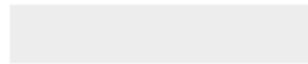
Table 8

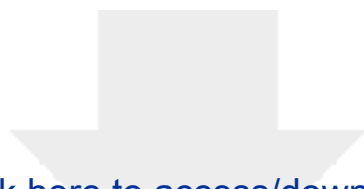




Click here to access/download

**Electronic Supplementary Material**  
ESD Tables 1 to 9.pdf





Click here to access/download  
**Electronic Supplementary Material**  
ESD Figures 1 to 4.pdf

