

Assessing Biophysical and Social Vulnerability to Natural Hazards in Uttarakhand, India



Colin Doyle
Jonathan Sullivan
Richa Mahtta
Bhartendu Pandey

Report Prepared for Ignacio Urrutia
The World Bank Group

June 30th, 2017

Disclaimer

This work was made possible by the financial contribution of the Water Partnership Program (WPP). The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

The World Bank does not necessarily own each component of the content contained within the work. The World Bank therefore does not warrant that the use of any third-party-owned individual component or part contained in the work will not infringe on the rights of those third parties. The risk of claims resulting from such infringement rests solely with you. If you wish to re-use a component of the work, it is your responsibility to determine whether permission is needed for that re-use and to obtain permission from the copyright owner. Examples of components can include, but are not limited to, tables, figures, or images.

All queries on rights and licenses should be addressed to the Publishing and Knowledge Division, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2625; e-mail: pubrights@worldbank.org.

Contact Information

This paper is available online at <http://www.worldbank.org/water>. Authors may also be contacted through the Water Help Desk at whelpdesk@worldbank.org.

Water Partnership Program <http://water.worldbank.org/water/wpp>.

Rights and Permissions

The material in this work is subject to copyright. Because The World Bank encourages dissemination of its knowledge, this work may be reproduced, in whole or in part, for noncommercial purposes as long as full attribution to this work is given.

Any queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2422; e-mail: pubrights@worldbank.org.

© 2017 The World Bank
1818 H Street NW
Washington DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

ABSTRACT

Given the devastating history of flooding and landslides, most notably in 2013, this report aimed to assess the current potential threat that biophysical hazards poses to Uttarakhand from four angles: 1) how geomorphic change in recent floods changes the location of rivers and surrounding floodplains; 2) what is the spatial distribution of biophysical hazards; 3) where is socially vulnerable population concentrated in the State; 4) How are infrastructure, population, and resources exposed to biophysical hazards and social conditions. To do so, the authors built spatial, statistical, and machine learning models in publicly available platforms to assess biophysical and social vulnerabilities. For biophysical vulnerability, the authors generated state-wide and spatially-detailed flood and landslide hazard maps. Validation efforts suggest that the flood hazard model and landslide hazard models will place a randomly chosen incidence location greater than a randomly chosen no incidence location with a frequency of 71% and 90%, respectively. The lower accuracy of the flood hazard assessment model is attributed to the fact that it is generated in the absence of spatially-explicit historical flood information. Furthermore, using the Spatial Database of South Asia and applying a robust principal component analysis-based methodology, the authors generated a social vulnerability index at the sub-district level, which is relevant for local planning. For social vulnerability, this report illuminates two underlying dimensions of the region that are believed to contribute to a region's propensity for loss, namely marginalized populations, and poor access to financial and health resources. Results highlight four sub-districts that have very high biophysical and social vulnerabilities and are located in Dehradun, Haridwar, Tehri Garhwal, and Bageshwar districts. Also, that sub-districts with very high vulnerabilities have 45.57% lower population count and 2.57% lower GDP compared to those with very low vulnerabilities. However, sub-districts with very high vulnerabilities have 31.82% higher road connectivity. Therefore, more investigation is required in how road connectivity is changing biophysical and social vulnerabilities in the State of Uttarakhand. These preliminary results represent a foundation upon which qualitative and quantitative fieldwork, such as focus groups with practitioners in disaster management in Uttarakhand and collection of geospatial data of observations of historic disaster events will be required to fine tune the results and estimate the uncertainty around the predictions reported here.

HIGHLIGHTS

- Using machine-learning techniques applied to remote sensing derived products, a scalable methodology for landslide hazard assessment was developed.
- A first spatially detailed landslide hazard map of the State of Uttarakhand was produced using machine-learning methodology.
- Social Vulnerability was assessed using a PCA-based methodology applied to socio-economic indicators, which were selected based on expert's consultation and literature review.
- An Analytic Hierarchy Process (AHP) based methodology was devised to validate social vulnerability output.
- A risk analysis was carried out considering population, infrastructure, and wealth dimensions at the sub-district level for regions with high biophysical and social vulnerabilities.

TABLE OF CONTENTS

1. INTRODUCTION	5
2. GEOMORPHIC CHANGE IN UTTARAKHAND	7
2.1 Google Earth Engine & Remote Sensing	7
2.2 Geomorphic Change Methodology.....	8
2.3 Results	10
2.4 Discussion	13
2.5 Further Work	14
3. FLOOD HAZARD IN UTTARAKHAND	15
3.1 Methodology	15
3.2 Results	18
3.3 Validation.....	19
3.4 Limitations and Further Work.....	23
4. LANDSLIDE HAZARD IN UTTARAKHAND	25
4.1 Introduction	25
4.2 Methodology	26
4.2.1 Data Collection.....	27
4.2.2 Model Development and Predicting Landslide Hazard	32
4.2.3 Validation	35
4.3 Results	36
4.4 Discussion and Further Work	39
5. SOCIAL VULNERABILITY IN UTTARAKHAND	40
5.1 Introduction	40
5.2 Methodology	44
5.3 Results	47
5.4 Data Limitations.....	51
5.5 Discussion and Further Work	52
6. COMBINED SOCIO-BIOPHYSICAL VULNERABILITY ASSESSMENT	53
7. CONCLUSIONS AND RECOMMENDATIONS	57
REFERENCES	58
APPENDICES	65

1. INTRODUCTION

In June of 2013, a cloudburst flood event in northern India and Nepal resulting in over 6,000 deaths, affecting over 200 villages with crop and property damage, and displacing 75,000 people (Nair and Singh 2014, Mishra 2015). The force of the floods and landslides reshaped the landscape and formed new river pathways that changed the river flow patterns in the region as well as the flood risk distribution. Given the frequency of floods and landslides in Uttarakhand, it is important that flood and landslide hazard models be updated to reflect changes in risk and vulnerability. The typical models, used to assess river geomorphology, hydrology, and lithology are physically based, time-intensive, and costly. The process of building new hydrologic and hydraulic models to reflect geomorphic changes can cost millions of dollars and take years to calibrate and validate. These barriers can prohibit development of hydro-geologic models in a timely manner, especially considering that new assessments are required each time a new major event comes through the area. Streaming satellite imagery and other spatial data sources is one alternative to generate biophysical hazard maps quickly and cheaply for immediate planning and decision-making post flood or landslide disaster, while precise hydro-geologic models are developed. In addition, most biophysical models do not consider the social dimensions of vulnerability, an equally important element for decision-making.

Flooding and landslides in 2012 and 2013, and ones experienced in more recent years, were major disasters for the region, not just because of the physical threat, but also because of the social, economic and political conditions of the people and communities that were affected. Climate change may increase vulnerability in all of these multi-faceted dimensions of biophysical hazards not just in Uttarakhand but elsewhere in India (Kundzewicz et al 2014). The social dimensions of Uttarakhand are rapidly changing with urbanization and climate change, which highlights the need for dynamic vulnerability assessments.

This report combines new Big Data analysis tools with the best available rapid assessment tools in social and physical science to explore the potential to understand and address information gaps about biophysical risk in Uttarakhand. We leveraged the modeling capabilities of Google Earth Engine (GEE) and R, an open-source statistical computing platform, to assess the biophysical and social vulnerability of the region based on satellite remote sensing data. Furthermore, we used state-of-art-tools to assess social vulnerability of the region based on multi-source socio-economic data.

GEE is a geographic data repository coupled with a cloud-computing platform that provides access to the historical library of public satellite imagery and other scientific map products, for example, Hansen et al. (2013), and analytical tools for the development of scientific algorithms. The GEE platform offers unique benefits to vulnerability assessment in flood prone developing countries for three primary reasons: 1) the amount of data it stores and provides access to, 2) its high-volume data processing capability, and 3) the use of a web browser interface. Earth Engine's (EE) data catalog is a multi-petabyte archive of georeferenced datasets essential for disaster assessment and prediction, including images from earth observing satellites (e.g.

Landsat, MODIS) and airborne sensors, weather and climate datasets, digital elevation models, and others. EE can process these high-volume datasets extremely quickly by parallelizing the processing among thousands of central processing units (CPU). Finally, this analytical power is accessible from any computer with a good Internet connection allowing regional to global analyses to be run on even the low-configuration desktop computers, evading the need for expensive software, processing, or data management systems. Lastly, the use of a web browser interface allows users to share data and analyses immediately by sending out a simple browser link. The license to use GEE is currently free for scientific, governmental, and even commercial use.

In addition to the technical strengths, the tool affords impressive communication capabilities. Not only are the resulting maps highly engaging, easy to understand, and interactive, but the results are presented with the generally recognizable Google Maps base layer. Overall the authors argue that platforms like GEE has the power to turn big data on its head, equipping non-experts with data and capacities rather than just extracting and crunching data from people. The localized science and analysis can help individuals understand the climate crisis and take control when preparing and responding to hazards. The platform streams the most recent satellite data collected, and so analysis can be updated with the mere refresh of a browser page – no downloading is required.

In regions undergoing rapid land-use change, like Uttarakhand both through urbanization and biophysical changes, the GEE platform provides a constantly up-to-date data catalog that can provide responsive analysis. This responsive analysis, combined with biophysical and social vulnerability assessments, provides actionable information of where to focus investments in disaster resilience. The alternative of waiting for hydrologic model updates from scientific expert may lead to slow results and an information gap at times when it is sorely needed. This report aims to build the foundations of a tool to assess geomorphic, biophysical, and social dimensions of risk that is flexible enough to include adjustments by local experts with knowledge and context of important variables of flood risk in their region. This tool can also be used to analyze how vulnerability changes over time by running the model over specific years and months, when land use, geomorphology, and human settlement patterns may have shifted.

This report aimed to assess the current potential threat that biophysical hazards poses to Uttarakhand from four aspects:

- new geomorphology of the region's rivers,
- biophysical hazard in the landscape,
- social vulnerability of the population, and
- exposure of infrastructure, populations, and resources to hazards and social conditions.

We ultimately hope to present the tool with local stakeholders in India to get feedback and make adjustments, and more importantly, to train them in how to use it, update it, and improve upon it themselves. Leveraging Big Data to create a tool that enables responsive vulnerability maps can play an important role in shifting disaster mitigation to where it is most needed. However, this

impact is small compared to the potential of empowering local decision makers with access to Big Data and computing in order to tailor their own tool building in ways that we believe have the potential to transform disaster management.

2. GEOMORPHIC CHANGE IN UTTARAKHAND

2.1 Google Earth Engine & Remote Sensing

In June 2013, extreme rains in the Uttarakhand region resulted in significant quantities of water overflowing and breaching dammed lakes. From a geomorphic perspective, the release of such quantities of water can alter the extent of floodplains and can result in avulsions that create new riverine pathways and sediment deposits (Kline & Cahoon, 2010; Syvitski, Overeem, Brakenridge, & Hannon, 2012). In regions like Uttarakhand that experience seasonal flash floods, riverine systems are dynamically changing after each flood event (Pande, 2010). For disaster managers, it is important to understand riverine changes and its relationship with changes in flood or avulsion risk across a watershed. However, land-use/ land-cover (LULC) change analyses are often slow, expensive, and lack additional context for managers to assess the significance of the changing natural environment. New satellite imagery analysis platforms are being built, such as GEE, that present an opportunity for greater streamlining of LULC analyses and more informed disaster management.

The richest source of spatial data stored within GEE for geomorphic assessment is the public satellite data that has been collected over the last 40 years from Landsat missions. The historical Landsat archive is one of the most valuable datasets for measuring and monitoring LULC change across the globe (Cohen, Fiorella, Gray, Helmer, & Anderson, 1998; Coiner, 1980; Coppin & Bauer, 1994; Seto et al., 2002). Access to the full Landsat data catalog is available through the [USGS Global Visualization Viewer](#) (GloVis) where individual tiles can be downloaded for analysis. The process of individually requesting satellite imagery tiles has proven to be a cumbersome exercise that often overloads local servers, inhibits facile sharing of datasets, and restricts interaction with scientific mapping products to academic journals. In this regard, GEE presents a breakthrough by providing a cloud-based data catalog that can be accessed through an application-programming interface (API) that allows processing of petabytes of Landsat data on Google infrastructure. In the GEE API, environment scientific applications and algorithms can be rapidly tested across the globe and across different time-periods, further expanding the LULC change science and providing quick information to decision-makers.

In addition to ease of access, sharing, and publishing the depth of the Landsat catalog provided in GEE, it allows for the rapid analysis of large datasets allowing highly sophisticated methods to be implemented with more ease. For example, a principal challenge to remote sensing has been the ubiquitous presence of clouds that lead to problems with inaccurate atmospheric correction and subsequent mistakes in land cover classification, leading to false detection in LULC change. The International Satellite Cloud Climatology Project-Flux Data (ISCCP-FD) dataset estimates the global annual mean cloud cover is approximately 66% (Zhang, Rossow, Lacis, Oinas, &

Mishchenko, 2004). Therefore, clouds and cloud shadow removal often present an initial step in most satellite imagery analyses (Arvidson, Gasch, & Goward, 2001; Simpson & Stitt, 1998). Typically, the collection of large Landsat datasets is necessary to implement such techniques as creating a cloud-free composite image from Fmask algorithms (Zhu & Woodcock, 2012), Automated Cloud Cover Assessment (ACCA) (Irish, 2000; Irish, Barker, Goward, & Arvidson, 2006), or the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Vermote & Saleous, 2007). With the GEE platform, these analyses can be run globally and over the entire historical Landsat collection of scenes within a few minutes.

With these elements of GEE in mind and the context of a seasonally shifting riverine system due to annual floods in Uttarakhand - the utility of a rapid LULC analysis to inform risk evaluation in a socio-biophysical context becomes clear. The objective of conducting a geomorphic change analysis within GEE is to demonstrate both the robust scientific tools that are supported in GEE and how access to important information can be generated quickly within the GEE platform.

2.2 Geomorphic Change Methodology

Geomorphic change in the floodplains of Uttarakhand was detected using land classification results from two time-periods surrounding the June 2013 flood event. A pre-flood time-period (October 2012 – March 2013) and a post-flood time-period (October 2013 – March 2014) were first used to create cloud-free composites and then apply supervised classification techniques to derive the extent of floodplains in each period.

For the pre-flood time-period, 96 orthorectified Landsat 7 ETM+ scenes were collected and adjusted to top of atmosphere reflectance. Using the internal cloud mask generated by the LEDAPS atmospheric correction tool a cloud mask was created for each individual tile within the Landsat 7 ETM+ collection for the pre-flood period (Vermote & Saleous, 2007). Cloud pixels were then masked from the collection and the median value across the collection was calculated to create a cloud-free composite (Figure 2.1).

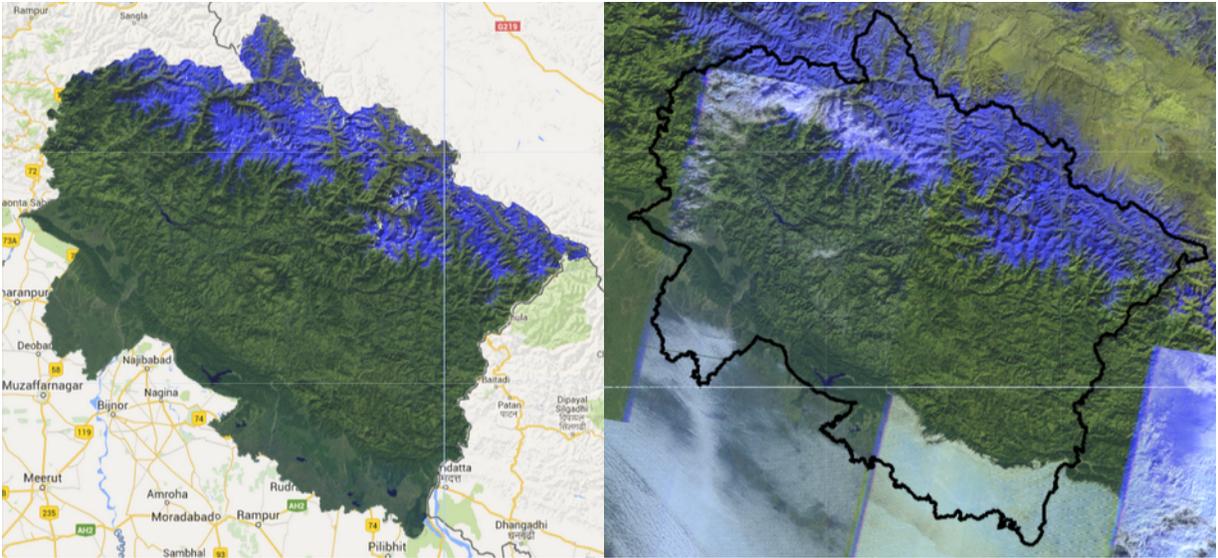


Figure 2.1. Cloud-free composite of Uttarakhand, India (left) using a Landsat 7 ETM+ collection (pre-cloud removal, right) and LEDAPS cloud mask from October 2012 – March 2013. Images are shown as false color composites in 7-5-2 band configuration through R-G-B channels.

Similarly, a cloud-free composite during the post-flood period was generated using a collection of 99 orthorectified Landsat 8 OLI scenes that were calibrated for top of atmosphere reflectance. Landsat 8 OLI has several new bands compared to Landsat 7 ETM+, namely band 9 (1.36 - 1.38 μm) that is helpful for detecting high altitude clouds and a quality assessment band that assigns a cloud probability to each pixel. Using the quality assessment band a cloud mask was generated and applied to each tile within the Landsat 8 OLI post-flood collection. The median was calculated across the collection for each pixel to produce a cloud-free composite (Figure 2.2).

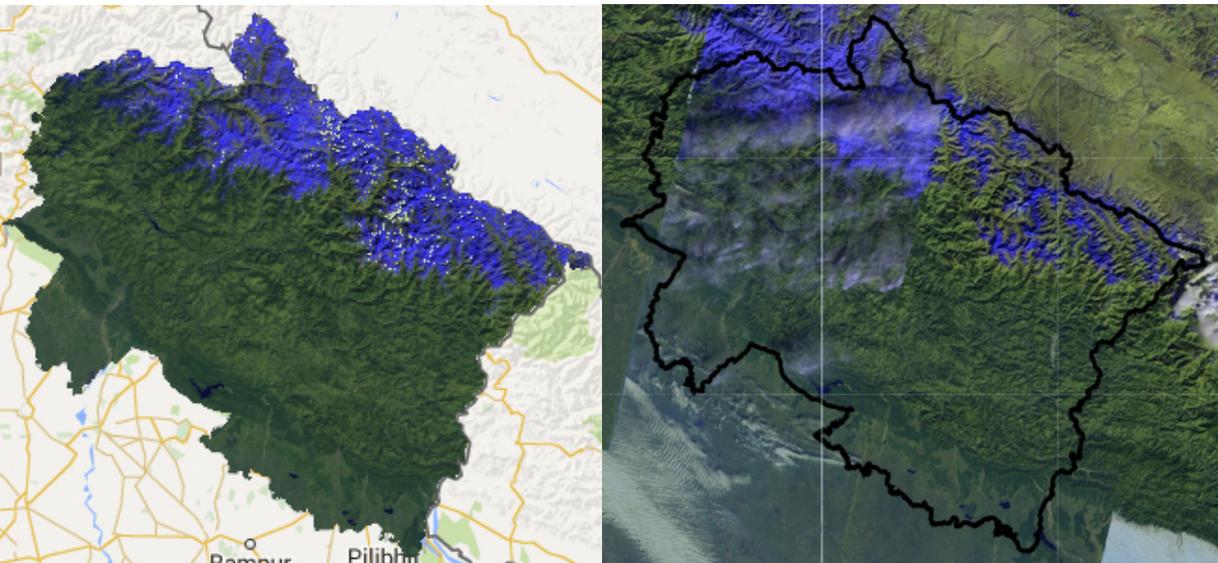


Figure 2.2. Cloud-free composite of Uttarakhand, India (left) using a Landsat 8 OLI collection and the built-in quality assessment band to mask clouds from a Landsat 8 OLI collection between October 2013 – March 2014 (pre cloud-removal right). Images are shown as false color

composites in 7-6-3 band configuration through R-G-B channels. Supervised classification methods were used to produce a map of six land classes: urban, sediment, water, snow, open field, and forest. Training data for these six classes for both periods (pre and post 2013 floods) was collected using the cloud-free composite generated and Google Earth imagery that extends over both periods (see training data in accompanying folder: WBREPORTGEODATA/GeoMorph).

A classifier for each cloud-free composite in the pre- and post-flood time-periods was trained using the random forests classifier (Breiman, 2001) with 10 trees. Several bands were used to generate the classifier including, blue, green, red, near-infrared, shortwave infrared 1, and shortwave infrared 2 bands for both collections. In addition, the Normalized Difference Vegetation Index (NDVI), Normalized Difference Snow Index (NDSI), near-infrared and shortwave infrared band ratio, slope (SRTM), and a distance from river layer (Hydrosheds 15 sec shapefile) were included. The NDVI layer is a common band ratio between red and near infrared used for detecting differences in vegetation cover such as forest versus agriculture. The near-infrared and short-wave infrared band ratio has been demonstrated to be helpful to distinguish bright rock and desert, much like sediment deposits, due to the fact that they tend to exhibit higher reflectance in short-wave infrared than near-infrared (Irish, 2000). Slope was used to help distinguish mountain shadows that are often mistaken for other dark features in images (e.g. water). Lastly, the distance from river variable was used to distinguish between sediment and other barren or rocky terrain found throughout the images.

2.3 Results

Figure 2.3 shows LULC classification results for pre-flood and post-flood time-periods. A closer view of the classification over Rishikesh district for the pre- and post-flood time-periods can be seen in Figure 2.4 and 2.5, respectively. No formal accuracy assessment was conducted due to lack of on-ground data but the classification performed well concerning all the land classes. Several issues were apparent, however, in distinguishing several classes for spectrally similar features in the area. In particular, the sediment class appeared to be classified as urban in some areas due to a similar spectral signature. The Hydrosheds data used to create the distance from river variable is not an accurate representation of waterways in Uttarakhand as it maps the natural flow of water and this variable is not representative in areas where these have been altered, such as due to irrigation.

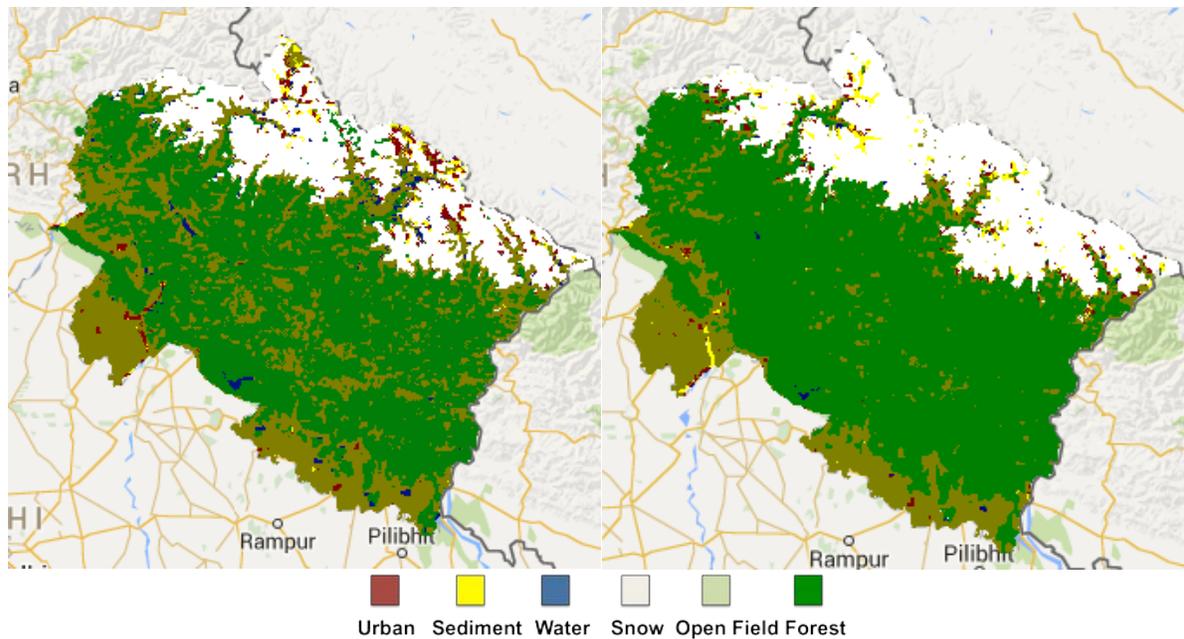


Figure 2.3. Pre-flood (left) and post-flood (right) land use / land cover (LULC) classification images.

The spectral signature for several variables used in the classification was averaged across the training classes (Figure 2.6 and 2.7). From this analysis, the similarity between the urban and sediment classes are evident from the spectral signatures collected. For improvement of the classification between these two classes, non-spectral variables such as more accurate spatial data on river systems within Uttarakhand should be included. Spectral variables that could be included in the future are the tasseled cap indices that have been demonstrated to distinguish “wet” features and may help improve a classification where river sediment may be wetter than urban areas (Crist & Cicone, 1984). Using the land classification developed for both time-periods, the water and sediment classes were combined to map the floodplain of the river system within Uttarakhand. By combining this information from pre- and post-flooding time periods a geomorphic change map is shown that highlight areas within Uttarakhand where new sediments or river courses have been created (green), where no change has occurred (yellow), and where sediment or water courses subsided between the two periods (red) (Figure 2.8).

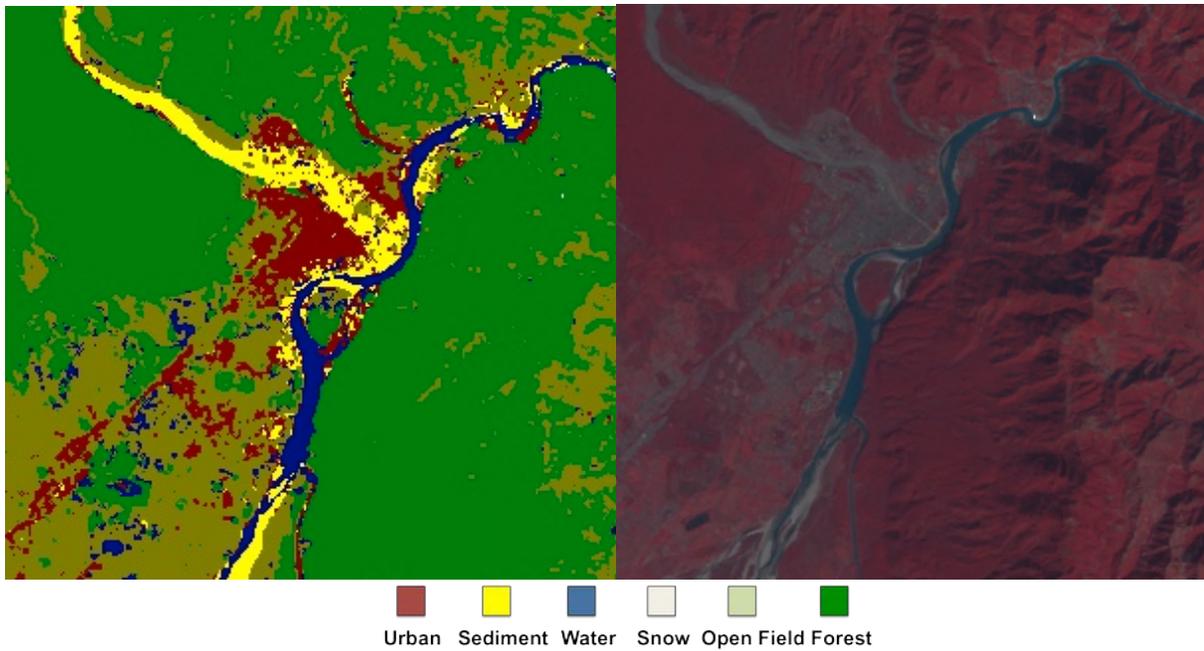


Figure 2.4. The pre-flood classification (left) and Landsat 7 ETM+ cloud-free composite (right) centered over Rishikesh (Image centered 78.3044, 30.1047).

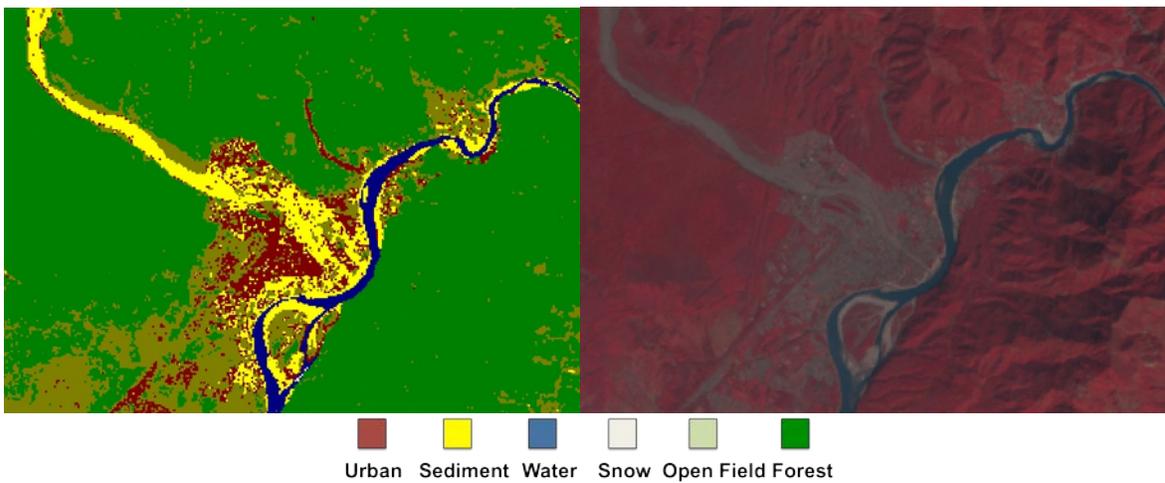


Figure 2.5. The post-flood classification (left) and Landsat 8 OLI cloud-free composite (right) centered over Rishikesh (Image centered 78.3044, 30.1047).

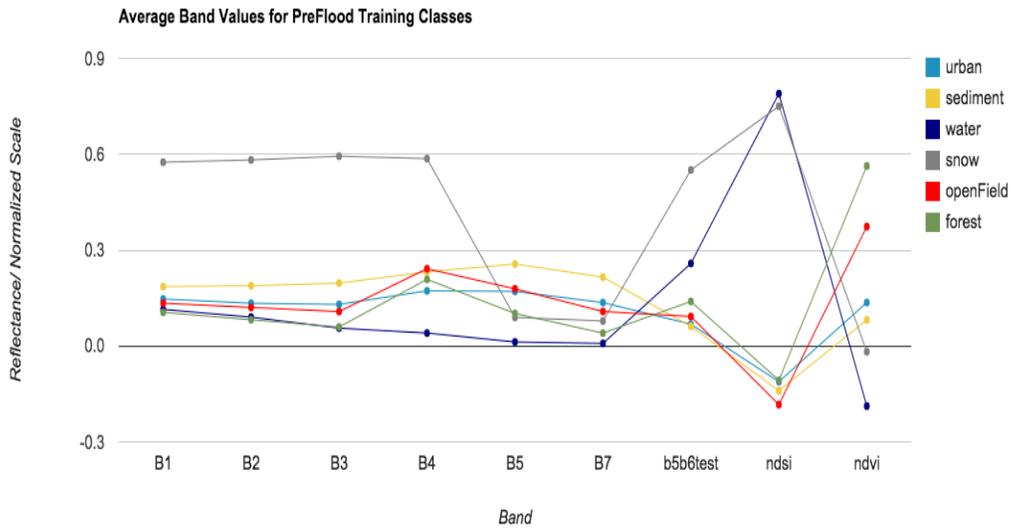


Figure 2.6. Average band values for the pre-flood training classes.

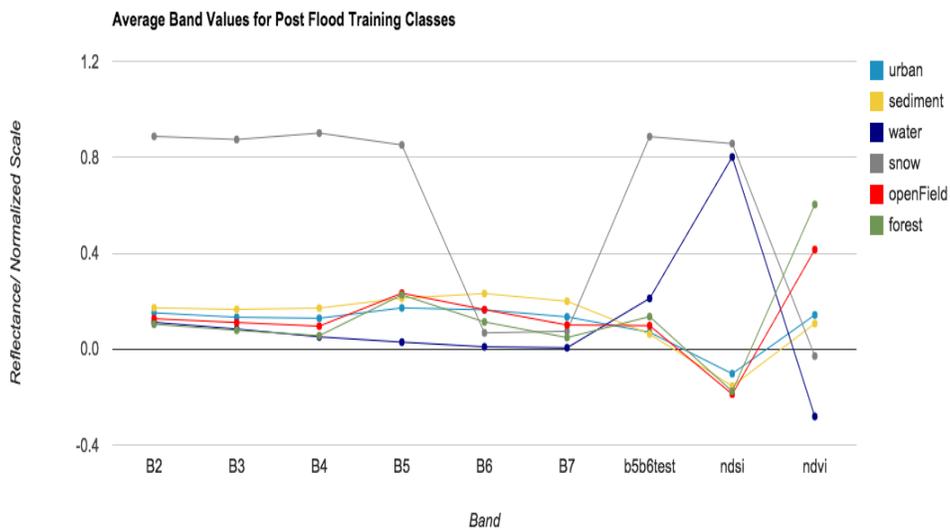


Figure 2.7. Average band values for the pre-flood training classes.

2.4 Discussion

The geomorphic change analysis that was demonstrated above has been developed in GEE's API platform. The significance of using this platform includes the ability to easily share the analysis (see Annex 1 for links to GEE code) and to calculate geomorphic change in different time-periods with only the collection of training data. The collection of training data, however, is something that can additionally be overcome in GEE by using pre-classification techniques where spectral signatures across a time series are measured to detect change. Using techniques

such as principal component analysis can further the automation of LULC analysis. Moreover, GEE continues to bring more sophisticated cloud-masking techniques such as the Fmask algorithm that is able to detect both clouds and cloud shadow (Zhu & Woodcock, 2012). GEE not only provides the scientific sophistication required to make accurate scientific predictions but also provides a platform for a continual feed of updated analyses and the ability to easily share important information with decision-makers.

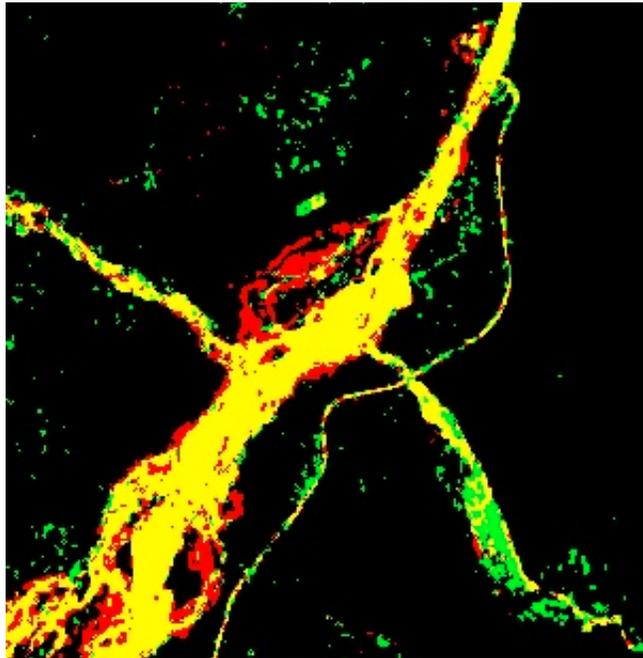


Figure 2.8. Geomorphic change map near Rishikesh showing areas that did not change between the pre- and post-flood periods (yellow), areas where new river courses or sediment deposits have emerged (red), and areas where old river courses or sediment deposits have subsided (green) (Image centered: 78.2566, 30.0456).

2.5 Further Work

There remain areas of improvement for dynamic land classification of Uttarakhand area for the purposes of producing flood-risk maps including the collection of reference data, improved cloud mask composites, and using pre-classification techniques to capture change. In regards to reference data, producing map accuracy estimates is a key component of remote sensing analysis that is lacking in this report. Organization of fieldwork or use of high-resolution imagery (RapidEye, Planet Labs), both present options to collect reference data needed to conduct an accuracy assessment, should be considered in the future. The cloud masking algorithms used in this analysis lack the capability to mask cloud shadows, an additional source of noise in imagery that can influence classification accuracy. Lastly, the methodology used in this analysis follows a post-classification technique that requires training data to define the land classes in both time-periods. Pre-classification techniques provide an additional option that uses spectral signatures across a time series to estimate areas of change and allows for greater automation of remote sensing results.

3. FLOOD HAZARD

3.1 Methodology

The flood hazard index estimates a general floodplain for the state of Uttarakhand. It is important to note that this biophysical hazard prediction is not a physical process based watershed hydrologic and hydraulic model that takes into account “flow” or quantitative precipitation estimates over a surface in real time to compute flood risk with differential equations. Examples of complex flood mapping models with 2- and 3-D differential and numerical equations include LISFLOOD, MIKE-SHE, and TOPKAPI that involve flow routing routines that are not possible in the parallelized computing environment in GEE. The parallelization process sends individual pixels (or potentially groups a region of pixels) in a grid to different parallel servers to process each pixel or image at each step of an algorithm, and then recompiles the new pixels. For hydrologic and hydraulic modeling that requires complex routing algorithms where pixels must communicate with one another in sequential fashion and in multiple dimensions (i.e. the D8 flow accumulation algorithm used to calculate contributing area and topographic index in TOPMODEL) is difficult to calculate with parallel computing (Qin and Zhan 2012). Our model simplifies the assumptions made in physically based hydrologic models, and instead uses map algebra on a pixel-by-pixel basis. This model takes advantage of Google’s cloud computing and is thus tailored towards a rapid assessment to understand dynamic changes to flood vulnerability; it is not meant to replace detailed 3-D hydrologic modeling.

To develop flood indicators for Uttarakhand, we updated and improved our models developed for previous biophysical flood indices in New York State and Senegal based on literature from Uttarakhand (Pande, 2010). For Uttarakhand-specific flood index, we removed several indicators from our original model, namely low elevation and low slope, as the relatively large gradient of topography in Uttarakhand led to excessive flood exposure in the lower extents of the valley only. New indices that capture the unique dynamic of flash flooding events and glacial lake outburst floods that occur at very high elevations and in small to medium sized rivers were added. This required adding 4 additional variables to our risk metrics (variables 3-6) shown in Table 3.1.

Each variable in the index (Table 1) was computed in GEE using datasets already available in the platform with the exception of glaciers, which we uploaded to Earth Engine via Fusion Tables, and the HAND index, which had to be calculated separately in ArcGIS.

The flood vulnerability index is time dependent via two main variables calculating impervious surface/low vegetation and estimating floodplains after the most recent extreme event. Therefore, the flood vulnerability was calculated after the major cloudburst events of 2013 in order to understand current vulnerability based on the most recent land use and geomorphology. While this basic risk model does not predict avulsions, this component could be added in a second iteration of the model.

Table 3.1. Flood Conditioning Factors Matrix

S. No.	Factor	Source	Unit of Analysis & Scale	Method	Classification Threshold	Literature References
1	Topographic Wetness Index (TWI)	HydroSHEDS flow accumulation, and slope derived from SRTM	Pixel, 15 arc-second, meters	$TWI = \ln(a/\tan\beta)$ where β is local slope, a is the upslope catchment area	>50= 5 50-35= 4 35-25=3 25-10-1 <1= 0	(Beven & Kirkby, 1979; Regmi, Giardino, & Vitek, 2010; Sørensen, Zinko, & Seibert, 2006; Tehrany, Pradhan, & Jebur, 2014a)
2	Impervious Surface and low vegetation (%)	Landsat 8 cloud free composite for 2014, built in Earth Engine	30m	NDVI (Normalized Vegetation Index) $\frac{(NIR-RED)}{(NIR+RED)}$		(Xian et al., 2011)
3	Distance from River (m)	HydroSHEDS river network (15 sec shapefile)	Raster, meters	Upload to fusion tables calculate Euclidean distance in earth engine	<10= 5 10-150= 4 150-400=3 400-1000-1 >2000= 0	(Tehrany, Pradhan, & Jebur, 2013, 2014b)
4	Height Above Nearest Drainage (HAND) (m) or Position Above the River	HydroSHEDS Hydrologically Conditioned DEM,	Pixel, meters	Simple pre-processing in ArcGis riparian topography toolbox	0-6= 5 6-15= 4 15-30=3 30-40-1 >40= 0	(Nobre et al., 2011)
5	Floodplain post 2013 floods	Landsat 8 Classification	30m	Supervised classification in Earth Engine	Sediment=5 Non-sediment=0	-
6	Glacial Flood Risk Index	GLIMS glacier database, TWI	Raster derived from map algebra	$GFRI = TWI/distance\ to\ nearest\ glacier$	>1000= 5 1000-500= 4 500-300=3 300-100-1 <50= 0	-

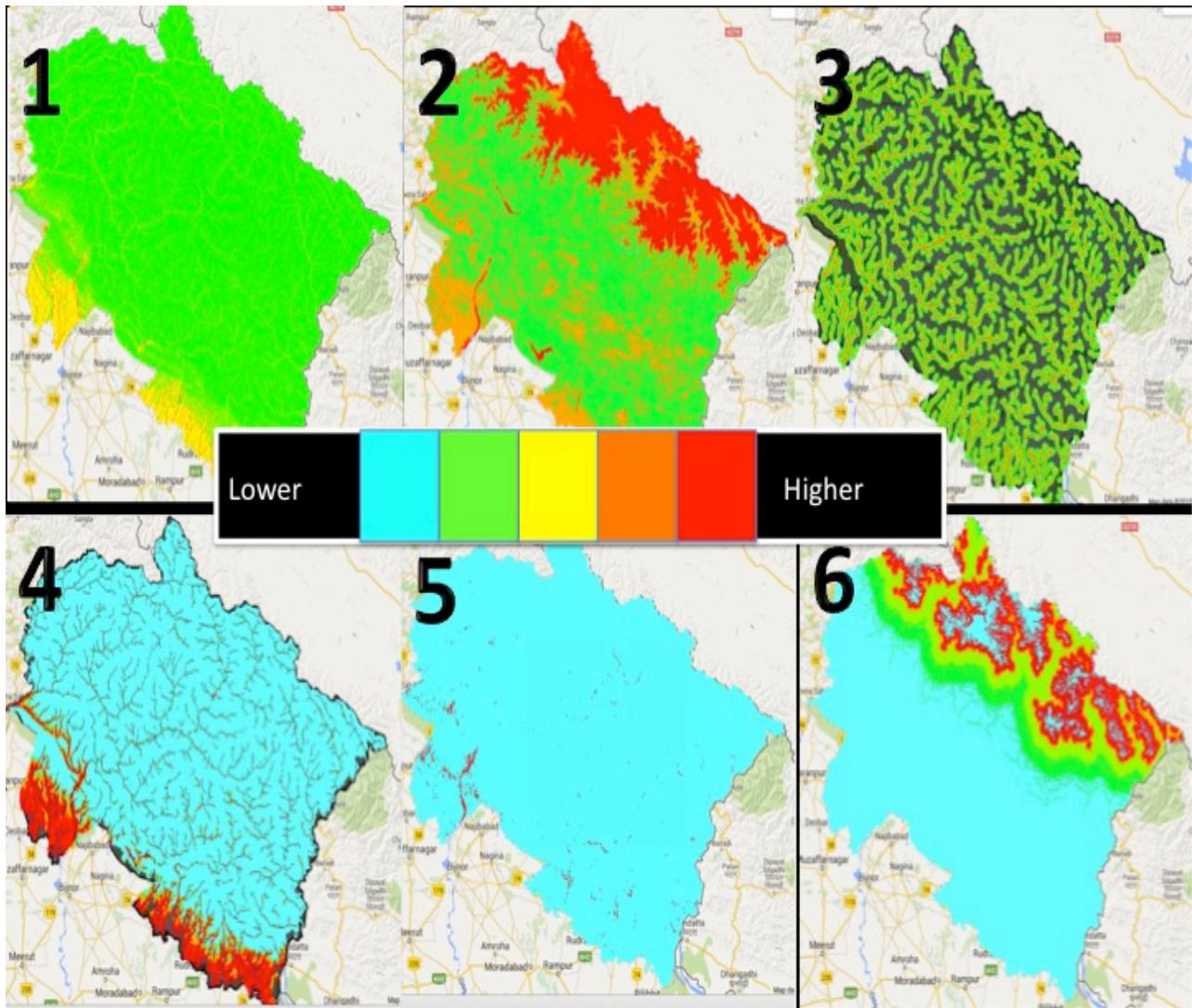


Figure 3.1. Map Depictions of Variables Used to Calculate Flood Risk

Figure 3.1 shows different types of flooding; overbank riverine flooding, flash floods, and glacial lake outbursts can cause floods in different parts of the state. This multi-dimensional index adds together various factors that might cause distinct kinds of flooding in order to create a comprehensive risk map. Each flood-risk factor is classified according to determined thresholds (see Table 3.1) and added together to create an index that could theoretically range from 0-30. The thresholds were identified from the literature and from the experience of the analyst, but could likely be improved by working with local hydrologists and geomorphologists in the region to tailor each threshold to local conditions. Thresholds could alternatively be set by spatial regression if geospatial locations of observed flooding are provided.

After the index is created and factors are added, the next important step is to mask out the areas that are already considered permanent water (rivers and lakes) or glaciers. This is important because the final risk map is determined by setting a threshold on the flood index based on the distribution of the input data. The input data should include all “floodable” areas, which would not include permanent water features. Table 3.2 shows the three data layers used to mask out existing water and ice.

Table 3.2. Masks to remove areas that are permanent water or ice

	Mask	Source	Unit of analysis	Method	Literature Reference
1	Glaciers	GLIMS glacier database	Vector, derived from 15-30m satellite imagery	Used to mask out existing glaciers	(Raup, Kääb, et al., 2007; Raup, Racoviteanu, et al., 2007)
2	Land Water Mask	MOD44W derived from MODIS and STRM	Pixel, 250 m resolution	Used to mask out existing water	(Sun, Yu, & Goldberg, 2011)
3	Water mask	Landsat 8 classification	30m	Supervised classification in Earth Engine	-

Finally, after permanent water and ice areas are masked out of the flood hazard map, the top 10% of the pixels with the highest score are used to classify the “flood zone.” For this region, the index threshold that represented the top 10% of values was 16, so all pixels with a value > 16 were included in the flood zone. The idea is to capture only the highest risk areas that could potentially flood in an intense rainfall event. Again, the thresholds for how much rainfall is required to flood a certain area can only be determined if that model can be calibrated with local flood extent and depth maps, which are not currently available.

3.2 Results

The flood hazard map, from which ranges from 4-26 is shown below in Figure 3.2. We estimated that approximately 3000 km² of the land surface in Uttarakhand is susceptible to flooding. This represents about 5% of the total land surface in the State. We identified the regions with the highest land area and populations exposed to flooding according to the flooding index we developed. These results are reported in Table 3.3. Our index estimated that the majority of flood risk exposure to people is in the southeastern portion of the state, near major populations and the rivers flowing into the Ganges basin. The second major risk area is in the northeast, in the valleys between and behind major glacial formations in the north. There is a third area of high population exposed to flooding in the south west, due to the more gently slope plains and hills that descend from the mountains in this regions. While this makes for fertile agriculture and attracts human settlements, its biophysical characteristics make it prone to flooding.

Table 3.3. High Flood Exposed Regions

	Top Regions with Flood Exposure by Land	Top Regions with Flood Exposure by Population
1	Joshimath	Dehradun
2	Bhatwari	Roorkee
3	Munsiari	Haridwar
4	Dharchula	Rishikesh
5	Roorkee	Vikasnagar
6	Haridwar	Laksar
7	Laksar	Joshimath
8	Mori	Kichha
9	Kapkot	Sitarganj
10	Sitarganj	Ramnagar

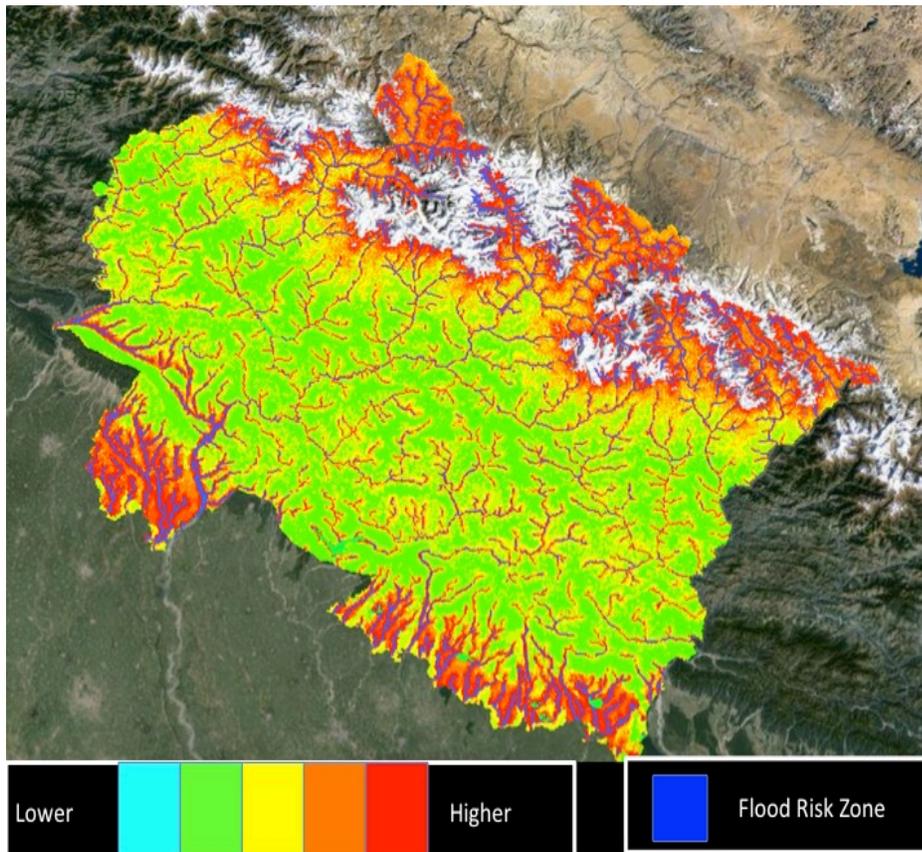


Figure 3.2. Biophysical flood index and flood area estimate

3.3 Validation

Table 3.4 Validation data of locations of past flood events

	Data	Source	Unit of analysis	Method	Reference
1	Place impacted	Google Crisis Maps-taken from local media reports	Points of village names	Ensure Villages are part of the flood map	http://google.org/crisismap/2013-uttarakhand-floods?gl=in
2	Dartmouth Colorado Flood Observatory Database		List of flood dates, number of affected people and fatalities	Gathering of media reports	(Sun et al., 2011)
3	Cloud Burst Events in the Himalayas Since 1970	-	-	-	(Asthana & Asthana, 2014)
4	Flood Inundation Layer (2013 Flood Event)	Bhuvan Geoportal	Random flood locations	-	http://bhuvan.nrsc.gov.in/

In India, data for biophysical hazards and disaster events are rarely available. In fact, Rautela (2016) emphasized lack of scientific recordkeeping of disaster incidences as the biggest hurdle in reducing disaster risk in India. Nonetheless, recent efforts have initiated compilation of disaster-related database. One example is the Bhuvan geoportal, which is a geoportal of the Indian Space Research Organization (ISRO). Information regarding disaster events, including, cyclones, earthquakes, and floods, is provided by Bhuvan disaster support services. In the case of flooding, regional information is provided in the form of recent floods, historic floods, flood annual layers, hazard zones, and aggregated flood. However, data is currently available for a few states. For the state of Uttarakhand, information is available only for one flood event, i.e., one that occurred on June 2013. Using this single flood-event information, we attempted to validate biophysical flood index layer generated from our flood hazard analysis. Specifically, random flood and non-flood locations were selected using a flood inundation map available in the Bhuvan geoportal (Figure 3.3). Overall, 31 flood locations and 56 non-flood locations were randomly selected and geographic coordinates were noted from the Bhuvan platform. We then analyzed the distribution of biophysical flood index values from these points for flood and non-flood locations (Figure 3.4). Results show that our flood index captured majority of the flood locations as is evidenced by the high kurtosis and positive skew of the distribution (Figure 3.4b). Nonetheless, the distribution of the flood index using non-flood locations has a high mean value (~13.3), emphasizing that our index over-predicts flood areas, but lower kurtosis, ascertaining non-

uniformity in over-prediction. Finally, using sensitivity and specificity analysis, and plotting a relative operating characteristics (ROC) curve, we computed the probability that our biophysical flood index will rank a random flood event higher than a random non-flood event. This probability estimate is the area under the curve from the ROC curves (Figure 3.5). Results show that the probability that our biophysical flood index will rank a random flood event higher than a random non-flood event, given similar characteristics of the 2013 flood event, is ~71%. Overall, this simple validation exercise based on ancillary information of a single flood event underlines the capacity of our index to capture the extent of a flood event.

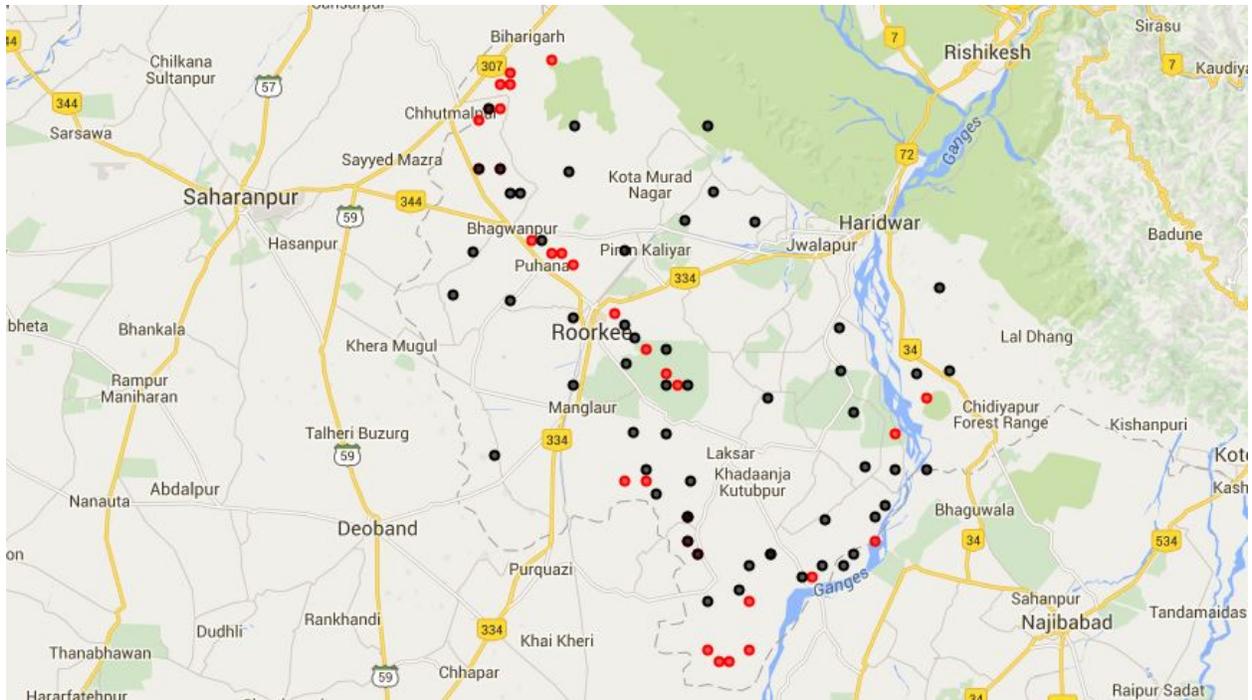


Figure 3.3. Map showing randomly selected flood (red) and non-flood points (black) in the Haridwar-Roorkee region.

We also tried to adjust the index such that affected villages in past events in Table 3.4 were included in the flood vulnerability map. For example, the fifteen villages recorded as having major flood damages in the June 2013 floods by Google Crisis Maps according to media reports, is shown below in Figure 3.6. Thirteen out of the fifteen villages mapped here in pink are part of the flood vulnerability zone, which serves as some validation that this index captured flood vulnerability in the most recent major event in the region.

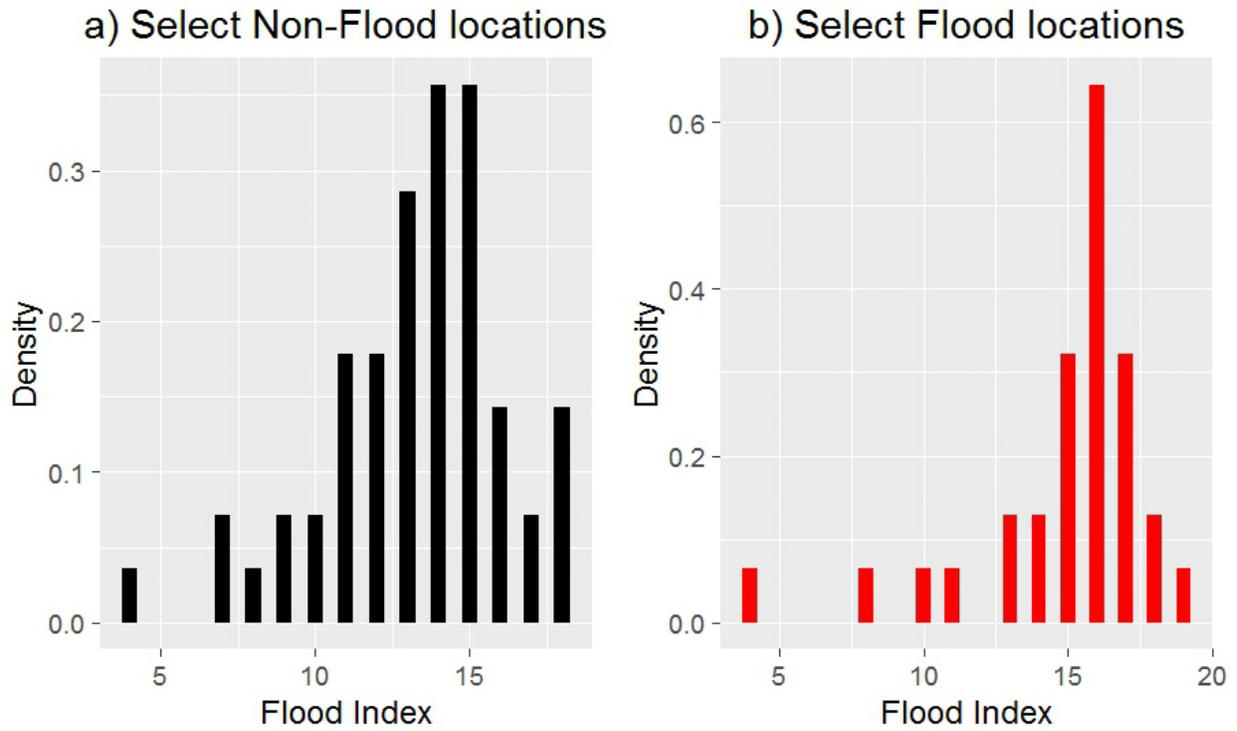


Figure 3.4. Distribution of biophysical flood index values generated using points for a) non-flood and b) flood locations

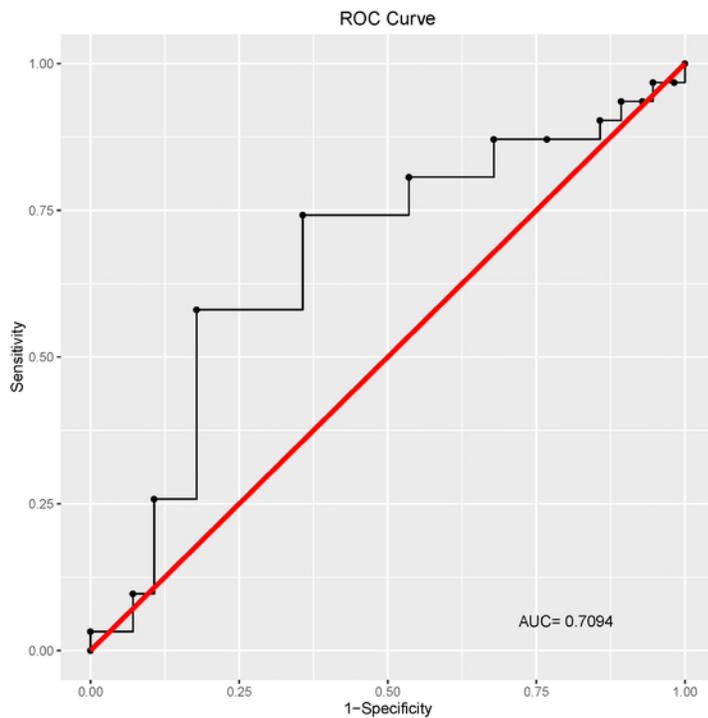


Figure 3.5. Relative operating characteristic (ROC) curve obtained using select flood and non-flood locations of the 2013-flood event and the biophysical flood index.

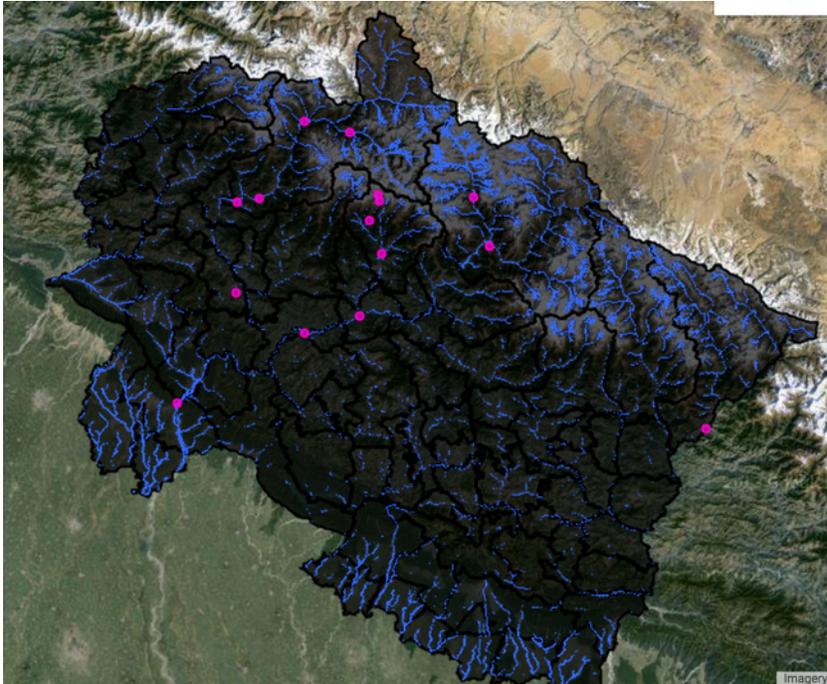


Figure 3.6 Flood index map with locations of villages highly impacted by June 2013 flooding

3.4 Limitations and Further Work

The current flood hazard assessment algorithm is a simplification of major physical processes that play a role in flooding. However, this could be further calibrated and validated with historic flood data to generate a locally-specific algorithm that takes into account not only general watershed topographic characteristics, but also flood control management structures such as levees and dams, and their ability to mitigate flood risk in different storms for each watershed. Not only would this capture the reality of flooding better, this could also result in a more dynamic algorithm that inputs quantitative precipitation data predictions. This approach could make use of the machine learning algorithms in Earth Engine to build predictive flood models. The uncertainty of this type of predictive model would depend on the number and quality of geospatial flood observations. This would require working with local governments to get access to these potential datasets. This kind of modeling would move the flood vulnerability map from being a “general” map of risk to a specific map of risk that relates flood events to rainfall duration and intensity. Training the model on observed flood data and quantifying uncertainty is an important next step to advance this modeling effort.

Additional datasets on infrastructure, in particular location of hydropower plants and roads, could help inform the degree to which flood vulnerability is related to geomorphic change accelerated by this type of infrastructure (Chin, 2006). While there is speculation that dams and road construction has increased flood risk in Uttarakhand (Kala, 2014), a more advanced model may be able to understand the contribution of these changes. Higher resolution would also aid in a better understanding of local over land flow patterns.

We were unable to detect geospatial flood footprints using the MODIS data available in GEE, which is how we typically inventory historical geospatial flood events. This is because the flooding in Uttarakhand happened over such a short time period that there were not enough cloud free images for the dates of the floods we were able to extract from the Colorado Flood Observatory Database (August 2014, June 2013, September 2012, August 2012, September 2010, and August-September 2007). The launch of the new European Sentinel satellite, or access to Indian Space Research Organization (ISRO) data could improve spatial flood mapping and allow for more robust model construction.

Finally, it is important to remember that this floodplain mapping exercise does not consider precipitation patterns in the region. While we have mapped the vulnerability due to the underlying biological and physical nature of the land surface, understanding how this interacts with local precipitation patterns, in particular the propensity for cloudburst systems, is essential. A floodplain only becomes “flooded” when filled with high water volumes from intense rainstorms upstream, which requires further analysis. This index is a first step towards building a more robust and locally determined model.

4. LANDSLIDE HAZARD

4.1 Introduction

Natural processes combined with increased human activity has led to increases in landslides in the Uttarakhand region of the Indian Himalayas (Sati, Sundriyal, Rana, & Dangwal, 2011). High-intensity rainfall and earthquakes form the majority of natural processes that lead to landslides (Martha et al., 2015). These natural processes combined with extensive transportation infrastructure, and settlement expansion has led to increased landslide frequency in the region (Islam, Chatteraj, & Ray, 2014).

Rautela (2015) argues that road accidents due to a lack of acquaintance with motor driving in the mountains surpasses natural disasters as a cause for resultant loss of human lives. However, this notion undermines that how disasters and their causes and consequences unfold are contingent on the coupled human-nature interactions. If lacking road safety measures is a concern, then solely improving it cannot solve the problem of disaster risks in the region. Therefore, it is equally important to consider natural disasters in disaster risk mitigation strategies.

In a recent year, i.e., 2013, an exceptional high-intensity rainfall triggered numerous landslides in the region and caused widespread damage and loss of lives. As per the official estimates, over 4000 people were reportedly dead, at least 13 districts in the state were affected with a net loss of 2163.2 million US dollars in damage to public properties (NIDM, 2015; Rautela, 2015). In Bhagirathi and Alaknanda catchments, Martha et al. (2015) mapped roughly 6000 landslides of which over 50% landslides were new and triggered by extreme rainfall in the month of June. This shows the strength with which natural processes can trigger landslide disasters in the region and result into large-scale devastation and loss of human lives, emphasizing that risks to natural disasters, and landslides in particular, cannot be neglected in disaster preparedness plans and risk mitigation.

Historically, the state of Uttarakhand has been susceptible to landslides and has experienced catastrophic outcomes (Mathew, Babu, Kundu, Kumar, & Pant, 2014). There exist governmental records on these disaster events including, damage data. Nonetheless, these datasets are unsystematic and lack consistency, and often miss local relevance and spatio-temporal details (Rautela, 2016). Satellite remote sensing and spatial analytics offers a promise in overcoming these data challenges (Martha, van Westen, Kerle, Jetten, & Vinod Kumar, 2013; Pardeshi, Autade, & Pardeshi, 2013; Tofani, Segoni, Agostini, Catani, & Casagli, 2013). Therefore, a remote sensing based monitoring framework coupled with advanced spatial analysis could be used to study landslide hazards in the region.

Indeed, several regional studies in the past have utilized remote sensing and spatial data to study landslide hazard in the region. These studies focused on parts of the State to derive landslide susceptibility index. For example, Pham, Pradhan, Tien Bui, Prakash, & Dholakia (2016) used machine learning methods to derive landslide risk map for a region that include regions of Tehri

Garhwal and Pauri Garhwal. Similarly, Martha et al. (2013) and Islam et al. (2014) studied landslide risk in the Rudraprayag district of the State. Several similar case studies exist. However, till date no landslide hazard assessment has been carried out at the State level using remote sensing and spatial analysis tools.

There exist several different approaches to derive landslide hazard map using remote sensing data inputs. These include, (1) Distribution Approach, (2) Bivariate Statistical Analysis, (3) Weights of Evidence Model, (4) Analytical Hierarchical Process, (5) Weighted Overlay, (6) Frequency Ratio Approach, (7) Logistical Regression, (8) Discriminant Analysis, (9) Machine Learning Approach, and (10) others. Pardeshi et al. (2013) provides a critical review of these methodologies. Pradhan (2013), Pham et al. (2016) and other studies show that machine learning methodologies are far more “efficient” compared to experts opinion based and other analytical methods.

In this section, we present a methodology to derive landslide hazard map using multi-source datasets, including satellite-derived products, and spatial analysis. In particular, we apply a widely used machine learning algorithm called support vector machines to develop a non-linear model establishing the relationships between predictor variables and binary landslide outcome variable. We validate the machine learning output with an independently collected landslide presence and absence dataset.

4.2 Methodology

Our methodology comprised of three main steps: (1) Data Collection, (2) Model Development, and (3) Validation (Figure 4.1). In the first step, we collected spatially-explicit data required for machine learning model development. The data included, a binary landslide incidence and no incidence variable and a set of predictor variables.

In the second step, we used the binary outcome variables and predictor variables to develop a SVM-based machine learning model that can accurately predict landslide susceptibility. It is worthy to note here that the machine learning model only generates spatial probabilities of landslide events. There is no temporal information on how sooner or later a location or pixel is likely to experience landslide. In fact, our machine learning model may best be described as a spatial landslide prediction model as opposed to a spatio-temporal landslide prediction model. That said, our model is not driven by dynamic processes that trigger landslides, for example, high-frequency rainfall, earthquake, road construction, and others. Rather our model predicts spatial probabilities of landslide events based on comparatively static predictor variables. In essence, our model tends to establish non-linear relationships between static predictor variables and the binary outcome variable. During model development, we used a rigorous Monte-Carlo Cross-validation approach for selecting optimal number of predictor variables.

In the third step, we validate our model output using an independently collected validation dataset. This data comprises of identified locations of landslide incidence and no incidence. The

landslide incidence locations fall into nine geological regions and the no incidence locations fall into all geological regions. By resampling the validation data and calculating relative operating characteristic (ROC) measures, we estimate the probabilities that our model will rank a randomly chosen landslide incidence higher than a randomly chosen location with no landslide incidence. Lastly, we evaluate how these probabilities vary under difference assumptions of the prevalence of landslide incidence in the region.

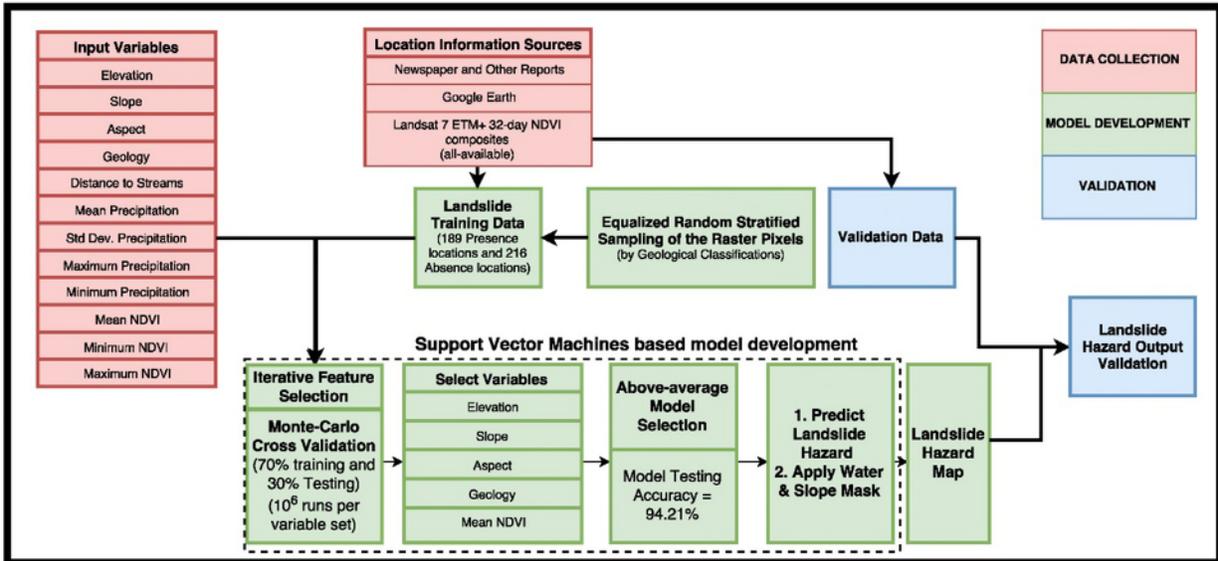


Figure 4.1. Three main steps followed in landslide hazard mapping: (1) Data Collection, (2) Model Development, and (3) Validation.

4.2.1 Data Collection

Binary Outcome: Landslide Incidence and no Incidence

We collected two independent sets of landslide incidence and no incidence data: (1) for model development and (2) for model validation. The data collected for model development essentially includes a small sample of landslides incidence points whose locations were identified by using a combination of data sources such as, newspaper and other reports, journal articles, trend analysis of all-available Landsat 7 ETM+ 32-day NDVI composites¹. After determining the general location coordinates, we use high-resolution time-series satellite images to pinpoint the location

¹ NDVI (Normalized Difference Vegetation Index) captures the level of vegetation activity over a region specified by the image pixel. There is generally an abrupt decline in NDVI over time after landslides occur.

of landslide incidence (Figure 4.2a). From following this approach, we identified 189 point locations of landslide incidence² (Figure 4.2b).

Next, we developed an independent validation dataset in which we randomly sampled 52 locations as polygons that identified landslide incidences in nine geological regions (Figure 4.2c). Additionally, we randomly sampled 272 locations as polygons that captured landslide no-incidences in all geological regions (Figure 4.2d). Overall, the landslide incidence data had 1650 and 1,172,606 pixel locations for landslide incidence and no-incidence, respectively.

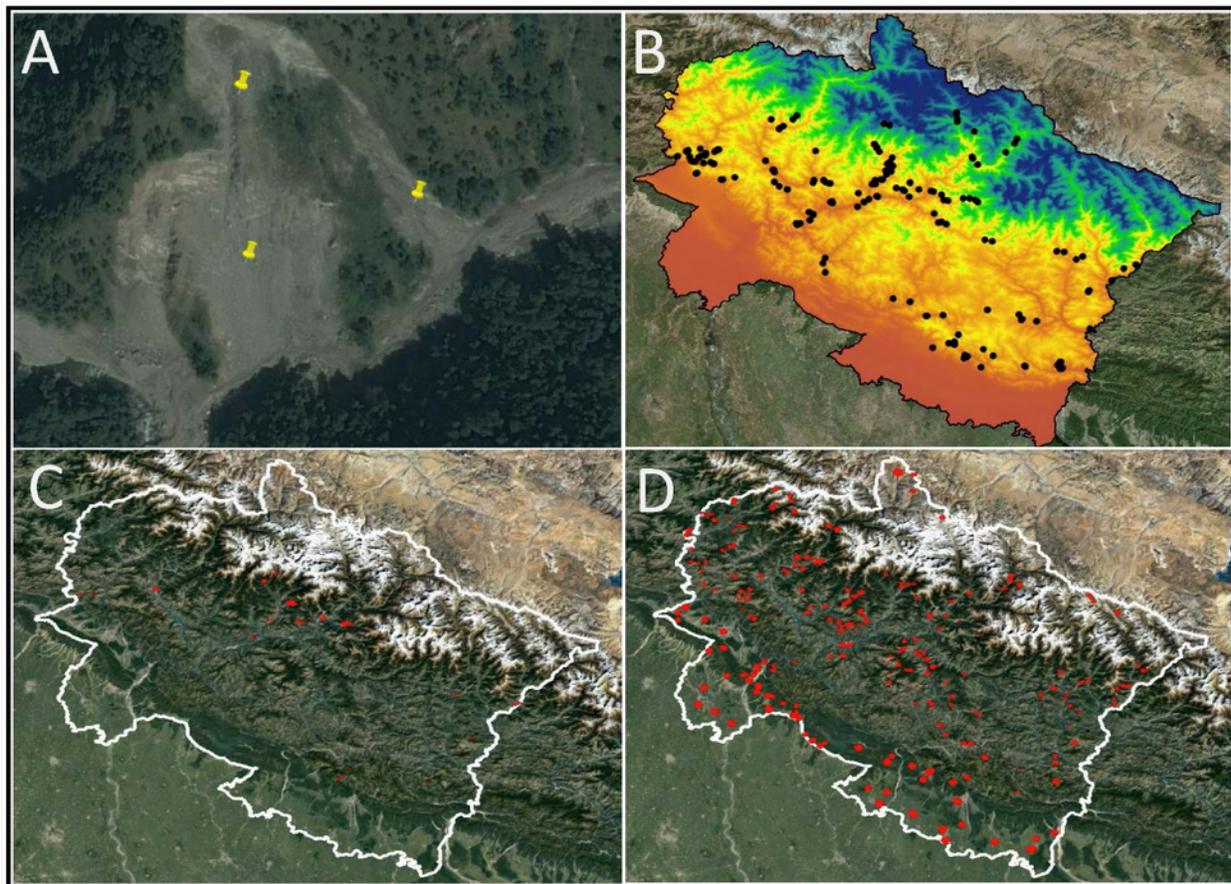


Figure 4.2. (a) Identifying landslide incidence location points using Google Earth Images, (b) Distribution of landslide incidence location points used in machine learning model development, (c) Distribution of landslide incidence locations sampled as polygons for use in model validation, and (d) Distribution of landslide no incidence locations sampled as polygons for use in model validation.

Predictor Variables

² Note that for model development, we collected landslide no incidence locations by randomly sampling pixels assuming that landslide incidence is a rare event. Therefore, no landslide no-incidence points were collected initially to be used in the training data.

The most widely used factors to assess landslide susceptibility are: Slope, Aspect, Soil Type & Geology, and Distance to Roads & Streams variables (Table 4.1). For analyzing landslide susceptibility at a catchment, basin, or watershed levels, the desired level of detail required is higher. This is because more variability in the predictor variables is necessary in order to precisely predict the location of landslides at these finer scales. Consequently, a large number of different topographic variables with detail on soil type, and lithology may be necessary to be considered as predictor variables. However, at coarser scales of analysis, for example, at the State level, a few topographical variables, such as Slope and Aspect, and variables that captures the differentiated nature of soil type, geology, and lithology may provide sufficient variability to identify hotspots of landslides (Pradhan, 2013). Consequently, we included, Slope, Aspect, Geology, and Distance to Streams identifying them as the key variables that influences landslide susceptibility significantly, in our analysis. Besides these four variables, we included four precipitation variables (maximum, minimum, mean, and standard deviation) and three NDVI variables (maximum, minimum, and mean). Therefore, we included twelve variables in our initial selection of predictor variables.

We generated the Elevation variable from Digital Elevation Model (DEM) tiles downloaded from Indian Space Research Organization's (ISRO) open data archive (<http://bhuvan.nrsc.gov.in/data>). Essentially, we used the CartoDEM version 3 product tiles and mosaicked these into a single spatially continuous raster layer. Then we cropped the mosaicked elevation raster using the administrative boundary of the State of Uttarakhand (Figure 4.3). Next, from the elevation raster we calculated Slope (in degrees) and Aspect variables using the *gdaldem* tool (Figure 4.3). In order to calculate distance to streams, we used Elevation raster and derived stream network using the Hydrology tools in ArcGIS™. Finally, we calculated minimum Euclidean Distance from stream network for each pixel. This generated our Distance to Streams variable (Figure 4.3).

We obtained the Geological Regions Map from *Water of Welfare* Initiative's website: (<http://www.ahec.org.in/wfw/intro.htm>). The map was obtained in the *Jpeg* file format and was Geo-referenced using the state administrative boundary. Next, we digitized the Geological map into vector format to distinctly identify the geological regions³. Finally, we converted the vector map into a categorical raster (Figure 4.3).

We calculated statistical summary raster layers using Global Precipitation Data from WorldClim (version 2) available at 1km resolution. This data has average monthly precipitation data for the 1970-2000 period. Using twelve average monthly precipitation raster layers, we calculated four statistical summary measures: mean, standard deviation, maximum and mean (Figure 4.3). Finally, we calculated three statistical summary measures, mean, maximum, and minimum from Landsat 7 ETM+ 32-day NDVI composites (2010–2015), available from Google Earth Engine. All the twelve raster layers thus far generated were then resampled to match the coordinate

³ Note that although the JPEG image file showed thirty-five distinct geological regions, only thirty-one were distinctly identifiable.

system and extent of the Elevation raster layer and a raster data cube was created at an equivalent of 30m spatial resolution.

Table 4.1. Factors commonly used to assess landslide susceptibility

S. No.	Factor	References
1.	Slope	(Ahmed, 2015; Kumar & Anbalagan, 2016; Mathew, Jha, & Rawat, 2009; Pham et al., 2016; Poonam et al., 2017; Pradhan, 2013)
2.	Aspect	(Kumar & Anbalagan, 2016; Mathew et al., 2009; Othman, Naim., M., & S., 2012; Pham et al., 2016; Poonam et al., 2017)
3.	Elevation	(Ahmed, 2015; Othman et al., 2012; Pham et al., 2016; Pradhan, 2013)
4.	Curvature	(Kumar & Anbalagan, 2016; Othman et al., 2012; Pham et al., 2016; Poonam et al., 2017)
5.	Plan Curvature	(Pham et al., 2016; Pradhan, 2013)
6.	Profile Curvature	(Pham et al., 2016)
7.	Soil Type, Geology, Lithology, and Landforms	(Ahmed, 2015; Kumar & Anbalagan, 2016; Mathew et al., 2009; Othman et al., 2012; Pham et al., 2016; Poonam et al., 2017; Pradhan, 2013)
8.	Land Use and Land Cover	(Ahmed, 2015; Kumar & Anbalagan, 2016; Mathew et al., 2009; Pham et al., 2016; Poonam et al., 2017)
9.	Rainfall	(Ahmed, 2015; Othman et al., 2012; Pham et al., 2016)
10.	Distance to Roads / Road Density/Distance to Settlement	(Ahmed, 2015; Kumar & Anbalagan, 2016; Mathew et al., 2009; Othman et al., 2012; Pham et al., 2016; Poonam et al., 2017; Pradhan, 2013)
11.	Distance to Rivers or Streams/ River or Streams Density	(Ahmed, 2015; Kumar & Anbalagan, 2016; Mathew et al., 2009; Othman et al., 2012; Pham et al., 2016; Poonam et al., 2017; Pradhan, 2013)
12.	Topographic Wetness Index	(Kumar & Anbalagan, 2016; Pradhan, 2013)
13.	Lineament/ Lineament Density/ Distance to Lineaments	(Kumar & Anbalagan, 2016; Mathew et al., 2009; Pham et al., 2016; Poonam et al., 2017)
14.	Thrust	(Poonam et al., 2017)
15.	NDVI	(Ahmed, 2015; Pradhan, 2013)
16.	Relative Relief	(Kumar & Anbalagan, 2016; Mathew et al., 2009)
17.	Stream Power Index	(Kumar & Anbalagan, 2016)
18.	Distance to Thrust-Fault	(Mathew et al., 2009)

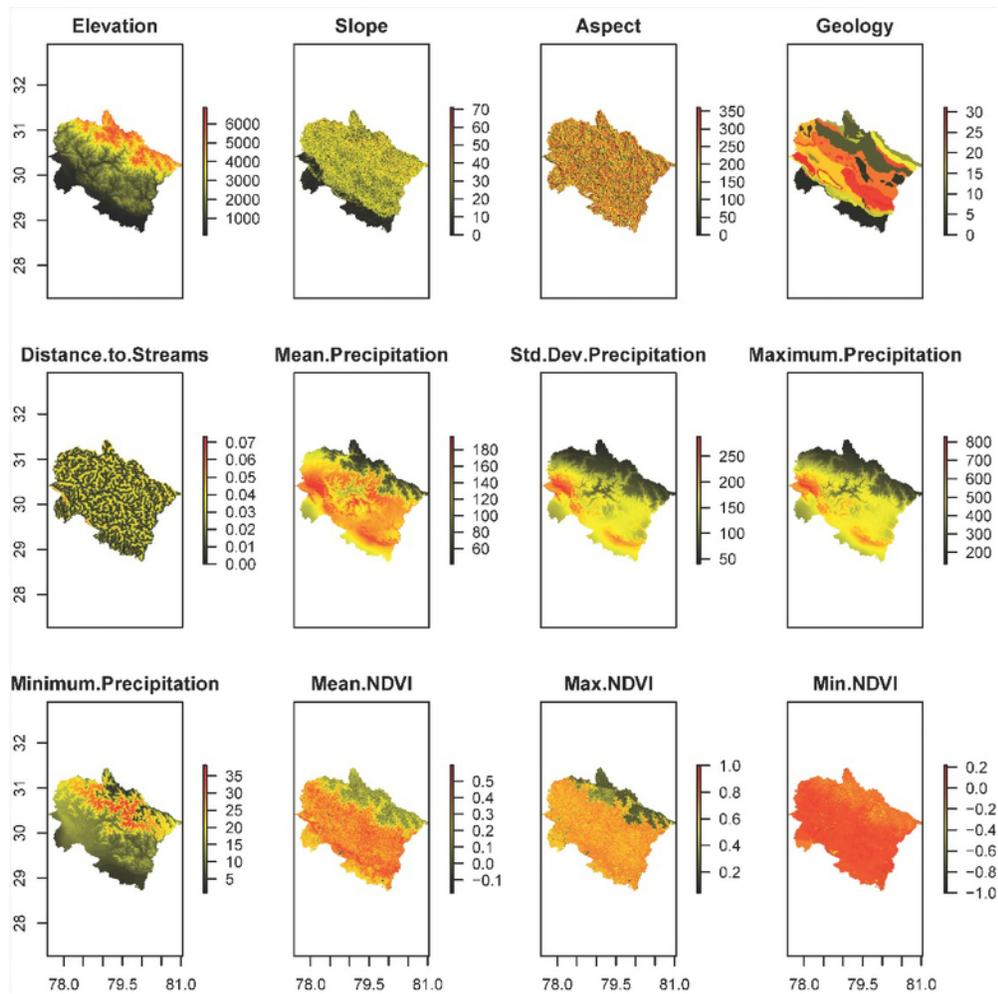


Figure 4.3. Twelve input predictor variables used in machine learning model development.

4.2.2 Model Development and Predicting Landslide Hazard

A key decision to make in a machine learning-based model development is selecting a suitable machine learning algorithm. Indeed, machine learning-based studies often rely on more than one algorithm to build a predictive model. In the case of landslide susceptibility modeling, studies have compared different machine learning algorithms but their results show inconclusive evidence as to which machine learning algorithm consistently outperforms others (Pradhan, 2013). However, there is strong evidence that support vector machines (SVM) based landslide susceptibility assessment yields fairly accurate results and that the algorithm either outperforms or yields similar results as other machine learning algorithms (Kavzoglu, Sahin, & Colkesen, 2014; Marjanović, Kovačević, Bajat, & Voženilek, 2011; Pham et al., 2016; Pradhan, 2013; Tien Bui, Pradhan, Lofman, & Revhaug, 2012). Therefore, in our study, we applied a SVM algorithm with radial basis function (RBF) kernel. We used the ‘*e1071*’ package in R statistical programming platform for SVM-based model development (Meyer, 2004).

SVM is a non-parametric statistical learning approach (Vapnik & Kotz, 1982). In essence, the algorithm finds a hyperplane that separates the data points into discrete classes, which are often predefined, such that there are minimal misclassifications (Mountrakis, Im, & Ogole, 2011). Therefore, in essence the algorithm attempts to maximize prediction accuracy in a given training set. Another critical aspect of the algorithm is that it relies on support vectors to determine the shape and width of the hyperplane and these support vectors are only a few of the observations in the training set. That said, often SVM-based models produces accurate results even when the total number of observations in the training set are small (Mountrakis et al., 2011). This in our case is an advantage since we can use our modest number of training observations of landslide incidences to generate fairly accurate results.

During SVM model development, we used the landslide incidence training data, i.e., with 189 point locations. In addition, we randomly sampled at-most seven pixels from each of the 31 geological regions. This yielded us with a stratified and (near) equalized random sample comprising of 186 incidence location and 216 no incidence locations. It is worthy to mention here that we collected the landslide no-incidence locations by randomly sampling pixels assuming that landslide incidence is a rare event and is likely will not be sampled in our random sample of landslide no-incidence locations. Using this landslide training data, we developed series of different SVM-based models to identify an optimum set of predictor variables. This we achieved by performing an iterative feature selection using a Monte-Carlo Cross Validation approach. A key reason for our use of Monte-Carlo Cross Validation Approach is to allow a stratified sub-sampling ensuring we get samples from all geological regions for training.

For a given set of predictor variables and a single SVM model run, we randomly generated a 70% training and 30% testing sample set, such that the 70% training set has observations for all thirty-one geological regions. Next, we trained the SVM model using the training set and calculated prediction accuracy from the testing set. In our Monte-Carlo Cross Validation approach, we repeated this process of randomly generating training and testing set, training the model, and subsequently calculating prediction accuracy from the testing set for 10^6 times per a given set of predictor variables. From the resultant 10^6 testing accuracies, we calculated the mean testing accuracy for the respective variable set. This forms

Using the cross validation approach, we performed backward variable elimination procedure. First, we calculated mean training accuracy of models trained using twelve predictor variables. Here we obtained a mean accuracy of 90.83%. Next, we dropped variables in the following order and obtained the following mean accuracies:

1. NDVI variables (Mean Accuracy = 84.22%),
2. Precipitation variables (Mean Accuracy = 91.41%),
3. Distance to Stream variable (Mean Accuracy = 91.40%),
4. Geology variable (Mean Accuracy = 88.89%), and

5. DEM variables (Mean Accuracy = 88.06%)

We observed a slight increase in mean accuracies when Precipitation variables (0.58% increase) and Distance to Stream variable were dropped (0.57%). This suggests that these variables were insignificant in assisting to predict landslide outcomes. Therefore, in our next variable set, we dropped these variables set, with five total number of variables, and obtained a mean accuracy of 91.68%. In our subsequent models, we trained the model by eliminating predictor variables from the remaining seven predictor variables.

From the remaining seven predictor variables⁴, we dropped variables in the following order and obtained the following mean accuracies:

1. Geology (88.19%), and
2. NDVI minimum (91.60%)
3. NDVI minimum and NDVI maximum (91.24%)

Here again, we noted that dropping NDVI minimum and NDVI maximum does not result into a significant decrease in mean accuracy (0.44% decrease). Therefore, we eliminated these variables from our variable set. Upon subsequent elimination of prediction variables, we noted a decline of greater than 2%, which we applied as a stopping criteria. This iterative feature selection procedure yielded us with a five-variable predictor set: Elevation, Slope, Aspect, Geology, and Mean NDVI, which yielded a mean accuracy of 91.24% computed with the testing set of observations.

After identifying the optimal set of input variables, we randomly selected a model that has the testing accuracy greater than the mean accuracy of 91.24%. An alternate to random selection of a model could have been to predict landslide susceptibility raster from a large number of random models and then to calculate per-pixel mean values of landslide susceptibility. However, this approach would have been computationally expensive. Therefore, our randomly selecting a model is computationally efficient and outputs a single landslide susceptibility raster predicted from a model of accuracy greater than mean accuracy. Furthermore, the rationale for selecting a model of accuracy greater than mean accuracy are two-fold. First, we assume that there is a certain level of bias in the input reference data due to which there is some variability in the accuracies obtained from different models. Second, we assume that randomly selecting a model of accuracy greater than the mean accuracy will ensure that the selected model is somewhat superior to other models. Using the random selection, we obtained a model with testing accuracy of 94.21%. We used this model to predict a landslide hazard raster.

⁴ These include, Elevation, Slope, Aspect, Geology, Mean NDVI, Maximum NDVI, and Minimum NDVI.

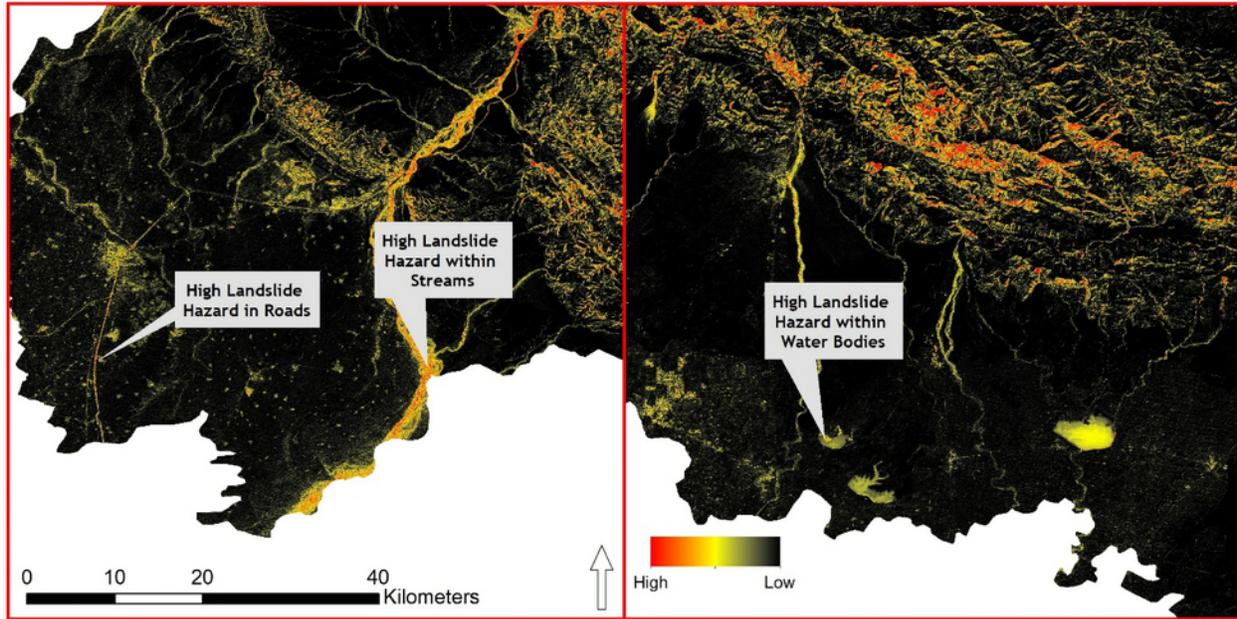


Figure 4.4. Errors in the predicted landslide hazard raster that needed slope and water masking.

From a visual examination of the predicted landslide hazard raster, we noticed two consistent errors (Figures 4.4). First, water bodies were labeled with hazard values of around 0.5. Second, in regions with slope approximately less than 10% there were high landslide hazard pixels identified such as in roads and within streams. In order to remove these two errors, we applied a water mask and a low-slope mask. We calculated a water mask by calculating mean NDVI from Landsat 7 ETM+ 32-day NDVI composite images over a period of five years from 2010 to 2015. We identified water pixels by thresholding pixel values lesser than -0.07^5 . We calculated the slope mask from thresholding the slope values of lesser than 10 degrees. Finally, we applied the slope and water masks to generate the final landslide hazard raster.

4.2.3 Validation

We validated our landslide hazard map using the independently collected validation data set. Our validation approach comprised of resampling the validation data set and calculating ROC measures for the resampled observations. The rationale for resampling is to examine the distributions of ROC measures. Resampling was done under the assumptions that 0.05%, 0.1%, 0.2%, 2%, and 20% of the locations are hazard-prone. Under each assumption, we calculated mean ROC after resampling the data 10^4 times. In a given assumption of, for example, 2% hazard-prone locations, we resampled 2% locations where landslide incidences were observed and 98% locations where no landslide incidences were observed. We compared the resampled

⁵ The threshold of -0.07 was identified by visual inspection of pixels with water.

locations with their values in the predicted landslide raster map and calculated the ROC value⁶. Finally, we examined how the mean ROC values vary under different assumptions of landslide hazard distribution.

4.3 Results

As also noted earlier, from our selected five-variable predictor set, we obtained models that exhibited mean testing accuracy of 91.24% (Figure 4.5). In fact, the 90% confidence interval showed that the testing accuracies ranged from 87% to 95% (Figure 4.5). Indeed, these testing accuracies are not precise estimates of how accurate the landslide hazard map will be when generated from the five-variable predictor set. However, it provides a preliminary evidence that the machine learning model is able to predict landslide hazard at acceptable level of accuracies. Therefore, an additional validation using an independently-collected reference data was necessary.

Using our independent reference data, we calculated ROC metrics under different assumptions of landslide hazard distributions. Here we obtained two key results. First, we noted that the ROC metrics increase to a certain extent when we assume that landslide hazards are widespread in the State compared that they are rare events (or that only a small fraction of the area is prone to landslides) (Figure 4.6). Second, we found that under our different assumptions of landslide hazard distribution the ROC metrics were greater than 0.90 (Figure 4.6). This suggest that our machine learning model will place a randomly chosen landslide incidence greater than a randomly chosen no landslide incidence location with at least a rate of 90%. This finding is very promising given that our model used only a five-variable predictor set for determining landslide hazard across the State of Uttarakhand. This also suggest that our select variables have sufficient variability to allow distinguishing landslide hazards across regions.

One main caveat we experienced while generating the output, as also noted earlier, was the landslide hazard identified in plain areas, i.e., with near-zero degree slopes and in water bodies and streams. Although we overcome this limitation by applying a water and a slope mask, this demonstrate a limitation of our machine learning model. We conjecture that by adding additional variables for example, NDWI and Distance to Roads, these errors may be minimized or perhaps eliminated. There are two reasons for why we did not include these variables in our current assessment. First, NDWI is somewhat correlated with NDVI, which is currently used in the variables, and based on our survey of existing studies, we did not find its use in assessing landslide hazards. Second, although distance to roads may be related to landslide hazards, but we assume it more to be an exposure variable that determines landslide risk rather than hazard. This assumption holds especially given that our model is static. Nonetheless, we agree that for a dynamic model distance to roads and changes in distance to roads might be a significant predictor of landslide incidences.

⁶ ROC value of a model output suggest how frequently the model will rank a randomly chosen landslide incidence higher than a randomly chosen location with no landslide incidence.

Our landslide hazard map show very distinct spatial profiles in our study area. First, our results show that in the downstream regions there is minimal landslide hazard (Figure 4.7). This is obvious because these regions have almost flat slopes. That said, these regions are like prone to flooding rather than landslides. These regions include areas from the following four districts: Dehradun, Nainital, Udham Singh Nagar, and Haridwar. In fact, we identify sub-districts of these districts as a “cold-spot” for landslide hazard (Figure 4.7b). Second, our results show that the landslide hazard map closely captured the Kedarnath landslide location as a hazard-prone region (Figure 4.8). This further emphasizes the validity of our landslide hazard map. Third, the regions and sub-districts in the north of the State that are Snow-covered are not identified to be landslide-prone (Figure 4.7 and 4.8). Intuitively, these regions are likely to be more prone to avalanches and therefore should not be identified as landslide-prone. Fourth, our results show that most of the landslide hazard is located in the central regions in or sub-districts of the State (Figure 4.7 and 4.8). In fact, we identified sub-districts forming the spatial hotspot of landslide hazard at 95% statistical significance (Figure 4.7b). These sub-districts are located in the following seven districts: Uttarkashi, Chamoli, Garhwal, Pithoragaph, Bageshwar, Rudraprayag, and Tehri Garhwal.

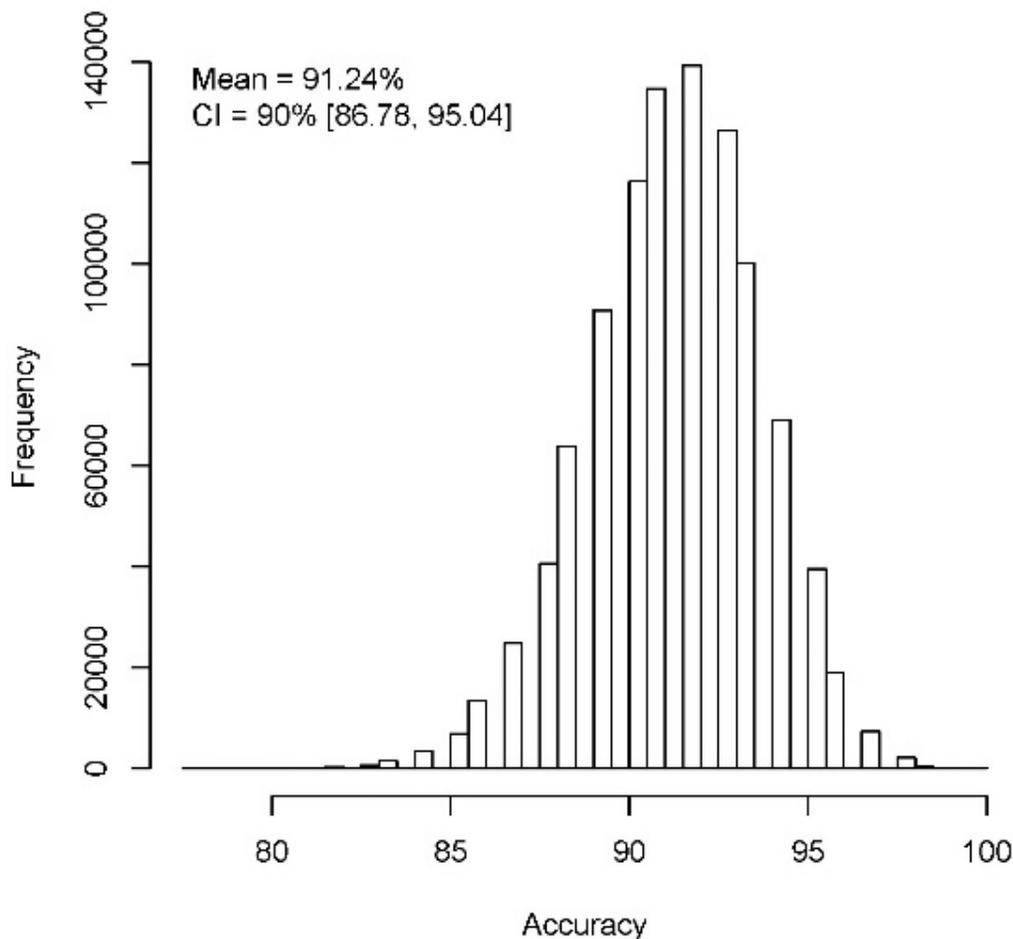


Figure 4.5. Distribution of accuracies obtained from models trained with five select predictor variables. Note that the mean accuracy of the models is 91.24%.

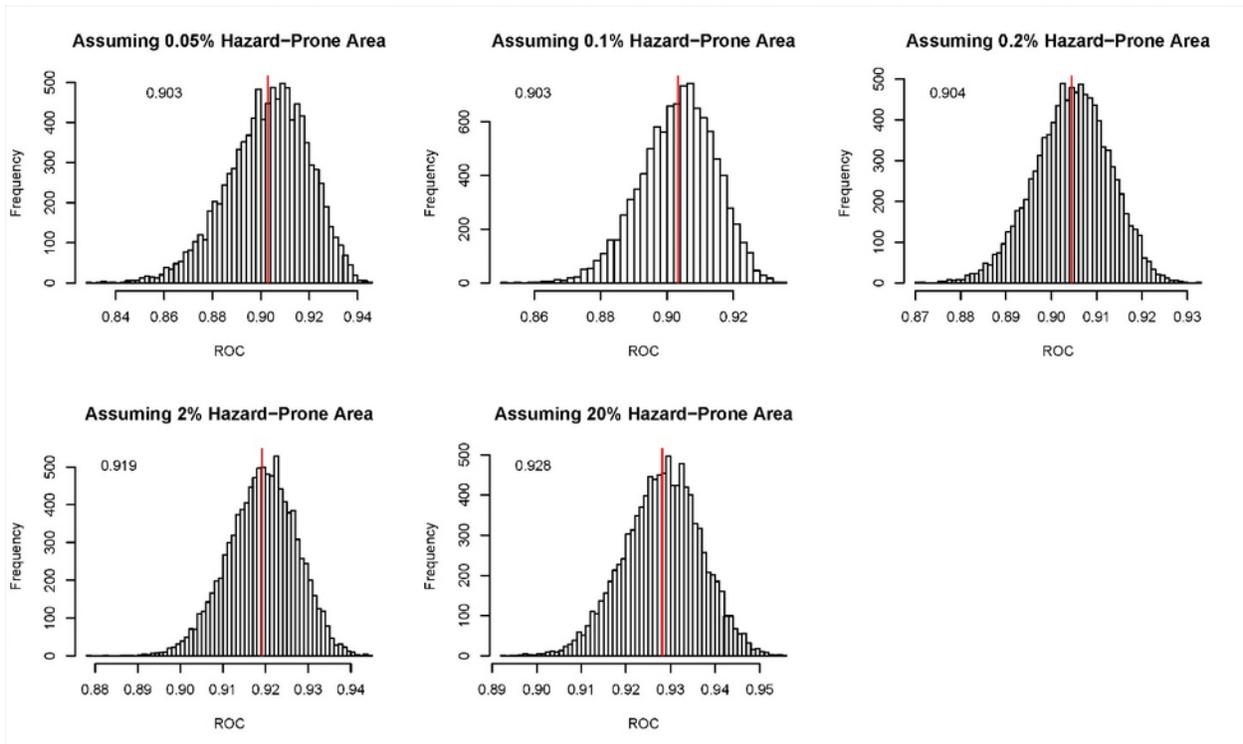


Figure 4.6. ROC metrics obtained under assumptions of different landslide hazard distribution.

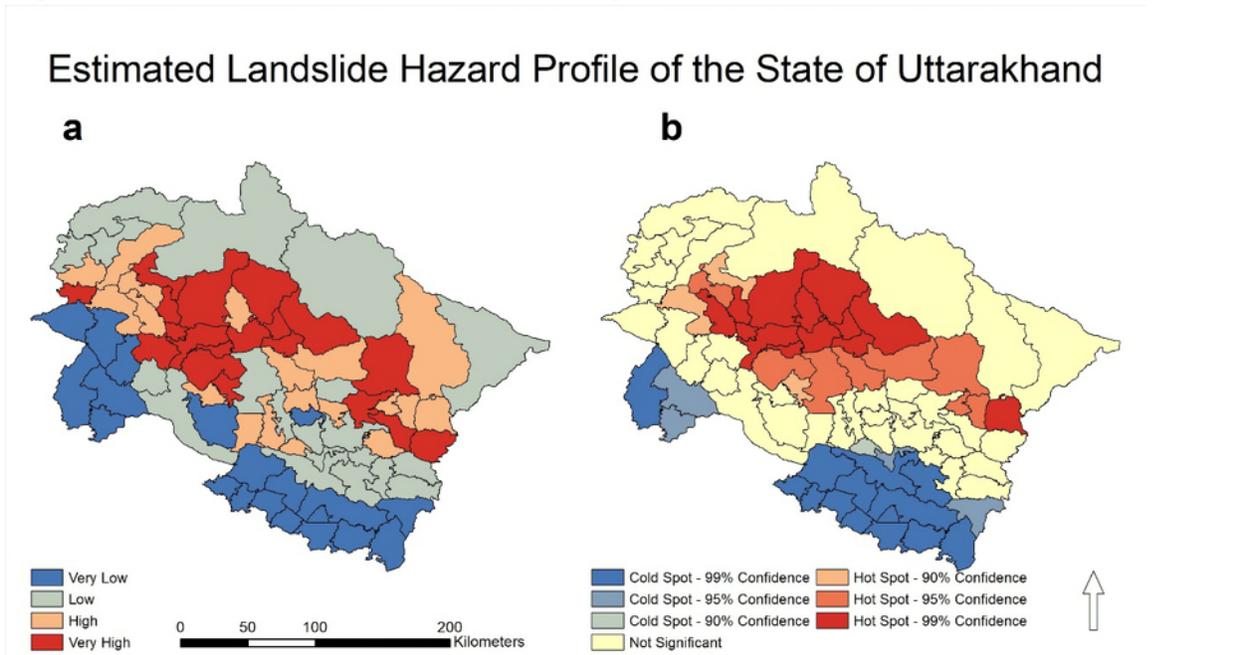


Figure 4.7. Landslide hazard profiles of the State of Uttarakhand at the sub-district level calculated as mean of all the pixels in a given sub-district: (a) Classification of sub-districts into four hazard classes. (b) Hotspots of landslides in the State.

LANDSLIDE HAZARD MAP OF UTTARAKHAND

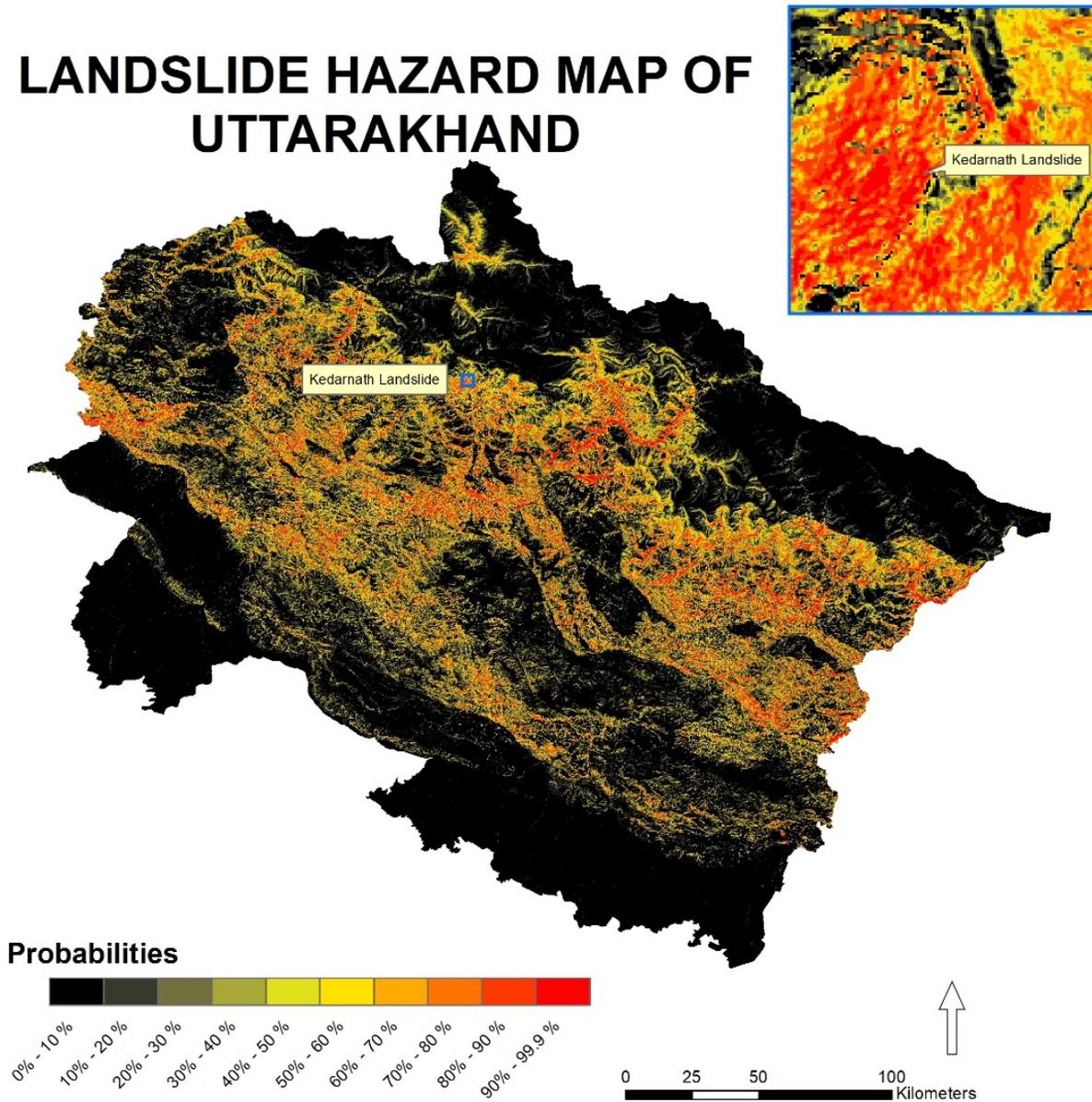


Figure 4.8. Detailed landslide hazard map of Uttarakhand.

4.4 Discussion and Further Work

This study provides a first state-wide and spatially-explicit assessment of landslide hazard using machine learning. Our study confirms that machine learning is an appropriate, accurate, and efficient tool for assessing landslide hazards over a region. Another important aspect of our analysis is that it relies on a few widely available but important input variables and is quick to be implemented. This makes our approach efficient to scale to other regions. Indeed, the “Geology” variable used in our model may not be available in other regions. Therefore, other global maps such as the “global Hydrogeology Maps (GLHYMPS) of permeability and porosity” may be

used to stratify the topography based on hydrologic and lithological factors (Gleeson, Moosdorf, Hartmann, & van Beek, 2014). Our validation suggests that our machine learning model is able to place a randomly chosen landslide incidence greater than a randomly chosen no landslide incidence location with a frequency of at least 90%. This detection rate is comparable to that reported in Pradhan (2013) and Pham et al. (2016). Pradhan (2013) reported ROC values ranging from 0.82 to 0.94 and Pham et al. (2016) reported values ranging from 0.91 to 0.95. Therefore, our results corroborate previous efforts in landslide hazard assessment.

Asides the three highlights of our study, we noted three main limitations of our approach. First, our machine learning model is static and fails to inform when a landslide is likely to occur. This suggests that this static assessment should be repeated on a time-scale at which significant landform and land-cover changes occur. Nonetheless, this limitation currently exist in all of the previous landslide hazard assessments at large spatial-scales. Another way to turn the limitation into an advantage is to use our landslide hazard map in conjunction with near-real time data on precipitation, which could trigger landslides, land-use changes, such as, settlement expansion and construction of roads, to arrive at dynamic landslide risks. Second, due to the limited availability of high resolution time-series images in Google Earth, we collected a limited number of input landslide incidence locations. This to a certain extent limited our ability to explore other machine learning algorithms and comprehensively validate our landslide hazard output. Third, our approach run as a black-box yielding no mathematical information on how predictor variables are related to the input variables. However, one can get a general idea of how important each predictor variable is in relation to the output by examining correlation coefficients.

Our study allude to two possible future works. First, to examine how landslide hazard maps coupled with dynamic variables could inform decision making and policy planning in landslide hazard-prone areas. This work will entail examining the extent to which combining a static hazard map with dynamic variables change our understanding of the process. Currently, we are unaware of any effort that has utilized static hazard maps produced from machine learning or other quantitative approaches, with near-real time data on precipitation and land-use changes. Second, our study highlight the need for an open access inventory of landslide histories. Such inventories are available at the global scale, for example, the Global Landslide Catalog. However, observations in these global inventories are not sufficient enough for making them locally-relevant. Therefore, an extensive locally-relevant inventory of landslide histories may be necessary for planning decisions and generating or updating landslide hazard maps.

5. SOCIAL VULNERABILITY IN UTTARAKHAND

5.1 Introduction

In the Chicago heat wave of 1995, black communities that had the same rates of violence, poverty, and were located in the same area experienced significantly different death rates (33 vs. 3 deaths for every 100,000 residents in one example); depending on how frequently their

community members interacted with each other. Knowing one another—from church, talking on the street, or meeting in local stores—proved to be lifesaving (Klinenberg, 2003). When Katrina hit land, New Orleans had a relatively low elderly population. Sixteen percent of the city’s residents were over 60 according to the 2005 US Census. Yet 75% of deaths from the hurricane were people in this age bracket (Brunkard, Namulanda, & Ratard, 2008).

Two communities hit by the same physical hazard, in land flooding in this case, will likely experience different amounts of loss in the short and long term. The field of social vulnerability investigates why this is the case and quantitative vulnerability uses the results to estimate or map who within a population is most at risk. The present assessment examines: 1) what social characteristics drive vulnerability in Uttarakhand and 2) by extension which towns in the state may be more likely to experience loss during extreme flooding and other fast on-set disasters.

The Intergovernmental Panel on Climate Change’s defines vulnerability as:

“the propensity or predisposition to be adversely affected. Vulnerability encompasses a variety of concepts and elements including sensitivity or susceptibility to harm and lack of capacity to cope and adapt” (IPCC, 2014)ⁱ

Social vulnerability specifically is the propensity to experience hazard that are socially created, as opposed to physically, ecologically, or otherwise (Cutter, Boruff, & Shirley, 2003)ⁱⁱ. After decades of research, there is some consensus in the social science research community around the demographic, behavioral, and psychological characteristics that make people and communities vulnerable at least in a general sense (Cutter et al., 2003). People who have more financial resources, who are not especially young or old, and have strong community support are less vulnerable. The elderly are vulnerable because of their health, disability, lack of transport, and lack of access to information and other resources (Ngo, 2001). Communities with a majority of their population above age 65 are likely to be more vulnerable than with a majority population between ages 30 and 45. Conversely, children, particularly infants and young children, are vulnerable because of their dependence on adults and their psychological impressionability (Peek, 2008). Crime can indicate reduced community cohesion, and prevent evacuation in rapid onset events like fires and floods. Governance may be weak in violent areas, leading to corruption of disaster aid, and preventing help from getting to those most in need (Tellman, Alaniz, Rivera, & Contreras, 2014). In many scenarios women are more vulnerable than men because of their lack of resources—both material and informational (A. Fothergill, 1996; Neumayer & Plümper, 2007). Psychological factors are increasingly recognized as significant at every stage of disaster response (Werg, Grothmann, & Schmidt, 2013). Furthermore, culture has a strong influence on risk perception and requires a very local and nuanced analysis to understand, which often demands qualitative study (Adger, Barnett, Brown, Marshall, & O’Brien, 2013).

These dimensions of vulnerability fall along a spectrum of universality or generalizability; some dimensions are fairly well documented and consistent across geographies (Cutter et al., 2003), while others vary significantly across time, place, and context. This report focuses primarily on the more generalizable dimensions because they are often easier to measure directly.

While the dimensions of social vulnerability have mostly been explored through qualitative methods, there has been an academic effort over the last two decades to quantify these dimensions so that we can estimate or even predict social vulnerability. These assessments have primarily come in the form of geospatial indices (de Sherbinin, 2014). Over the last two decades, social vulnerability researchers have begun to distill the dimensions of social vulnerability into empirically based indicators. When combined in summary indices, typically using demographic information, these tools describe who is most vulnerable and where the most vulnerable are located before, during, and after a crisis (Tate, 2012). If measured using benchmarks and monitored over time, these indicators may serve as diagnostic tools.ⁱⁱⁱ

Our methodology is based on the Social Vulnerability Index (SoVI) developed by Dr. Susan Cutter at the University of South Carolina. SoVI uses Principal Components Analysis (PCA) based factor analysis on a large set of county- or tract-level US Census variables in order to determine a set of underlying dimensions of vulnerability *e.g.* Hispanic ethnicity, special needs individuals, Native American ethnicity, and service industry employment (Cutter et al., 2003). An updated model in 2010 added new dimensions such as family structure, language barriers, vehicle availability, medical disabilities, and healthcare access in the preparation for and response to disasters. Other approaches to social vulnerability or similar themes that use other methods and datasets exist, but those are not as widely relied upon within the scientific or practitioner community. The factor analysis based approach and other methods have been used in a small set of countries and regions around the world. The Asian Cities Climate Change Resilience Network, a project of the Rockefeller Foundation, created vulnerability assessments to identify and understand the current location and dynamics of vulnerable urban populations in three Indian cities: Gorakhpur, Indore, and Surat (The Rockefeller Foundation, 2008).

Although social vulnerability and resilience sciences have advanced immensely in the last two decades, the social science—particularly for developing countries—lags considerably behind the science of the geophysical dimensions of hazards. Yet, it is important to understand what makes developing communities vulnerable where the climatic changes are likely to hit hardest and where existing inequality is often the greatest. India and Uttarakhand in particular is well suited as a case study as a disaster prone, developing region with social inequity.

Qualitative and statistically descriptive assessments of India consistently describe several attributes that make certain groups more vulnerable in general. Poverty and marginalized communities, which are chiefly concentrated in the northern region, are consistently estimated to be at higher risk and subject to a variety of other threats like violence that compound existing

vulnerability. Women in the state also face specific gendered risks and vulnerabilities: they receive up to 30% lower wages than men in casual labor receive (World Bank, 2008). Women in the region generally lack stable employment. For instance, most agricultural workers are female, though most women do not own land (National Alliance of Women, 2008) and are generally more responsible for domestic needs than men. A history of invasion in the violent Northern area leaves communities in those regions less stable than the southern portion of the region and may lead to heightened vulnerability. The significant religious, cultural and linguistic differences can further isolate the South from the North (Brenkert & Malone, 2005). A strong sense focus on extended family unit (Gannon, 1994) can also determine a neighborhood or individual's ability to withstand shock. Ray-Bennett (2009) suggested a complex interplay of caste, class, and gender in surviving disasters. Besides these attributes, there are several more that make certain communities or groups more vulnerable than other. A key challenge in assessing social vulnerability in India lies in accounting for social and economic diversity. This diversity makes vulnerability profiles of different regions to differ significantly with each other.

In India, social vulnerability has been studied in different contexts. These contexts differ in that these could be for specific hazards such as, coastal threats and cyclone hazards (Mahapatra, Ramakrishnan, & Rajawat, 2015; O'Hare, 2001). Or these could be for general hazards termed as, natural and anthropogenic disasters (Nisha & Punia, 2014), climate change (Maiti et al., 2015), and multiple disasters (Ray-Bennett, 2009). With these different contexts, and also regional variations in social and economic conditions, there is no fixed method to assess social vulnerability in India (Yenneti, Tripathi, Wei, Chen, & Joshi, 2016). However, some attributes are consistently used in social vulnerability assessments. For example, gender ratio or female-to-male ratio is constantly used in social vulnerability assessments across contexts and regions with an understanding that higher gender ratio implies a higher sensitivity of the population to be affected by a given hazard (A. Fothergill, 1996; Holmes, Sadana, & Rath, 2010; Ray-Bennett, 2009). Similarly, female literacy rate is another attribute that is central to social vulnerability in India (Bahinipati, 2014; Brenkert & Malone, 2005; Maiti et al., 2015; Yenneti et al., 2016). Besides gender-specific attributes, poverty and concentration of marginalized groups is a determinant of higher social vulnerability as these groups are considered more sensitive than others and have less adaptive capacity (Alice Fothergill & Peek, 2004; Holmes et al., 2010; O'Hare, 2001). In the context of cyclone hazard, O'Hare (2001) asserted that poverty and social organization in India make individuals and social groups to have different limits on their risk-alleviation abilities. This assertion can well be extrapolated to other contexts and used generally. Therefore, while contexts may differ, some social attributes can be generalized to measure social vulnerability quantitatively. This report explores a quantitative vulnerability assessment of some of the generalizable dimensions of vulnerability through an exploratory model.

5.2 Methodology

Social vulnerability cannot be measured directly, at least in full, so we use variables that have been measured and monitored, and try to model underlying relationships, both positive and negative, in the data (Cutter et al., 2003). The goal of this factor analysis-based social vulnerability model is to reduce the measurable (and available) characteristics of Uttarakhand to the latent dimensions that may determine social vulnerability to disaster.

The analysis primarily relies on a multi-scale and multi-source national database of social, economic, and infrastructure variables. For the State of Uttarakhand, the database contains sixty-five social, economic, and infrastructure-related variables for all the seventy-eight sub-districts of the State of Uttarakhand. The full list of these variables is represented in Appendix 2. Indicators of vulnerability from the dataset were selected based on the IPCC definition (above), IPCC framework for vulnerability, and based on pairwise comparison to identify highly correlated variable pairs. Furthermore, our selection of variables was informed by feedback obtained from interviewing experts. In effect, we used the IPCC framework to measure social vulnerability considering sensitivity and adaptive capacity as the two characteristic dimensions that determine overall social vulnerability using the following conceptual formulation (McCarthy, 2001):

$$\text{Social Vulnerability} = \text{Sensitivity} + \text{Adaptive Capacity}$$

Following the aforementioned variable selection criteria, four variables were selected at the sub-district level. In addition, we selected four variables at the district level, which we found important for social vulnerability based on our literature review. Each of the district-level variable was disaggregated linearly at the sub-district level based on weights derived from a weight variable. Primarily, nightlight intensity per area, inverse of nightlight intensity per area and total population variables were used as weight variables. While this disaggregation approach will inevitably inflate the strength of the correlation between weight variables and the disaggregated variables, a PCA-based factor analysis approach with varimax rotation will ensure maximum separation between different factors. A list of all the selected variables along with literature references are represented in Table 5.1.

PCA is a statistical dimension reduction method that determines which variables in a dataset group together after an orthogonal linear transformation in N dimensional space (Abdi & Williams, 2010). These groups or components can be used to categorize the original variables based on the variables relatedness to each new component (Cutter et al., 2003). The first component corresponds to and therefore describes the most variables in the dataset and the second orthogonal component describes the second biggest grouping of input variables (Abson, Dougill, & Stringer, 2012).

We performed PCA based factor analysis on the twenty-seven variables shown in Table 5.1 in the R programming platform (Revelle, 2016). Using the value 1 as a cut off threshold for eigenvalues, we selected seven components (Figure 5.1). Using a varimax rotation, we extracted

seven factors that explain ~81% of the variance in the data. Sub-districts in Uttarakhand were ranked in an overall index by calculating an equally weighted sum of factors for the sub-districts. Here, we then applied a directional to each component based on our interpretation of how each factor will affect overall social vulnerability. After the overall social vulnerability was obtained, the index was normalized to range from 0 to 1. Lastly, we identified the top twenty sub-districts with high social vulnerability.

Table 5.1. Input variable matrix and literature overview

S. No.	Name	Variable Name	Disaggregated	Weight Variable	Representative Dimension	Reference
1	Gender Ratio	sex_rat_t	No	-	Sensitivity	(A. Fothergill, 1996; Holmes et al., 2010; Ray-Bennett, 2009)
2	Scheduled Caste	sc	No	-	Sensitivity	(Ray-Bennett, 2009)
3	Scheduled Tribe	st	No	-	Sensitivity	(Ray-Bennett, 2009)
4	Illiteracy rate, 7+ years (Total)	edu_lit_7_t	No	-	Adaptive Capacity	(Brenkert & Malone, 2005; Maiti et al., 2015; Yenneti et al., 2016)
5	Dependency ratio	dep_rat_t	Yes	Population	Sensitivity	(Brenkert & Malone, 2005; Newport & Jawahar, 2003; Ngo, 2001)
6	Unemployment rate, 15+ years, usual (total)	une_15_u_t	Yes	Population	Sensitivity	(Yenneti et al., 2016)
7	Poverty rate at national poverty line	pov_nl	Yes	Inverse of Nightlight intensity per area	Sensitivity	(Alice Fothergill & Peek, 2004; Holmes et al., 2010; O'Hare, 2001)
8	Infant mortality rate (Total)	imr_t	Yes	Inverse of Nightlight intensity per area	Sensitivity	(Bahinipati, 2014)

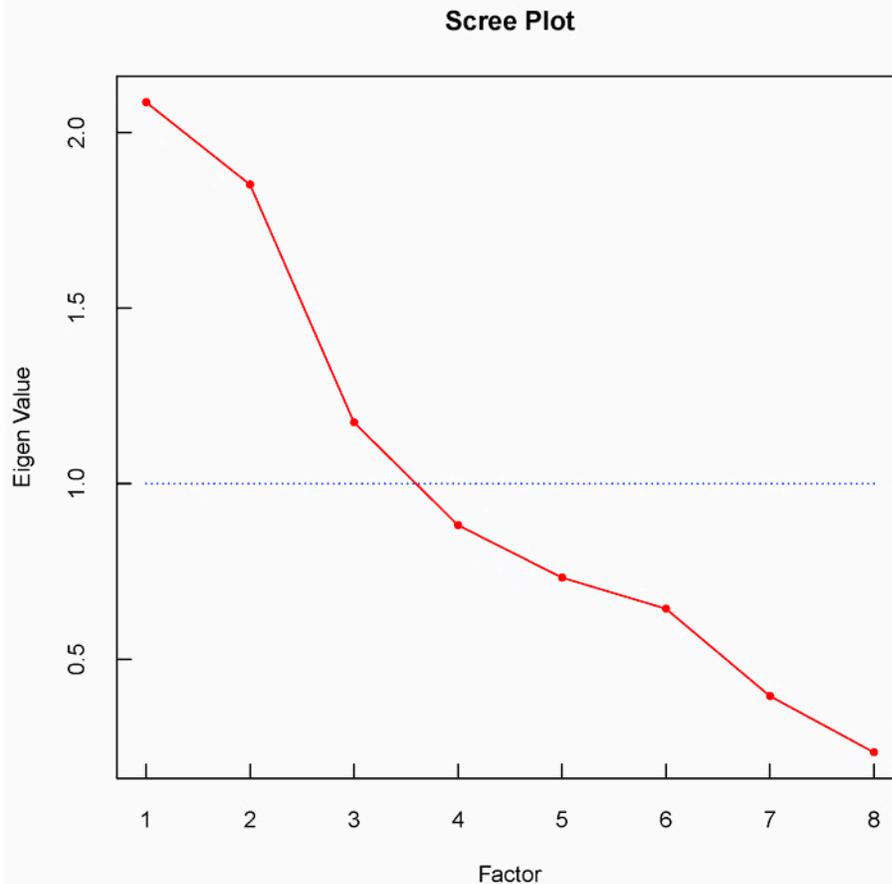


Figure 5.1. Eigen values for individual factors obtained from principal component analysis based factor analysis. Blue horizontal line shows the cutoff threshold.

5.3 Results

A vast majority of the social vulnerability indicators in the model can be explained by three composite factors that differentiated sub-districts based on their relative social vulnerability or based on underlying dimensions of the data. All the three factors explains equivalent amount of variance and are reflective of the important yet independent dimensions of social vulnerability. Table 5.2 lists the three factors.

Marginalized Population - I

The strongest variable loaded in the first factor is *Percentage of Scheduled Tribe Population* with a correlation of 84%. Also, this factor is correlated with Illiteracy rate variable ($r = 0.69$). Upon examining the data we found that the *Percentage of Scheduled Tribe Population* and *Illiteracy rate* variables are correlated ($r = 0.46$) in the State. Often illiteracy and presence of tribal groups, make a region socially more sensitive and less adaptive to external shocks (Ray-Bennett, 2009; Yenneti et al., 2016). Therefore, this factor captures both the sensitivity and adaptive capacity

aspects of vulnerability and, therefore, positively influences social vulnerability. Overall, this factor variable represents 23% of the variation in our data.

Marginalized Population - II

Our second factor captures some important dynamics operating the region. Essentially, it suggests that high gender ratio, or high female population relative to male population, is inversely correlated with unemployment. In simple terms, unemployment is concentrated in regions with predominantly male population. In the State of Uttarakhand, with a majority rural population (~70% of the total population), females often assume domestic and subsistence agriculture work (National Alliance of Women, 2008; Singh & Garia, 1999). In fact, Singh & Garia (1999) suggested that females account for 85%, 81%, and 81% of the total working hours spent by a household in agriculture, animal husbandry, and domestic works. The male population, on the other hand, often migrate to cities in search of jobs, wherein unemployment levels are unusually high. Now, unemployment and high gender ratio influences the social vulnerability of a region by making the population more sensitive. For example, based on our interviews of regional experts in social vulnerability, we noted that female population is generally more vulnerable to disasters than total population. Additionally, unemployment decreases access to resources for the population hence reducing their adaptive capacity. Here, we assume that gender ratio plays a major role in social vulnerability and this in the simultaneity of marginalized scheduled caste population suggest that our second factor also positively influences social vulnerability. This factor explains 21% of the total variance in our data.

Poor Access to Resources

Our third factor explained 21% variation in our data and captures aspects of *Poverty rate*, *Infant mortality rate*, and *Dependency ratio*. High poverty rates reflect less access to resources, which make a population more sensitive and vulnerable to external shocks, including, disasters. Similarly, high infant mortality rate is a proxy measure of poor health services (Bahinipati, 2014; Jain, 1985) and high female infant mortality, specifically, of low socio-economic development (George, Abel, & Miller, 1992). Another dimension positively and significantly loaded in our third factor variable is *Dependency Ratio*. High dependency of children and old-age population to working population suggests less per-capita access to resources. Therefore, in all, this variable captures aspects of poor access to resources.

Table 5.2. Possible dimensions of Social Vulnerability in Uttarakhand

Factor	Interpreted Name	% variation explained	Governing Variables			Correlation Coefficients			Expectation (Directional)
			First	Second	Third	First e	Second	Third	
1	Marginalized Population - I	23%	Schedule Tribe	Illiteracy rate, 7+ years (Total)	Gender Ratio	0.84	0.69	-0.58	+
2	Marginalized Population - II	21%	Unemployment rate, 15+ years, usual (total)	Gender Ratio	Scheduled Caste	-0.69	0.64	0.57	+
3	Poor Access to Resources	21%	Poverty rate at national poverty line	Infant mortality rate (Total)	Dependency ratio	0.79	0.75	0.60	+

Based on the final social vulnerability score, we identified top twenty highly vulnerable sub-districts Table 5.3. These most vulnerable sub-districts are identified majorly in the border areas of the State predominantly in the southwest, northwest, central, and eastern regions. The sub-districts in the southwest direction lie in downstream region and have relatively high population count and built density—therefore high sensitivity from poverty and unemployment—due to which these sub-districts have higher vulnerabilities.

Overall, however, *Illiteracy rate* and *Infant mortality rate* variables have the highest correlation with our social vulnerability index. This emphasizes the key role education plays in increasing the capacity to deal with hazard or disaster events. Finally, infant mortality is identified to have a major positive influence on final social vulnerabilities, underlining the need to improve health services besides social and economic development in order to reduce social vulnerability. A map of overall social vulnerability is shown in Figure 5.2.

Table 5.3. Top twenty sub-districts with high social vulnerability

Rank	Sub District
1	Laksar
2	Chakrata
3	Tyuni
4	Kalsi
5	Roorkee
6	Mori
7	Kanda
8	Puraula
9	Hardwar
10	Kapkot
11	Jakholi
12	Ghansali
13	Garud
14	Rajgarhi
15	Chiniyalisaur
16	Bageshwar
17	Lohaghat
18	Pati
19	Champawat
20	Dunda

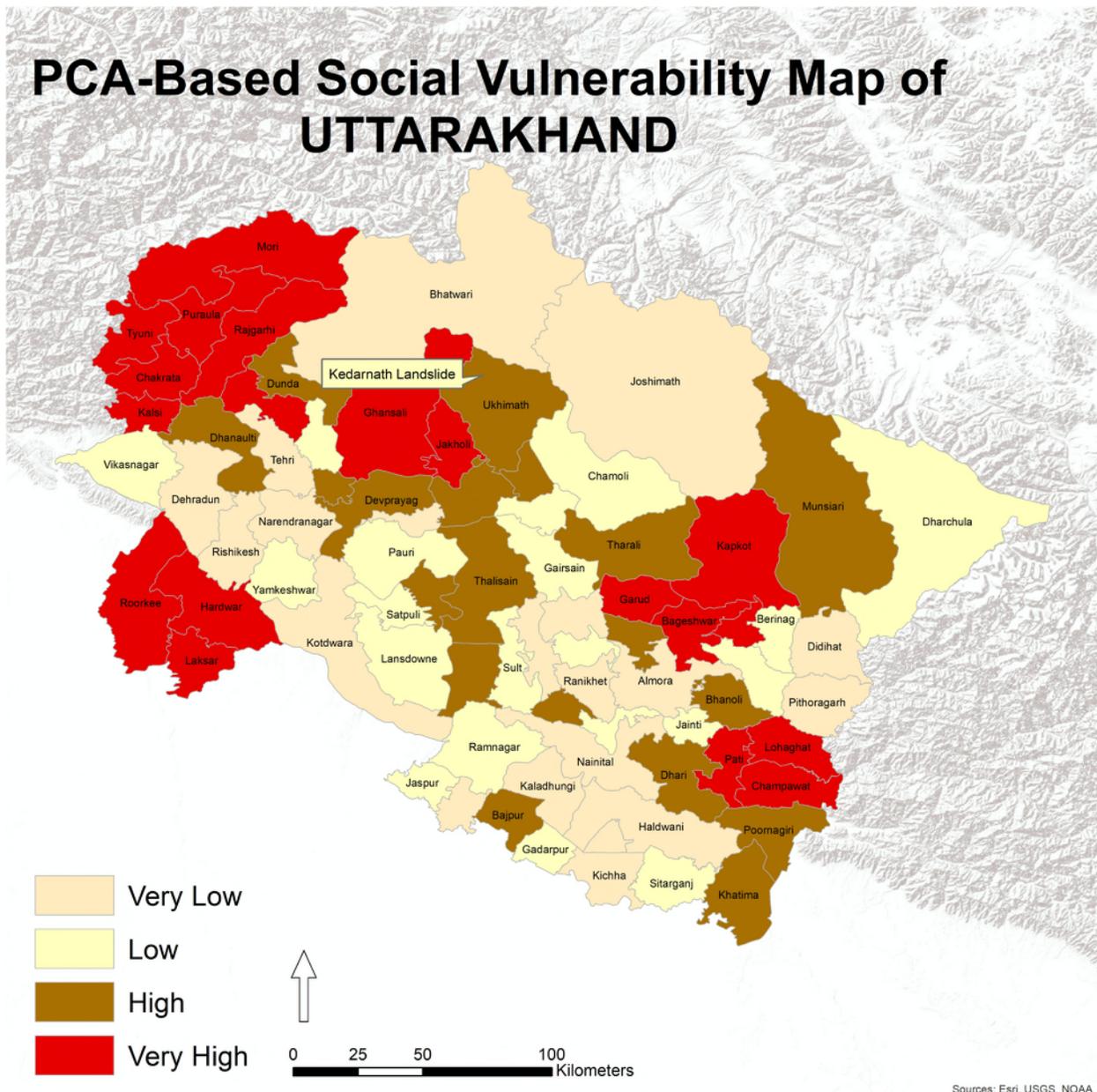


Figure 5.2. Social vulnerability index for Uttarakhand generated using principal component analysis based factor analysis with eight selected socio-economic indicators.

5.4 Data Limitations

Though the dimension reduction was very successful, the conclusions that we can make from the results are limited for several identifiable reasons. Variable selection is arguably the most important part of a PCA-based approach. The data used in a social vulnerability index needs to be current, robust, and qualitatively verified by experts. Likewise, the outputs must be interpreted and used with intimate knowledge of the region and with the statistical limitations of the data in mind.

There could be important dimensions of vulnerability that were not represented as variables in the available data. We suspect these omitted variables could be important to social vulnerability in Uttarakhand and other contexts. Likewise, our analysis may suffer from the modifiable areal unit problem. Therefore, similar vulnerability analysis needs to be conducted at finer spatial scale such as village, town, and city level. Furthermore, there are most likely aspects, that we and the consulted experts are not aware of, which are critical for social vulnerability in the region. However, with the comprehensive nature of the available data at the sub-district level, all the major aspects and dimensions of social vulnerability have been incorporated.

Beyond these data considerations, it is critical to note that PCA is not a predictive statistical model. The results presented here describe the natural groupings of the variables input into the model and not validated externally with disaster outcome data statistical sense. However, this essential point reinforces the importance of variable selection for this approach to social vulnerability. Again, the model is only as good as the variables that go into it and only describes the information it is given.

5.5 Discussion and Further Work

Of the several ways to improve the scientific understanding of social vulnerability, the most important is incorporating more, locally contextualized input data. As we explain in the introduction of this section, many of the factors that make people communities vulnerable differ widely across different cultural, political and other contexts, and at different stages of the disaster cycle. For instance, what qualifies as relatively low-income varies between communities and across time. Also, culture has a strong influence on risk perception and requires a very local and nuanced analysis to understand, which often demands qualitative study (Adger et al., 2013).

The variables for a social vulnerability index must be constructed and reviewed in collaboration or consultation with local scientists, practitioners, and/or community members. We recommend a deep literature review on the region, or India in general, with organization or groups like local World Bank staff or the Uttarakhand Disaster Mitigation and Management Centre (DMMC) and possibly with experts on social vulnerability from outside the country. To cover the full set of indicators, we would almost certainly have to include ground information from conventional national surveys. Further research should explore more remote sensing, public opinion polling (Leiserowitz & Thaker, 2012), call record data (CDR), social network analysis, and more.

We further recommend the exploration of other assessment models. Primarily, other statistical techniques, such as other data reduction methods and predictive models could be useful. Regression-based modeling is the next frontier in social vulnerability analysis (Fekete, 2009) because it ensures that the models are describing an external reality of disasters, rather than just interpreting characteristics of input data. The authors are building such a model for the U.S. over the course of 2015. This model was not created for the sake of this report for two reasons: 1) the approach has not been widely developed and verified for social vulnerability in the academic

literature and 2) the geospatial damage data necessary to build such a model were not available to the authors at the time of this analysis.

Finally, two other critical potentials for moving forward with vulnerability analysis are determining the appropriate scale of analysis and exploring the responsive or even real time applications of vulnerability analysis. Though conducting quantitative social vulnerability analysis, at never before seen local levels in India, offers new insight into the social conditions on the ground, the geographic scale at which those new results are most meaningful remains a largely unexplored research area. We recommend using variograms or another scale sensitivity-analysis technique to determine the areas in which the data is most different.

Despite its conceptual and scientific limitations, integrating the social dimensions of hazards into the disaster cycle is necessary for fully successful emergency planning and response (National Academy of Sciences, 2012). Indices can help reduce people's social risk before a disaster hits if appropriately integrated into planning and response. When fully developed, they can identify areas most in need of assistance when a disaster strikes. The appropriate index can also suggest areas most in need of recovery assistance post-disaster by knowing who had a low coping capacity prior to the disaster, and where these town and villages are. A validated and fully functioning version of the model presented here may serve these purposes.

Lastly, there are a few ways to customize the social vulnerability assessment through the model based on the user's needs and interests. Some important decisions concerns the variables to include, accurate interpretation of the identified factors, assigning appropriate weights to the factors to derive final social vulnerability scores. These decisions could be made through expert consultation and incorporating decision making methodologies (Saaty, 2008). To that end, we have devised a survey instrument⁷ relevant for our current social vulnerability analysis that may be used to derive a vulnerability index or to validate the existing social vulnerability index from PCA-based approach.

6. COMBINED SOCIO-BIOPHYSICAL VULNERABILITY ASSESSMENT

Our combined vulnerability assessment first labels sub-districts with differentiated levels of biophysical and social vulnerabilities. By applying a Quantile classification, we classified all the sub-districts into four classes based on their biophysical and social vulnerabilities: Very Low, Low, High, and Very High (Figure 6.1a and 6.1b). Our results show four sub-districts that have very high biophysical and social vulnerabilities (Figure 6.1c). Two of these sub-districts are located in relatively urban districts namely, Dehradun and Haridwar and the remaining two are in Tehri Garhwal and Bageshwar Districts. High biophysical vulnerability in three of the sub-districts located in Tehri Garhwal, Bageshwar, and Dehradun districts are due to high landslide hazard. In Laksar sub-district, however, the biophysical is flooding. We further estimated the

⁷ See Social Vulnerability Survey Form C2S.pdf

total population, road infrastructure, and wealth concentrated in sub-districts with different levels of biophysical and social vulnerability, using the Spatial Database for South Asia (Table 6.1). We found that the four sub-districts cumulatively have a population of ~0.47 million, total road length of ~600 km, and generates \$52 million (USD) as gross domestic product. We further calculated these estimates for sub-districts with high to very high social and biophysical vulnerabilities (Table 6.1). We found that sub-districts with high social and high biophysical vulnerabilities cumulatively have a population of ~1.9 million, total road length of ~1500 km, and generates \$532 million (USD) as gross domestic product. Indeed not everyone, every road, and every economic unit is vulnerable to biophysical hazards, yet these numbers tell us aspects of vulnerability at the sub-district level. By a comparison of these estimates between sub-districts with very and high biophysical and social vulnerabilities, we observe that sub-districts that are more vulnerable are those that do not generate high economic output, have relatively less population, and less road connectivity⁸. We noted that sub-districts with very high biophysical and social vulnerabilities have 45.57% lower population count and 2.57% lower GDP compared to those with very low vulnerabilities. However, highly vulnerable sub-districts have 31.82% higher road connectivity compared to sub-districts with very low vulnerability. Road connectivity perhaps are associated with enhanced biophysical and social vulnerabilities but more investigation is required in this direction. These regions, therefore, require increased efforts in ensuring disaster resilience planning and management, enhancing social security schemes, increasing access to health and education, and promoting ecologically-informed infrastructure and economic development.

Finally, our analysis also showed that the Ukhimath sub-district, wherein the 2013 disaster event unfolded; located in the Rudraprayag district, has high biophysical vulnerability due to it being a landslide prone region. Furthermore, our analysis labeled the sub-district with high social vulnerability, which suggests that the sub-district requires greater disaster resilience planning and development. Nonetheless, there are other sub-districts with similar vulnerability profiles located in Chamoli, Rudraprayag, Tehri Garhwal, Garhwal, Pithoragarh, and Udham Singh Nagar districts.

⁸ Considering total road length as a proxy measure of road connectivity levels.

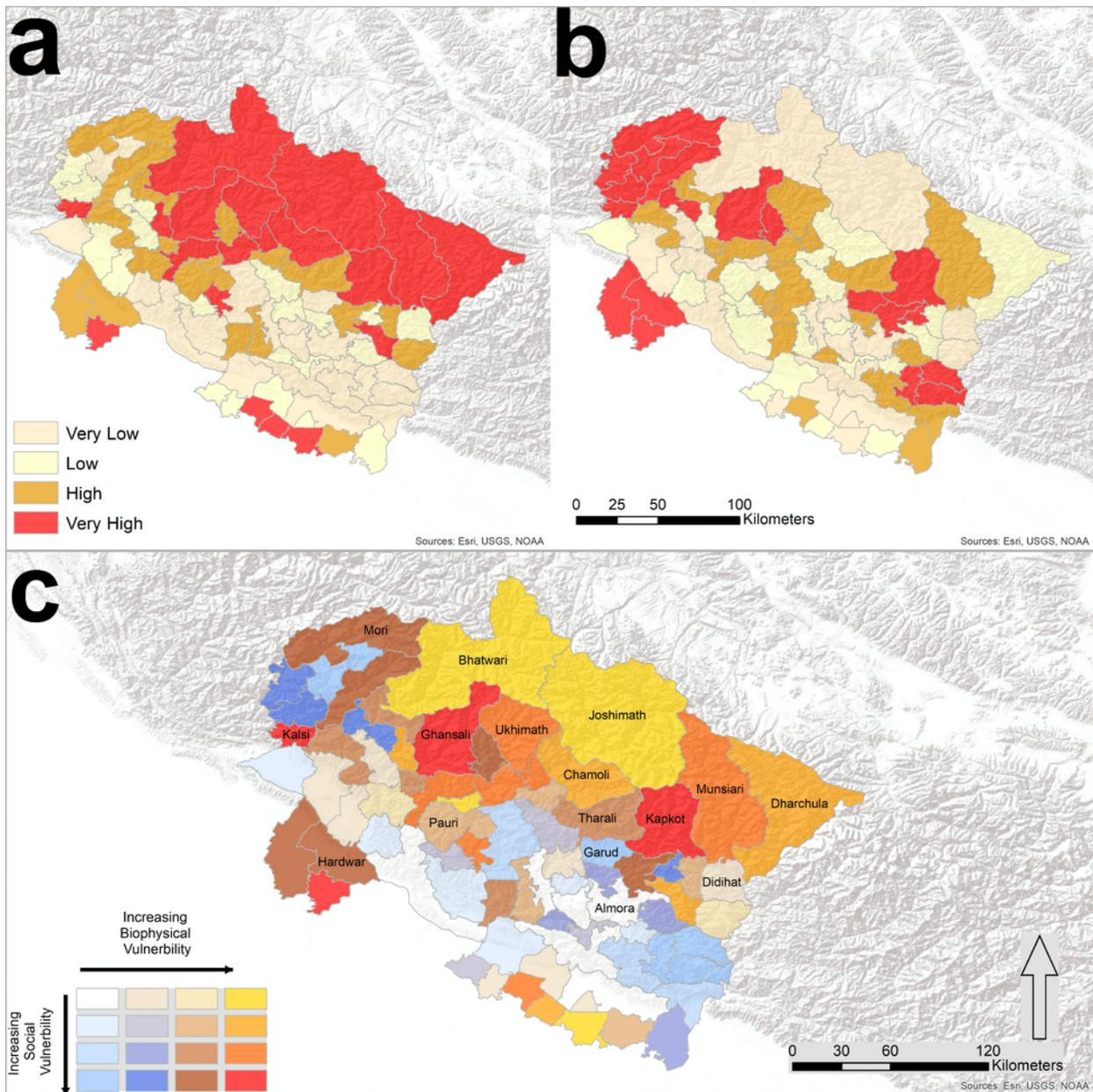


Figure 6.1 a) Biophysical hazard map prepared from combining flood and landslide hazard maps, b) Social vulnerability, and c) Combined biophysical and social vulnerability map of Uttarakhand.

Table 6.1. Risk estimates of sub-districts with very high and high biophysical and Social Vulnerability based on data for the year 2011.

VERY HIGH BIOPHYSICAL AND SOCIAL VULNERABILITY				
Sub District	District	Total Population	Road Length (km)	GDP (USD)
Ghansali	Tehri Garhwal	121,000	107	2,449,088
Kalsi	Dehradun	47,300	105	1,680,950
Kapkot	Bageshwar	64,900	173	707,820
Laksar	Haridwar	236,000	215	46,861,359
	SUB TOTAL	469,200	600	51,699,217
VERY HIGH BIOPHYSICAL AND HIGH SOCIAL VULNERABILITY				
Dunda	Uttarkashi	60,300	103	8,078,855
Tharali	Chamoli	89,100	227	5,013,051
Jakhnidhar	Tehri Garhwal	32,800	44	3,245,660
Dhanaulti	Tehri Garhwal	74,100	137	3,610,662
Dhoomakot	Garhwal	39,200	136	1,086,469
	SUB TOTAL	295,500	647	21,034,697
HIGH BIOPHYSICAL AND VERY HIGH SOCIAL VULNERABILITY				
Pokhari	Chamoli	35,900	53	6,194,203
Ukhimath	Rudraprayag	87,000	193	4,496,318
Rudraprayag	Rudraprayag	91,900	156	17,415,050
Devprayag	Tehri Garhwal	94,400	54	7,501,242
Chaubattakhal	Garhwal	38,400	33	2,171,164
Munsiari	Pithoragarh	46,500	57	22,460
Bajpur	Udham Singh Nagar	188,000	34	13,594,553
	SUB TOTAL	582,100	580	51,394,990
HIGH BIOPHYSICAL AND HIGH SOCIAL VULNERABILITY				
Mori	Uttarkashi	40,500	57	31,918
Rajgarhi	Uttarkashi	70,600	157	2,086,506
Jakholi	Rudraprayag	63,400	72	8,737,673
Bageshwar	Bageshwar	100,000	136	14,600,382
Roorkee	Haridwar	996,000	325	282,343,867
Haridwar	Haridwar	658,000	740	224,349,800
	SUB TOTAL	1,928,500	1,488	532,150,146

7. CONCLUSIONS AND RECOMMENDATIONS

Given the devastating history of flooding and landslides, notably in 2012 and 2013, this report aimed to assess the current potential threat that flooding poses to Uttarakhand from four aspects:

- 1) How geomorphic change in recent floods changes the location of rivers and surrounding floodplains?
- 2) What Biophysical factors contribute to landslide and flood hazards?
- 3) What is the social vulnerability of the population?
- 4) What are the characteristics of sub-districts with biophysical and social vulnerabilities?

To do so, the authors built four different models using satellite remote sensing, statistical analysis, and machine learning. Our first model rely on satellite image analysis to detect geomorphological changes, GIS analysis for estimating flood hazard, machine learning for estimating landslide risk, and statistical analysis for estimating a social vulnerability index. Together these outputs provide the foundations for what could be a comprehensive picture of the underlying risk that floods and landslides pose to the region and its people. This can answer questions such as,

1. Given an intense rainfall prediction, which areas will be affected hardest and why?
2. Where should government spend limited resources in disaster mitigation?
3. What are the most active areas of geomorphic change where we might expect future avulsions?
4. Where should emergency relief go first in the case of a disaster and which type of people are most likely in need during those events?

Despite the insight produced by the exploratory models behind this report, this analysis is a foundational investigation and lays the groundwork for a fully functional, highly responsive, accessible, and comprehensive method for determining flood and landslide risk in this region. If fully developed and operationalized appropriately, our framework would essentially be a living window into vulnerability for the region as it most likely experiences increasing threats from extreme flooding and landslides. This dynamic approach could be “refreshed” after an extreme event, or as an analyst chooses to add new data, and adjust different model dimensions. The dynamism provided by the platforms used, and accessibility of underlying model code sets our approach apart from traditional risk modeling approaches. Traditional approaches yield static PDF maps, whose code and data repository system used to the calculate such outputs may remain inaccessible to local governments due to lack of computing, data management, or even technical skills necessary to update traditional risk models. We hope to pioneer our model as a new approach to responsive vulnerability modeling using Uttarakhand as a case study.

REFERENCES

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459.
- Abson, D. J., Dougill, A. J., & Stringer, L. C. (2012). Using principal component analysis for information-rich socio-ecological vulnerability mapping in Southern Africa. *Applied Geography*, 35(1), 515–524.
- Adger, W. N., Barnett, J., Brown, K., Marshall, N., & O'Brien, K. (2013). Cultural dimensions of climate change impacts and adaptation. *Nature Climate Change*, 3(2), 112–117. <https://doi.org/10.1038/nclimate1666>
- Ahmed, B. (2015). Landslide susceptibility mapping using multi-criteria evaluation techniques in Chittagong Metropolitan Area, Bangladesh. *Landslides*, 12(6), 1077–1095. <https://doi.org/10.1007/s10346-014-0521-x>
- Arvidson, T., Gasch, J., & Goward, S. N. (2001). Landsat 7's long-term acquisition plan—An innovative approach to building a global imagery archive. *Remote Sensing of Environment*, 78(1), 13–26.
- Asthana, A. K. L., & Asthana, H. (2014). Geomorphic control of cloud bursts and flash floods in Himalaya with special reference to Kedarnath area of Uttarakhand, India. *International Journal of Advancement in Earth and Environmental Sciences*, 2(1), 16–24.
- Bahinipati, C. S. (2014). Assessment of vulnerability to cyclones and floods in Odisha, India: a district-level analysis. *Current Science*, 107(12), 1997–2007.
- Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal*, 24(1), 43–69.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Brenkert, A. L., & Malone, E. L. (2005). Modeling vulnerability and resilience to climate change: a case study of India and Indian states. *Climatic Change*, 72(1–2), 57–102.
- Brunkard, J., Namulanda, G., & Ratard, R. (2008). Hurricane Katrina Deaths, Louisiana, 2005. *Disaster Medicine and Public Health Preparedness*, 2(04), 215–223. <https://doi.org/10.1097/DMP.0b013e31818aaaf55>
- Chin, A. (2006). Urban transformation of river landscapes in a global context. *Geomorphology*, 79(3), 460–487.
- Cohen, W. B., Fiorella, M., Gray, J., Helmer, E., & Anderson, K. (1998). An efficient and accurate method for mapping forest clearcuts in the Pacific Northwest using Landsat imagery. *Photogrammetric Engineering and Remote Sensing*, 64(4), 293–299.
- Coiner, J. C. (1980). Using LANDSAT to monitor changes in vegetation cover induced by desertification processes [West Africa; Mali]. In *Proceedings... International Symposium on Remote Sensing of Environment*. Retrieved from <http://agris.fao.org/agris-search/search.do?recordID=US19830841582>
- Coppin, P. R., & Bauer, M. E. (1994). Processing of multitemporal Landsat TM imagery to optimize extraction of forest cover change features. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4), 918–927.
- Crist, E. P., & Cicone, R. C. (1984). A physically-based transformation of Thematic Mapper data—The TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote Sensing*, 3(3), 256–263.

- Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly*, 84(2), 242–261.
- de Sherbinin, A. (2014). Climate change hotspots mapping: what have we learned? *Climatic Change*, 123(1), 23–37.
- Fekete, A. (2009). Validation of a social vulnerability index in context to river-floods in Germany. *Natural Hazards and Earth System Sciences*, 9(2), 393–403.
- Fothergill, A. (1996). Gender, risk, and disaster. *International Journal of Mass Emergencies and Disasters*, 14(1), 33–56.
- Fothergill, Alice, & Peek, L. A. (2004). Poverty and disasters in the United States: A review of recent sociological findings. *Natural Hazards*, 32(1), 89–110.
- Gannon, M. J. (1994). *Understanding global cultures: Metaphorical journeys through 17 countries*. SAGE Publications, Incorporated.
- George, S., Abel, R., & Miller, B. D. (1992). Female Infanticide in Rural South India. *Economic and Political Weekly*, 27(22), 1153–1156.
- Gleeson, T., Moosdorf, N., Hartmann, J., & van Beek, L. P. H. (2014). A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*, 41(11), 2014GL059856. <https://doi.org/10.1002/2014GL059856>
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... others. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160), 850–853.
- Holmes, R., Sadana, N., & Rath, S. (2010). Gendered Risks, Poverty and Vulnerability in India: Case Study of the Indian Mahatma Gandhi National Rural Employment Guarantee Act (Madhya Pradesh). *The Overseas Development Institute Research Report, October*. Retrieved from <https://www.odi.org/resources/docs/6254.pdf>
- IPCC. (2014). *Climate Change 2014–Impacts, Adaptation and Vulnerability: Regional Aspects*. Cambridge University Press. Retrieved from <https://books.google.com/books?hl=en&lr=&id=aJ-TBQAAQBAJ&oi=fnd&pg=PA1142&dq=Summary+for+Policymakers.+In:+Climate+Change+2014:+Impacts,+Adaptation,+and+Vulnerability.+Part+A:+Global+and+Sectoral+Aspects.&ots=v0RzOM54HF&sig=MAp9dTJL-dnthKOBuQuZDbOrITk>
- Irish, R. R. (2000). Landsat 7 automatic cloud cover assessment. In *AeroSense 2000* (pp. 348–355). International Society for Optics and Photonics. Retrieved from http://proceedings.spiedigitallibrary.org/data/Conferences/SPIEP/35040/348_1.pdf
- Irish, R. R., Barker, J. L., Goward, S. N., & Arvidson, T. (2006). Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. *Photogrammetric Engineering & Remote Sensing*, 72(10), 1179–1188.
- Islam, M. A., Chattoraj, S. L., & Ray, C. P. (2014). Ukhimath landslide 2012 at Uttarakhand, India: causes and consequences. *International Journal of Geomatics and Geosciences*, 4(3), 544.
- Jain, A. K. (1985). Determinants of Regional Variations in Infant Mortality in Rural India. *Population Studies*, 39(3), 407–424. <https://doi.org/10.1080/0032472031000141596>
- Kala, C. P. (2014). Deluge, disaster and development in Uttarakhand Himalayan region of India: Challenges and lessons for disaster management. *International Journal of Disaster Risk Reduction*, 8, 143–152. <https://doi.org/10.1016/j.ijdr.2014.03.002>

- Kavzoglu, T., Sahin, E. K., & Colkesen, I. (2014). Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, *11*(3), 425–439. <https://doi.org/10.1007/s10346-013-0391-7>
- Kline, M., & Cahoon, B. (2010). Protecting River Corridors in Vermont. *JAWRA Journal of the American Water Resources Association*, *46*(2), 227–236.
- Klinenberg, E. (2003). *Heat Wave: A Social Autopsy of Disaster in Chicago*. University of Chicago Press. Retrieved from <https://books.google.com/books?hl=en&lr=&id=r22xueipNegC&oi=fnd&pg=PR9&ots=NmQWOB-our&sig=kmkaKhHk8TUKFCBqTLIJwA198C8>
- Kumar, R., & Anbalagan, R. (2016). Landslide susceptibility mapping using analytical hierarchy process (AHP) in Tehri reservoir rim region, Uttarakhand. *Journal of the Geological Society of India; Bangalore*, *87*(3), 271–286. <https://doi.org/http://dx.doi.org/10.1007/s12594-016-0395-8>
- Leiserowitz, A., & Thaker, J. (2012). Climate change in the Indian mind. *Yale University, Yale Project on Climate Change Communication, New Haven, CT*.
- Mahapatra, M., Ramakrishnan, R., & Rajawat, A. S. (2015). Coastal vulnerability assessment using analytical hierarchical process for South Gujarat coast, India. *Natural Hazards*, *76*(1), 139–159. <https://doi.org/10.1007/s11069-014-1491-y>
- Maiti, S., Jha, S. K., Garai, S., Nag, A., Chakravarty, R., Kadian, K. S., ... Upadhyay, R. C. (2015). Assessment of social vulnerability to climate change in the eastern coast of India. *Climatic Change*, *131*(2), 287–306. <https://doi.org/10.1007/s10584-015-1379-1>
- Marjanović, M., Kovačević, M., Bajat, B., & Voženilek, V. (2011). Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology*, *123*(3), 225–234. <https://doi.org/10.1016/j.enggeo.2011.09.006>
- Martha, T. R., Roy, P., Govindharaj, K. B., Kumar, K. V., Diwakar, P. G., & Dadhwal, V. K. (2015). Landslides triggered by the June 2013 extreme rainfall event in parts of Uttarakhand state, India. *Landslides*, *12*(1), 135–146. <https://doi.org/10.1007/s10346-014-0540-7>
- Martha, T. R., van Westen, C. J., Kerle, N., Jetten, V., & Vinod Kumar, K. (2013). Landslide hazard and risk assessment using semi-automatically created landslide inventories. *Geomorphology*, *184*, 139–150. <https://doi.org/10.1016/j.geomorph.2012.12.001>
- Mathew, J., Babu, D. G., Kundu, S., Kumar, K. V., & Pant, C. C. (2014). Integrating intensity–duration-based rainfall threshold and antecedent rainfall-based probability estimate towards generating early warning for rainfall-induced landslides in parts of the Garhwal Himalaya, India. *Landslides*, *11*(4), 575–588. <https://doi.org/10.1007/s10346-013-0408-2>
- Mathew, J., Jha, V. K., & Rawat, G. S. (2009). Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. *Landslides*, *6*(1), 17–26. <https://doi.org/10.1007/s10346-008-0138-z>
- McCarthy, J. J. (2001). *Climate change 2001: impacts, adaptation, and vulnerability: contribution of Working Group II to the third assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=RT71Q24quc4C&oi=fnd&pg=PA1&dq=Impacts,+Adaptation+and+Vulnerability+&ots=osY9-ujnP-&sig=6V5ObZCNJZ5hOup_C4uA-wfKfWk

- Meyer, D. (2004). Support vector machines: The interface to libsvm in package e1071. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.492.7482>
- Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- National Academy of Sciences. (2012). *Disaster Resilience: A National Imperative*. Washington, D.C.: National Academies Press. Retrieved from <http://www.nap.edu/catalog/13457>
- National Alliance of Women. (2008). *Engendering the Eleventh Five-Year Plan 2007-2012: Removing Obstacles, Creating Opportunities*. New Delhi: NAWO.
- Neumayer, E., & Plümper, T. (2007). The Gendered Nature of Natural Disasters: The Impact of Catastrophic Events on the Gender Gap in Life Expectancy, 1981–2002. *Annals of the Association of American Geographers*, 97(3), 551–566. <https://doi.org/10.1111/j.1467-8306.2007.00563.x>
- Ngo, E. B. (2001). When disasters and age collide: Reviewing vulnerability of the elderly. *Natural Hazards Review*, 2(2), 80–89.
- NIDM. (2015). Uttarakhand Disaster 2013.
- Nisha, & Punia, M. (2014). Socio-economic vulnerability and sustainable development in context of development vs. conservation debate: A study of bhagirathi basin, uttarakhand, India (Vol. XL-8, pp. 77–84). Presented at the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. <https://doi.org/10.5194/isprsarchives-XL-8-77-2014>
- Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., ... Saleska, S. (2011). Height above the nearest drainage—a hydrologically relevant new terrain model. *Journal of Hydrology*, 404(1), 13–29.
- O’Hare, G. (2001). Hurricane 07B in the Godivari Delta, Andhra Pradesh, India: Vulnerability, mitigation and the spatial impact. *Geographical Journal*, 167(1), 23–38.
- Othman, A. N., Naim., W. M., M., W., & S., N. (2012). GIS Based Multi-Criteria Decision Making for Landslide Hazard Zonation. *Procedia - Social and Behavioral Sciences*, 35, 595–602. <https://doi.org/10.1016/j.sbspro.2012.02.126>
- Pande, R. K. (2010). Flash flood disasters in Uttarakhand. *Disaster Prevention and Management: An International Journal*, 19(5), 565–570.
- Pardeshi, S. D., Autade, S. E., & Pardeshi, S. S. (2013). Landslide hazard assessment: recent trends and techniques. *SpringerPlus*, 2(1), 523. <https://doi.org/10.1186/2193-1801-2-523>
- Peek, L. (2008). Children and disasters: Understanding vulnerability, developing capacities, and promoting resilience—An introduction. *Children Youth and Environments*, 18(1), 1–29.
- Pham, B. T., Pradhan, B., Tien Bui, D., Prakash, I., & Dholakia, M. B. (2016). A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environmental Modelling & Software*, 84, 240–250. <https://doi.org/10.1016/j.envsoft.2016.07.005>
- Poonam, Rana, N., Champati ray, P. K., Bisht, P., Bagri, D. S., Wasson, R. J., & Sundriyal, Y. (2017). Identification of landslide-prone zones in the geomorphically and climatically sensitive Mandakini valley, (central Himalaya), for disaster governance using the Weights of Evidence method. *Geomorphology*, 284, 41–52. <https://doi.org/10.1016/j.geomorph.2016.11.008>

- Pradhan, B. (2013). A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, *51*, 350–365.
- Raup, B., Kääb, A., Kargel, J. S., Bishop, M. P., Hamilton, G., Lee, E., ... others. (2007). Remote sensing and GIS technology in the Global Land Ice Measurements from Space (GLIMS) project. *Computers & Geosciences*, *33*(1), 104–125.
- Raup, B., Racoviteanu, A., Khalsa, S. J. S., Helm, C., Armstrong, R., & Arnaud, Y. (2007). The GLIMS geospatial glacier database: a new tool for studying glacier change. *Global and Planetary Change*, *56*(1), 101–110.
- Rautela, P. (2015). Traditional practices of the people of Uttarakhand Himalaya in India and relevance of these in disaster risk reduction in present times. *International Journal of Disaster Risk Reduction*, *13*, 281–290. <https://doi.org/10.1016/j.ijdr.2015.07.004>
- Rautela, P. (2016). Lack of scientific recordkeeping of disaster incidences: A big hurdle in disaster risk reduction in India. *International Journal of Disaster Risk Reduction*, *15*, 73–79.
- Ray-Bennett, N. S. (2009). The influence of caste, class and gender in surviving multiple disasters: A case study from Orissa, India. *Environmental Hazards*, *8*(1), 5–22. <https://doi.org/10.3763/ehaz.2009.0001>
- Regmi, N. R., Giardino, J. R., & Vitek, J. D. (2010). Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. *Geomorphology*, *115*(1), 172–187.
- Revelle, W. (2016). An overview of the psych package. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.190.7429&rep=rep1&type=pdf>
- Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83–98.
- Sati, S. P., Sundriyal, Y. P., Rana, N., & Dangwal, S. (2011). Recent landslides in Uttarakhand: nature's fury or human folly. *Current Science(Bangalore)*, *100*(11), 1617–1620.
- Seto, K. C., Woodcock, C. E., Song, C., Huang, X., Lu, J., & Kaufmann, R. K. (2002). Monitoring land-use change in the Pearl River Delta using Landsat TM. *International Journal of Remote Sensing*, *23*(10), 1985–2004.
- Simpson, J. J., & Stitt, J. R. (1998). A procedure for the detection and removal of cloud shadow from AVHRR data over land. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(3), 880–897.
- Singh, A. K., & Garia, P. S. (1999). Female work participation and involvement in decision-making process: A study in Uttarakhand. *Indian Journal of Agricultural Economics; Bombay*, *54*(3), 300.
- Sörensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences Discussions*, *10*(1), 101–112.
- Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS One*, *10*(2), e0107042.
- Sun, D., Yu, Y., & Goldberg, M. D. (2011). Deriving water fraction and flood maps from MODIS images using a decision tree approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *4*(4), 814–825.

- Syvitski, J. P., Overeem, I., Brakenridge, G. R., & Hannon, M. (2012). Floods, floodplains, delta plains—a satellite imaging approach. *Sedimentary Geology*, 267, 1–14.
- Tate, E. (2012). Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. *Natural Hazards*, 63(2), 325–347. <https://doi.org/10.1007/s11069-012-0152-2>
- Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2013). Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology*, 504, 69–79.
- Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014a). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology*, 512, 332–343.
- Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014b). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology*, 512, 332–343.
- Tellman, B., Alaniz, R., Rivera, A., & Contreras, D. (2014). Violence as an obstacle to livelihood resilience in the context of climate change. *United Nations University-Institute for Environment and Human Security*, 3. Retrieved from <https://works.bepress.com/ralaniz/10/>
- The Rockefeller Foundation. (2008). Asians Cities Climate Change Resilience Center (ACCCS). Retrieved from <http://www.acccrn.org/>
- Tien Bui, D., Pradhan, B., Lofman, O., & Revhaug, I. (2012). Landslide Susceptibility Assessment in Vietnam Using Support Vector Machines, Decision Tree, and Naive Bayes Models. *Mathematical Problems in Engineering*, 2012, e974638. <https://doi.org/10.1155/2012/974638>
- Tofani, V., Segoni, S., Agostini, A., Catani, F., & Casagli, N. (2013). Use of remote sensing for landslide studies in Europe. *Nat Hazards Earth Syst Sci*, 13(1), 299–309.
- Vapnik, V. N., & Kotz, S. (1982). *Estimation of dependences based on empirical data* (Vol. 40). Springer-Verlag New York. Retrieved from <https://hal.inria.fr/docs/00/04/54/30/PDF/tel-00003024.pdf>
- Vermote, E., & Saleous, N. (2007). LEDAPS surface reflectance product description. *College Park: University of Maryland Department of Geography*. Retrieved from <https://dwrgis.water.ca.gov/documents/269784/4654504/LEDAPS+Surface+Reflectance+Product+Description.pdf>
- Werg, J., Grothmann, T., & Schmidt, P. (2013). Assessing social capacity and vulnerability of private households to natural hazards – integrating psychological and governance factors. *Nat. Hazards Earth Syst. Sci.*, 13(6), 1613–1628. <https://doi.org/10.5194/nhess-13-1613-2013>
- World Bank. (2008). *Gender in Agriculture Sourcebook*. World Bank.
- Xian, G., Homer, C., Dewitz, J., Fry, J., Hossain, N., & Wickham, J. (2011). Change of impervious surface area between 2001 and 2006 in the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 77(8), 758–762.
- Yenneti, K., Tripathi, S., Wei, Y. D., Chen, W., & Joshi, G. (2016). The truly disadvantaged? Assessing social vulnerability to climate change in urban India. *Habitat International*, 56, 124–135. <https://doi.org/10.1016/j.habitatint.2016.05.001>
- Zhang, Y., Rossow, W. B., Lacis, A. A., Oinas, V., & Mishchenko, M. I. (2004). Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global

data sets: Refinements of the radiative transfer model and the input data. *Journal of Geophysical Research: Atmospheres*, 109(D19). Retrieved from <http://onlinelibrary.wiley.com/doi/10.1029/2003JD004457/full>

Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, 83–94.

APPENDICES

Appendix 1. Geomorphic Change GEE Source Code

Pre-Flood Landsat 7 ETM+ Land Classification

<https://ee-api.appspot.com/3376d35d615e542626a2364211e13192>

Post-Flood Landsat 8 OLI+ Land Classification

<https://ee-api.appspot.com/d0c716204e4ca311e7b0965c78719253>

Geomorphic Change

<https://ee-api.appspot.com/d34d4d3e3b712660971e36a70de03bde>

Appendix 2. Indian Social, Economic, and Infrastructure Variables

S. No.	Theme	Subtheme	Variable	Sub-variable
1	Urban Extent	Administrative	Area	
2	Urban Extent	Administrative	Population	
3	Urban Extent	Administrative	Population Density	
4	Urban Extent	Built Up	Area	
5	Urban Extent	Built Up	Population	
6	Urban Extent	Built Up	Population Density	
7	Urban Extent	Lit at night	Area	
8	Urban Extent	Lit at night	Population	
9	Urban Extent	Lit at night	Population Density	
10	Demographics	Age and Sex	Gender ratio	All ages
11	Demographics	Social Background	Scheduled Caste	
12	Demographics	Social Background	Scheduled Tribe	
13	Jobs	Labor Force	Working Population (7 + years)	Total
14	Jobs	Labor Force	Working Population (7 + years)	Female
15	Jobs	Labor Force	Working Population (7 + years)	Male
16	Jobs	Labor Force	Employment rate, 7+ years, main + marginal	Total
17	Jobs	Labor Force	Employment rate, 7+ years, main + marginal	Female
18	Jobs	Labor Force	Employment rate, 7+ years, main + marginal	Male
19	Jobs	Employment	Farmers	Total
20	Jobs	Employment	Farmers	Female
21	Jobs	Employment	Farmers	Male
22	Jobs	Employment	Self-employed	Total
23	Jobs	Employment	Self-employed	Female
24	Jobs	Employment	Self-employed	Male
25	Jobs	Employment	Casual wage earners	Total
26	Jobs	Employment	Casual wage earners	Female
27	Jobs	Employment	Casual wage earners	Male
28	Jobs	Employment	Regular wage earners	Total
29	Jobs	Employment	Regular wage earners	Female
30	Jobs	Employment	Regular wage earners	Male
31	Economic Activity	Nightlight Intensity	Light intensity per 1000 people	
32	Economic Activity	Nightlight Intensity	Light intensity per area	
33	Infrastructure	Connectivity	Road Length	Total
34	Infrastructure	Connectivity	Road Length	Major highways, primary and secondary
35	Infrastructure	Connectivity	Road Length	Tertiary and rural
36	Infrastructure	Connectivity	Road Length	Other
37	Infrastructure	Connectivity	Road Intensity	Total
38	Infrastructure	Connectivity	Road Intensity	Major highways, primary and secondary

39	Infrastructure	Connectivity	Road Intensity	Tertiary and rural
40	Infrastructure	Connectivity	Road Intensity	Other
41	Infrastructure	Connectivity	Number of Stations	Total
42	Infrastructure	Connectivity	Number of Stations	Railway
43	Infrastructure	Connectivity	Number of Stations	Metro
44	Infrastructure	Connectivity	Density of Stations	Total
45	Infrastructure	Connectivity	Density of Stations	Railway
46	Infrastructure	Connectivity	Density of Stations	Metro
47	Infrastructure	Energy	Households' access to electricity	Total
48	Infrastructure	Energy	Households' use of fuel for cooking	Biomass
49	Infrastructure	Energy	Households' use of fuel for cooking	Coal/lignite/charcoal
50	Infrastructure	Energy	Households' use of fuel for cooking	Kerosene
51	Infrastructure	Energy	Households' use of fuel for cooking	LPG/PNG
52	Infrastructure	Energy	Households' use of fuel for cooking	Electricity
53	Infrastructure	Energy	Households' use of fuel for cooking	Biogas
54	Infrastructure	Water and Sanitation	Households' access to improved water	Total
55	Infrastructure	Water and Sanitation	Households' access to improved sanitation	Total
56	Infrastructure	Water and Sanitation	Households' access to enhanced improved sanitation	Total
57	Infrastructure	Housing	Use of biomass to cook indoors	Total
58	Information Technology	Computer	Households' access to a computer with Internet	Total
59	Information Technology	Cellphone	Households' access to cellphone	Total
60	Finance	Access	Households' access to banking services	Total
61	Living Standards	Wealth	Access to all key durable assets	
62	Living Standards	Wealth	Lack of access to durable assets	
63	Education	Attainment	Literacy rate, 7+ years	Total
64	Education	Attainment	Literacy rate, 7+ years	Female
65	Education	Attainment	Literacy rate, 7+ years	Male

ⁱ Physical hazards combine with existing vulnerabilities to create a disaster. Resilience, borrowed from ecology, broadly refers to the ability of a system to recover after shock (Holling, 1973; Pimm 1993), as opposed to vulnerability, which is typically employed to identify specific social conditions pre-disaster and define post-disaster impacts. The IPCC defines resilience as “The capacity of social, economic, and environmental systems to cope with a hazardous event or trend or disturbance, responding or reorganizing in ways that maintain their essential function, identity, and structure, while also maintaining the capacity for adaptation, learning, and transformation” (IPCC AR5 WGII). As the IPCC vulnerability definition includes “capacity to cope and adapt”, it accounts for what many mean by resilience.

ⁱⁱ However, these different vulnerabilities are intertwined, and their distinctions and relationships are not clearly determined in the literature.

ⁱⁱⁱ The more contextual dimensions of vulnerability, the third category presented above, do not lend themselves to generalizable proxies like census data and are therefore cannot be full captured by a quantitative index.