Supporting Information for

A Comprehensive Socioeconomic Vulnerability Analysis Using Robust DEA Technique at the Finest Resolution of Sub-District Scale in Entire Maharashtra State of India: Highlighting Significant Vulnerability Drivers

Isha Dev[a], Ankan Chakraborty[b], Subhankar Karmakar[a, b, *]

[a] *Environmental Science and Engineering Department, Indian Institute of Technology Bombay, Mumbai 400076, India*

[b] *Centre for Climate Studies, Indian Institute of Technology Bombay, Mumbai 400076, India*

[*]Corresponding Author at Environmental Science and Engineering Department, Indian Institute of Technology Bombay, Mumbai 400076, India

E-mail address of the corresponding author: skarmakar@iitb.ac.in (Subhankar Karmakar)

Supplementary Tables

Table S1. List of derived composite socioeconomic vulnerability indicators considered in the study based on the sensitive scenario of vulnerability assessment

Indicators	Justification	References	Influence to Vulnerability
Percentage of Non-Working Population wrt Total Population (%)	They are dependent adult members in their respective families, and they will be more vulnerable during a disaster due to their low coping ability	Yoon (2012); Sherly et al. (2015); Vittal et al. (2020)	Increases vulnerability (sensitive/positive)
Percentage of Non-Working Female Population wrt Total Female Population (%)	The coping capacity will be very lower in a family during hazards if the female belongs to a non-working population due to financial issues and family liabilities	Wood et al. (2010); Sherly et al. (2015)	Increases vulnerability (sensitive/positive)
Percentage of Marginal Working Population wrt Total Population (%)	They hold temporary jobs and are mostly landless labourers and do not have financial securities	Sherly et al. (2015)	Increases vulnerability (sensitive/positive)
Percentage of Marginal Female Working Population wrt Total Female Population (%)	They hold temporary jobs and are mostly landless labourers and don’t have financial securities and also they are more liable to families	Sharma et al. (2020)	Increases vulnerability (sensitive/positive)
Percentage of HH Latrine Facility NA wrt Total HH (%)	During disasters such as floods, if the households do not have proper drainage and latrines, then the possibility of diseases will increase mainly due to unhygienic conditions	Vittal et al. (2020)	Increases vulnerability (sensitive/positive)
Percentage of HH Electricity Facility NA wrt Total HH (%)	The household with electricity may have accessible information to the early warning systems and can withstand the impacts of a disaster	Vittal et al. (2020)	Increases vulnerability (sensitive/positive)
Percentage of Female Population wrt Total Population (%)	An increase in the female population increases difficulties in evacuation during hazards, as they feel more responsible towards family and children	Schmidtlein et al. (2011); Khan (2012); Lee (2014); Sherly et al. (2015)	Increases vulnerability (sensitive/positive)
Percentage of Rural Population wrt Total Population (%)	Usually poorer and with a higher percentage of farmers; thereby having direct adverse effect during a disaster	World Bank (2002); Shewmake (2008); Fekete (2009)	Increases vulnerability (sensitive/positive)
Percentage of Illiterate Population wrt Total Population (%)	Illiterates tend to have a smaller set of employable skills and have reduced access to information with a low level of risk assessment; this increases their vulnerability	Wongbusarakum and Loper (2011); Sherly et al. (2015); Dumenu and Obeng (2016)	Increases vulnerability (sensitive/positive)

Indicators	Justification	References	Influence to Vulnerability
Percentage of Population below 6 Years wrt Total Population (%)	They require special attention during evacuation due to their high care needs and susceptibility to health problems due to immaturity; thereby slowing down the evacuation processes during disaster situation	Guillard-Gonçalves et al. (2015); Kotzee and Reyers (2016)	Increases vulnerability (sensitive/positive)
Percentage of Female Illiterate Population wrt Total Female Population (%)	Illiterate females may find it more difficult to follow any evacuation warning and care for family during a disaster	Sherly et al. (2015), Vittal et al. (2020)	Increases vulnerability (sensitive/positive)
Percentage of HH with 7+ Members wrt Total HH (%)	They need special and quick attention because a large number of members in a single family will slow the process	Blaikie et al. (1994), Morrow (1999)	Increases vulnerability (sensitive/positive)
Percentage of Bad HH wrt Total HH (%)	They may face tremendous problems during a disaster event because the household conditions are bad	Vittal et al. (2020)	Increases vulnerability (sensitive/positive)
Percentage of SC and ST Population wrt Total Population (%)	They are categorised as backward communities by the government of India and tend to be weaker economic and social sections	Sherly et al. (2015), Vittal et al. (2020)	Increases vulnerability (sensitive/positive)
Percentage of Main Cultivator and Main Agricultural Population wrt Total Population (%)	They are connected to agriculture and are usually poor and face a direct adverse effect during a disaster event. Their coping ability will also be less, as they need to find alternate jobs to fulfil their financial needs	World Bank (2002), Sherly et al. (2015), Vittal et al. (2020)	Increases vulnerability (sensitive/positive)
Percentage of HH with Drinking Water Facility Away wrt Total HH (%)	If the availability of drinking water is away from the facility, then the residents may have a huge impact during the disaster event since readily water is not available for them	Vittal et al. (2020)	Increases vulnerability (sensitive/positive)
Percentage of Temporary HH wrt Total HH (%)	They may face tremendous problems during a disaster event as these are made for temporary purposes and with inferior quality materials	Brooks et al. (2005)	Increases vulnerability (sensitive/positive)
Percentage of HH with No Drainage Facility wrt Total HH (%)	During disasters such as floods, if the households do not have proper drainage, then the possibility of diseases will increase mainly due to unhygienic conditions	Vittal et al. (2020)	Increases vulnerability (sensitive/positive)

Percentage of Rented HH wrt Total HH (%)	People who rent do so because they are either transient or do not have the financial resources for home ownership. They often lack access to information about financial aid during recovery. In the most extreme cases, renters lack sufficient shelter options when lodging becomes uninhabitable or too costly to afford.	Morrow (1999); Cutter et al. (2003)	Increases vulnerability (sensitive/positive)
Percentage of HH with Banking Facility wrt Total HH (%)	Those people can have the extra facility of having flood insurance and other monetary relief from the government directly during any disaster	[-]	Decreases vulnerability (adaptive/negative)
Percentage of HH with Communication Facility wrt Total HH (%)	HH with communication facility will have extra facility for the people to get aware about the early warning about the upcoming natural disasters	[-]	Decreases vulnerability (adaptive/negative)

Supplementary Figures

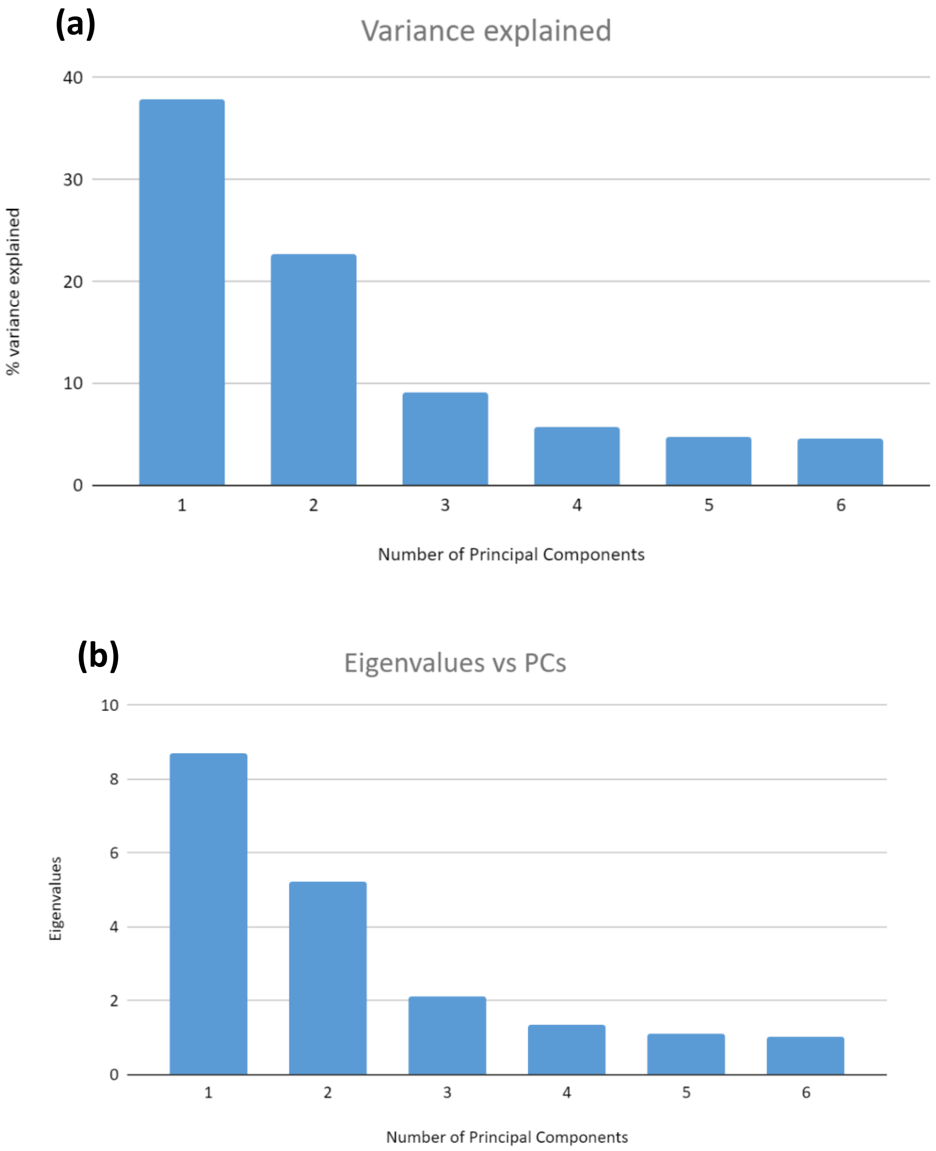

Figure S1. Results of PCA applied to the derived composite socioeconomic vulnerability indicators obtained from the 2011 census sub-district level demographic data: (a) Principal Component (PC) vs Variance and (b) Principal Component (PC) vs Eigenvalue.

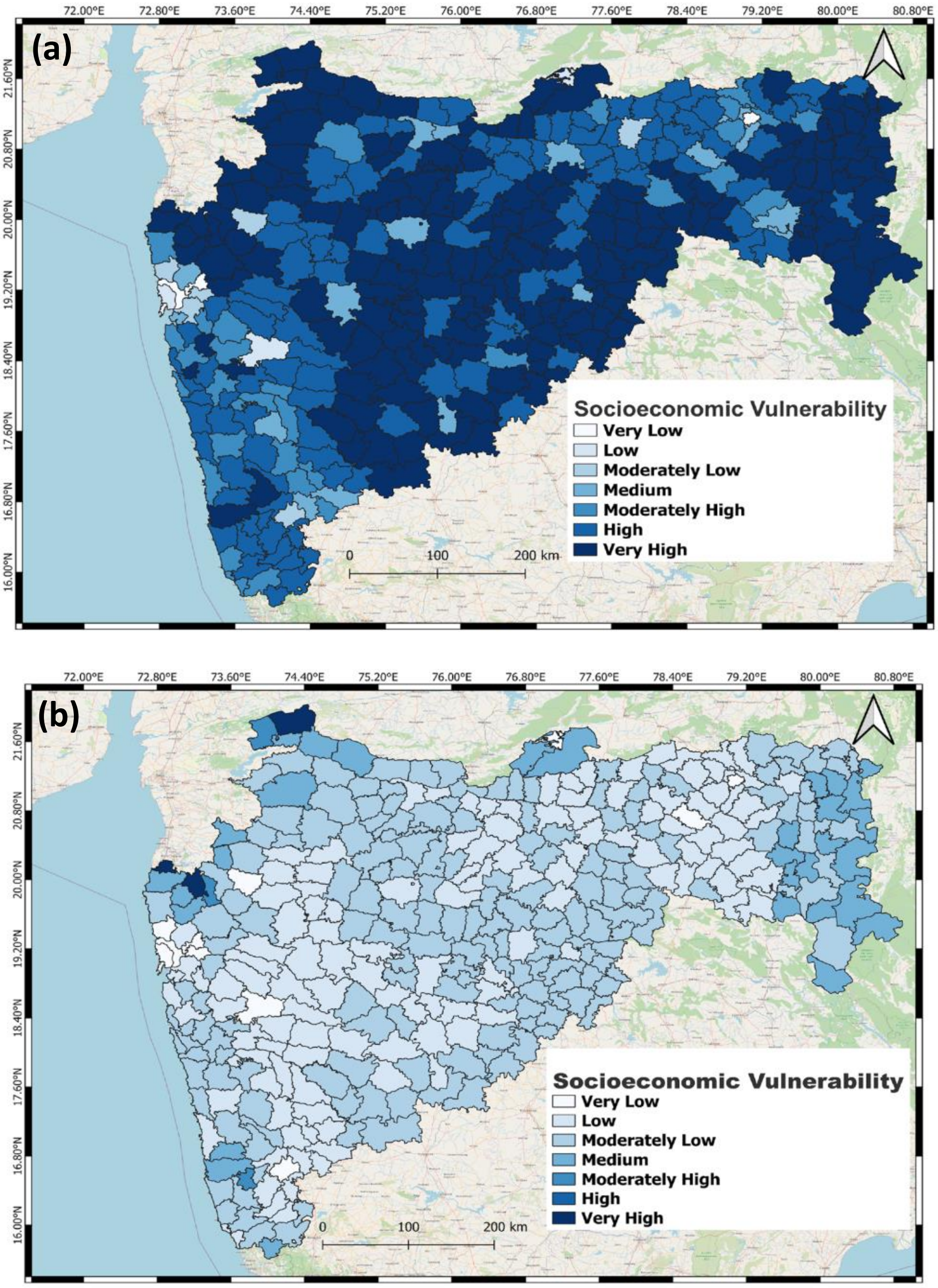

Figure S2. Socioeconomic vulnerability was computed using different methods for the 2011 census sub-district level demographic data across the entire Maharashtra state: (a) PC1-based method (i.e., considering only the first principal component) and (b) Simple average method.

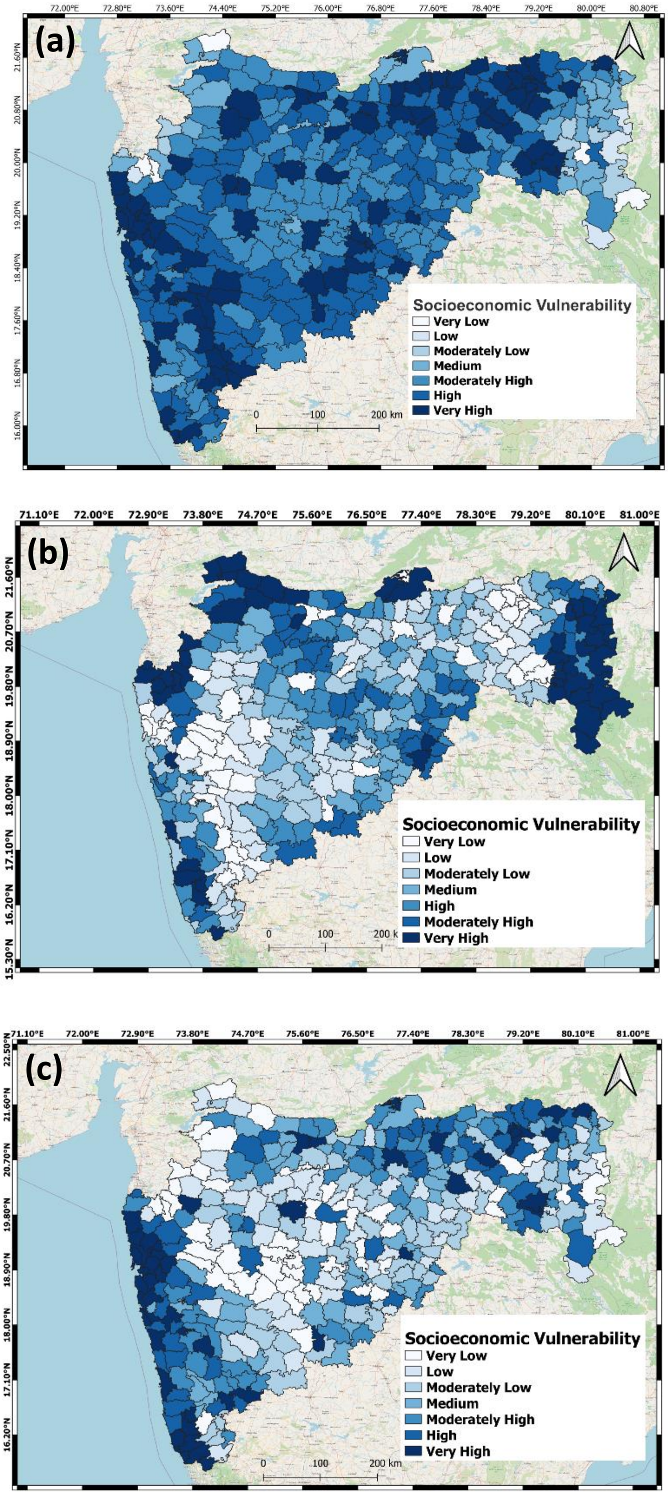

Figure S3. Socioeconomic vulnerability was computed using different methods, considering only the significant socioeconomic indicators derived from the factor analysis results on the sub-district level demographic data obtained from the Census of India, 2011 for the entire Maharashtra state: (a) Only PC1-based method, (b) Simple average method and (c) DEA (BCC model).

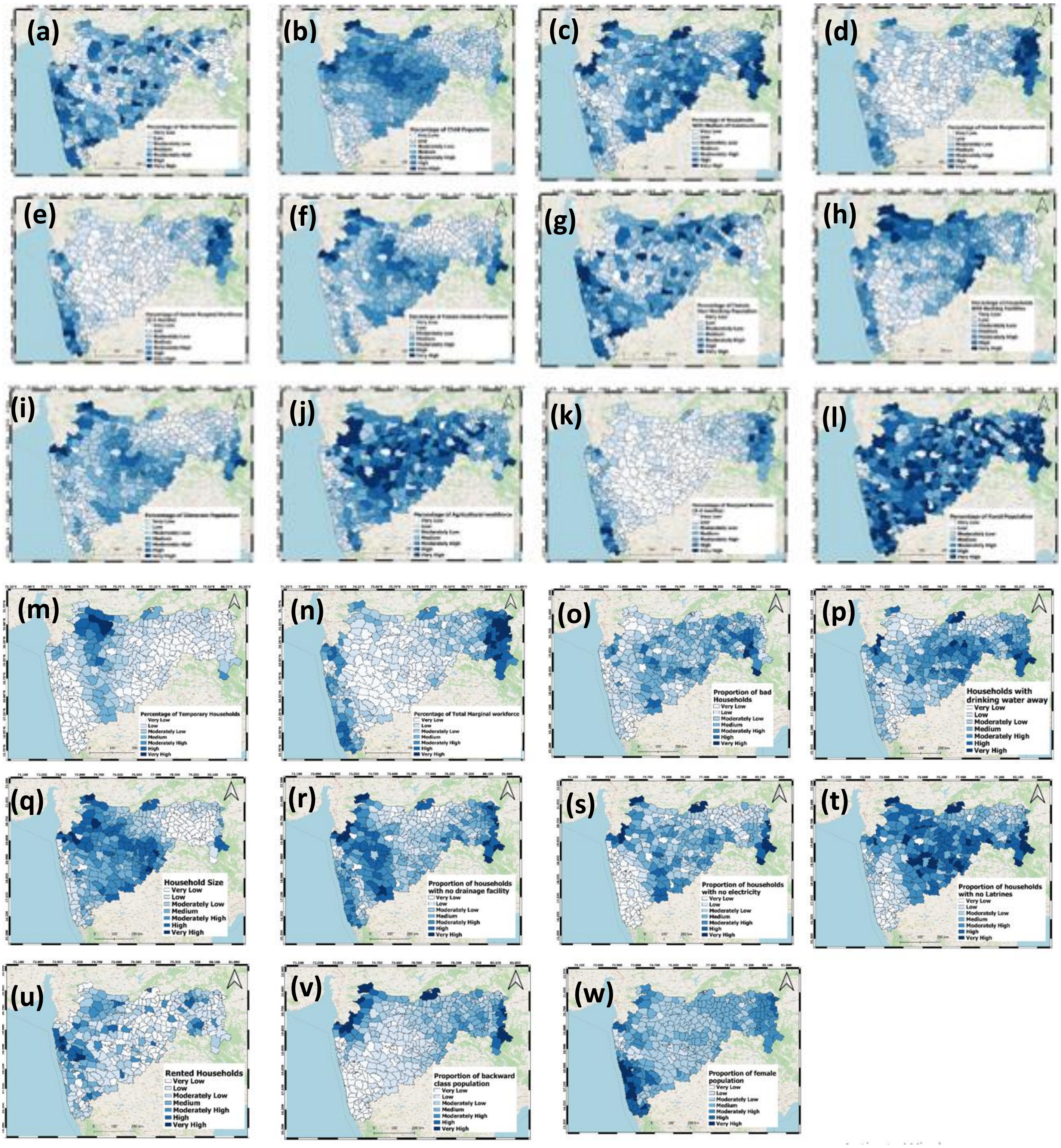

Figure S4. Spatial distribution of the selected socioeconomic vulnerability indicators for sensitive scenarios computed on the sub-district level demographic data procured from Census of India, 2011, for the entire Maharashtra state showing the normalized values for sub-districts across all 23 indicators: (a) Percentage of Non-working population; (b) Percentage of child population; (c) Percentage of households with communication

facility; (d) Percentage of female marginal workforce; (e) Percentage of female marginal workforce for 0 to 3 months; (f) Percentage of the female illiterate population; (g) Percentage of female non-working population; (h) Percentage of households with banking facilities; (i) Percentage of the illiterate population; (j) Percentage of the agricultural workforce; (k) Percentage of the marginal workforce; (l) Percentage of rural population; (m) Percentage of temporary households; (n) Percentage of the total marginal workforce; (o) Proportion of bad households; (p) Households with drinking water facilities away; (q) Household size; (r) Proportion of households with no drainage facilities; (s) Proportion of households with no electricity; (t) Proportion of households with no latrines; (u) Rented households; (v) Proportion of backward class population; and (w) Proportion of the female population.

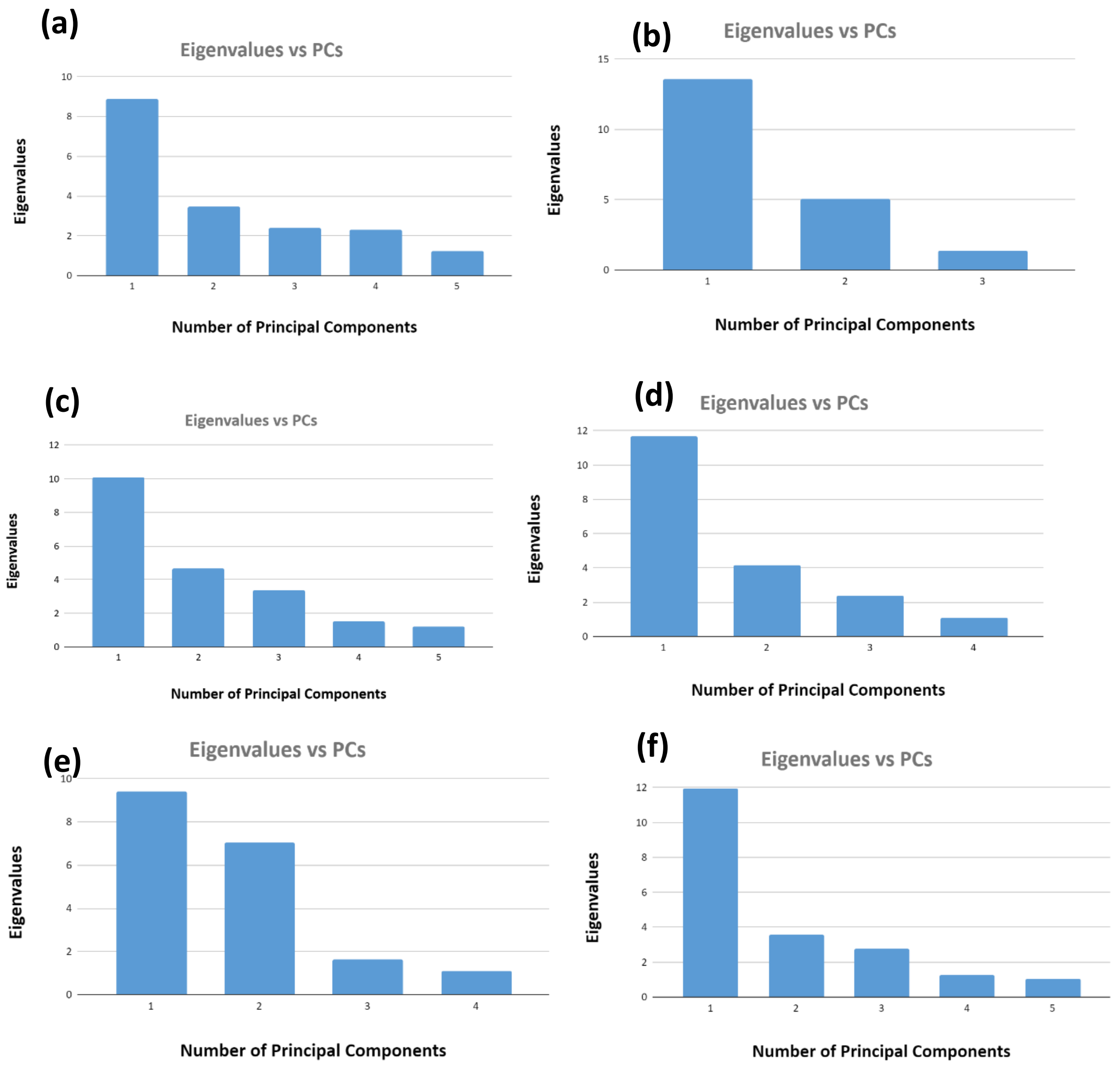

Figure S5. PCA results of the selected socioeconomic vulnerability indicators for the sensitive scenario computed on the sub-district level demographic data obtained from the Census of India, 2011, for the entire Maharashtra state across 6 different clusters ((a) to (f)), based on administrative revenue zones (Total 23 composite indicators for each cluster).

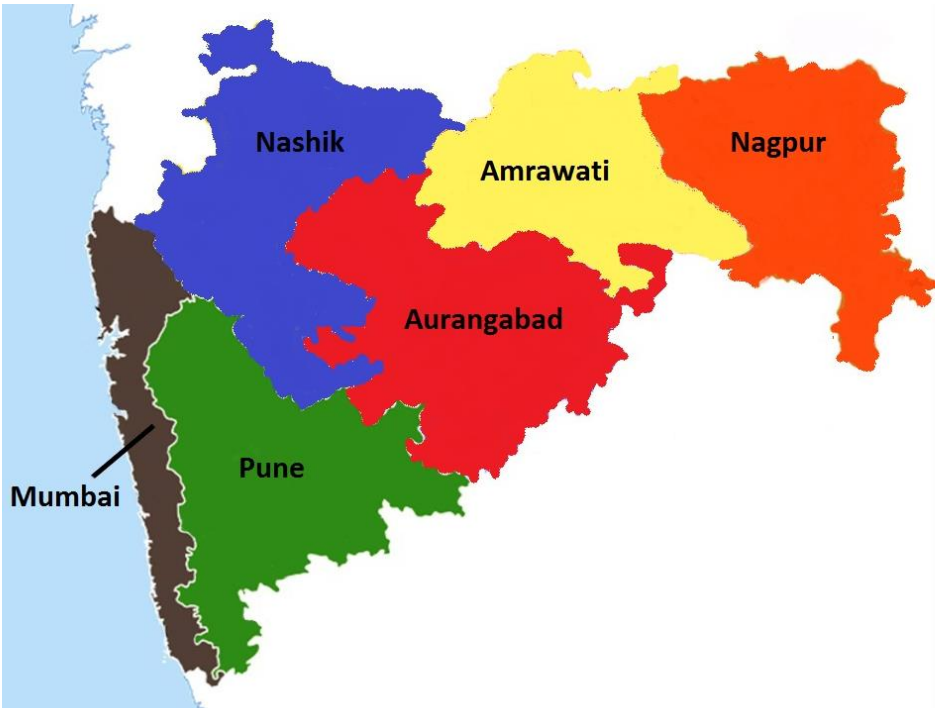

Figure S6. Entire Maharashtra state with 6 different clusters based on administrative revenue zones

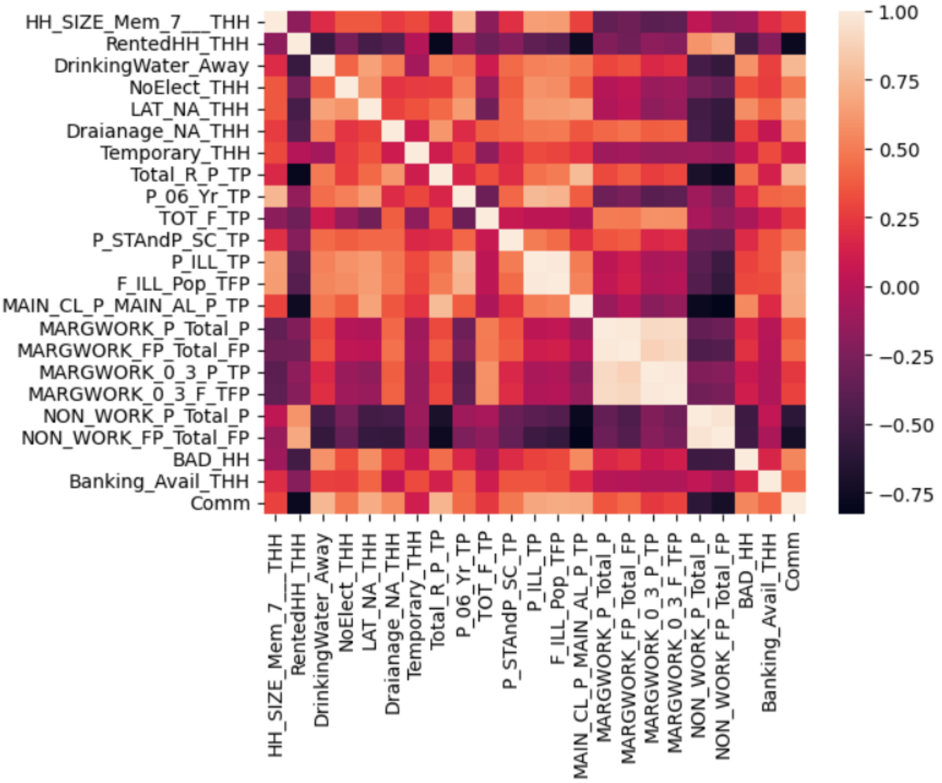

Figure S7. Spearman correlation coefficients of derived composite indicators considering socioeconomic vulnerability for census year 2001 sub-district level demographic data for entire Maharashtra state (23 x 23 dimensional square matrix)

Summary of DEA

In order to determine the relative efficiency of units based on numerous inputs and multiple outputs, Charnes et al. established the Data Envelopment Analysis (DEA) methodology in 1978. According to the study, efficiency is the weighted sum of the inputs divided by the weighted sum of the outputs. The performance of a group of peer entities known as Decision Making Units (DMUs), which transform many inputs into multiple outputs, is assessed using this relatively new, data-oriented methodology. At first, DEA is a methodology that focuses on boundaries as opposed to central tendencies. One floats a piecewise linear surface to rest on top of the observations rather than attempting to fit a regression plane through the middle of the data, as in statistical regression. As a result, DEA proves adept at uncovering relationships that remain hidden from other methodologies. Since the initial development of DEA, numerous models have been introduced to enhance its performance and applicability. DEA models can be categorised using various terminologies, one of which distinguishes between radial and non-radial approaches. The classical DEA models provided the foundation for subsequent variations and remain integral to understanding DEA. The three primary classical models are CCR, BCC, and SE. This study employs the BCC model of DEA.

For accurate efficiency assessment, the number of DMUs must be greater than or equal to the number of inputs and outputs, and the variables should exhibit low correlation. A high correlation among variables can reduce the model's capacity to effectively calculate efficiency. Saein and Saen (2012) applied DEA to assess seismic vulnerability in the twentieth district of Tehran, defining vulnerability as the state of a system prior to a hazard. Their analysis introduced a dummy output with a unity value for all DMUs. Similarly, our study adopts this approach, considering an output of unity value, as also utilised by Saein and Saen (2012) and Sherly et al. (2015). The Banker-Charnes-Cooper (BCC) model (Banker et al., 1984) is implemented in this study due to its widespread popularity among DEA models. The slack-based, input-oriented BCC model incorporates variable returns to scale (VRS) and is expressed as:

$$Minimise\ [\theta - \varepsilon(\textstyle\sum_{\{i=1\}}^{m} S_i^- + \sum_{\{r=1\}}^{s} S_r^+\ ) \tag{1}$$

$$\textstyle\sum_{\{j=1\}}^{n} \lambda_j X_{ij} + S_i^- = \theta X_{ip}\ ,\ i = 1, \dots, m \tag{1a}$$

$$\textstyle\sum_{\{j=1\}}^{n} \lambda_j y_{rj} - S_r^+ = y_{rp}\ ,\ r = 1, \dots, s \tag{1b}$$

$$\textstyle\sum_{\{j=1\}}^{n} \lambda_j = 1 \tag{1c}$$

$$\lambda_j, \varepsilon \geq 0 \tag{1d}$$

where p is the district being evaluated; Θ ($0 < \Theta \leq 1$) is the technical efficiency of the district; y_{rj} is the amount of output r provided by district j; x_{ij} is the amount of input i used by district j; λ_j is the weight assigned to district j; $S_i -$ and S_r^+ are the slack and remnant variables respectively; and $\mathcal{E}$ is the non-Archimedean infinitesimal, generally assigned $\mathcal{E} = 10^6$

Summary of Factor Analysis

Factor analysis is a robust statistical technique that allows researchers to identify a set of latent factors meaningfully and parsimoniously representing a group of observed indicators. This process relies on researchers' judgment and interpretation to make sense of the identified factors (Goretzko et al., 2021; Hair et al., 2019; Howard, 2016; Watkins, 2018). It estimates the number of latent factors underlying the observed indicators and measures the strength of association between each indicator and the latent factors, known as factor loadings. The study's second objective was to identify the major drivers of socioeconomic vulnerability—factor analysis was performed to evaluate the contribution of key indicators to significant principal components (PCs). The significance of an observation within a component was determined using the ratio of the observation's squared factor loading to the eigenvalue associated with that component. This ratio represents the observation's contribution to the component (Williams, 2010).

This variance-based method evaluates the contribution of individual indicators only within the significant PCs (in this case, six PCs). The analysis identified key indicators that significantly influence socioeconomic vulnerability across the state of Maharashtra. These indicators highlight areas where targeted policies addressing social and economic disparities could substantially improve conditions for vulnerable populations.

The percentage contribution of an indicator (i) is calculated as:

$$ctr_{i,l} = \frac{f_{i,l}^2}{\sum f_{i,l}^2} = \frac{f_{i,l}^2}{\lambda_l} \qquad (2)$$

Where $ctr_{i,l}$ refers to the contribution of observation to the component. Formally, the contribution of observation i to component l is denoted by$ctr_{i,l}$. Similarly, $f_{i,l}^2$ represents the square factor loadings of ith observation to lth component and **λ** represents the eigenvalue of the l_{th} component. The contribution value ranges between 0 and 1. For a specific component, the total contributions of all observations sum up to 1. A higher contribution value indicates that the observation contributes more significantly to the component.

References

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*(9), 1078-1092. https://doi.org/10.1287/mnsc.30.9.1078

Goretzko, D., Reiser, M., & Bühner, M. (2021). The impact of the number of items on the factor structure of a scale: A Monte Carlo study. *Psychological Methods*, *26*(2), 247-263. https://doi.org/10.1037/met0000272

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Howard, M. (2016). *The analysis of unidimensional and multidimensional constructs: Theoretical issues and practical applications*. Wiley.

Saein, S., & Saen, R. F. (2012). Seismic vulnerability assessment using Data Envelopment Analysis: The case of the 20th district of Tehran. *Structural Safety*, *34*(1), 23-34. https://doi.org/10.1016/j.strusafe.2011.11.004

Sherly, T., Jain, R., & Joshi, S. (2015). Data Envelopment Analysis for vulnerability assessment: A case study of regional resilience. *Environmental Hazards*, *14*(4), 345-361. https://doi.org/10.1080/17477891.2015.1055815

Watkins, M. W. (2018). *Exploratory factor analysis: A guide to best practice*. Oxford University Press.

Williams, T. A. (2010). The contribution of social factors to disaster vulnerability in the context of hazard risk assessment. *Disasters, 34*(2), 207-225. https://doi.org/10.1111/j.1467-9523.2009.01179.x