Landscape Ecol (2007) 22:323–331
DOI 10.1007/s10980-006-9062-6

REPORT

Using landscape analysis to assess and model tsunami damage in Aceh province, Sumatra

Louis R. Iverson · Anantha M. Prasad

Received: 10 April 2006 / Accepted: 22 October 2006 / Published online: 8 December 2006

Abstract The nearly unprecedented loss of life resulting from the earthquake and tsunami of December 26, 2004, was greatest in the province of Aceh, Sumatra (Indonesia). We evaluated tsunami damage and built empirical vulnerability models of damage/no damage based on elevation, distance from shore, vegetation, and exposure. We found that highly predictive models are possible and that developed areas were far more likely to be damaged than forested zones. Modeling exercises such as this one, conducted in other vulnerable zones across the planet, would enable managers to create better warning and protection defenses, e.g., tree belts, against these destructive forces.

Keywords Tsunami damage · Prediction · Random forests · Indonesia · Forests · Developed areas · Modeling damage · Classification and regression trees · Tsunami warning · Mangroves

Introduction

The Indian Ocean earthquake and tsunami disaster of 26 December, 2004 claimed nearly 275,000 lives and destroyed billions of dollars' worth of property (Barber 2005). One can scarcely comprehend the enormity of this disaster and its lasting impact on the affected countries and the world in general.

The development of models capable of assessing vulnerable locations before a tsunami hits could save countless lives (Geist et al. 2006). Most modeling exercises necessarily incorporate a complexity associated with the shape of the ocean floor and how it intersects with the coastline. For example, the Method of Splitting Tsunami (MOST) model accounts for tsunami generation and propagation in the ocean, followed by inundation on the land (Titov and Gonzalez 1997). However, these authors also state that at the time of their paper the inundation portion of the model was the least well developed partly due to poor bathymetric and topographic data. Titov and Synolakis (1998) estimated that 150-m resolution data were needed for "adequate prediction" and 50-m resolution for predicting extreme run-up. However, the data have since been improved. For example, the Shuttle Radar Topographic Mission of 2000 (http://www2.jpl.nasa.gov/srtm) has improved the topographic data available, especially for lesser-developed areas. Other studies, such as those by Borrero et al. (2003, 2004), have used the MOST model with success in Papua New Guinea and southern California.

L. R. Iverson (✉) · A. M. Prasad
United States Department of Agriculture, Forest Service, Northern Research Station, 359 Main Road, Delaware, OH 43015, USA
e-mail: liverson@fs.fed.us

Other studies use empirical data from past tsunamis to estimate the size and location of vulnerable sites (e.g., Kulikov et al. 2005 for Peruvian tsunamis). This is the general approach taken in the current study. Our objectives are to use the damage information gleaned from the Indian Ocean tsunami along with good, high resolution data and a new statistical tool for modeling relationships among variables to better understand the core relationships and vulnerabilities along the coast of Sumatra and ultimately elsewhere. We also are assessing the relative protective value of forests along the coast. While the role of trees and mangroves in mitigating damage from tsunamis is not new, studies such as ours may be able to spatially delineate zones where damage is most likely to occur and hence equip managers with better information while planning protective defenses.

Methods

Data collection and preparation

The following data sets, primarily from the geographic area 95–96°E and 4.5–5.5°N, were used in this analysis: (1) damage polygons (as defined by interpreters of high resolution imagery when it was clearly visible where the water damaged developed and forested areas) as acquired by the United States in the initial days following the disaster (Figs. 1, 2a); (2) elevation at 30-m resolution (SRTM shuttle missions) (Fig. 2b); (3) Landsat GeoCover "natural view" 15-m Thematic Mapper provided by Earth Satellite Corporation (http://www.earthsat.com); and (4) boundaries of Banda Aceh Province. All data were projected to Universal Transverse Mercator Zone 47 at 30-m raster resolution. Processing was done in ERDAS (gis.leica-geosystems.com), ArcGIS, and Grid (http://www.esri.com).

We limited our zone of interest to the 10 km of land next to the coastline and two zones of analysis: (1) a small region representing the zone receiving a direct hit from the tsunami and (2) a large region representing all of the small region and extending down the coastline in both directions to encompass all of the damage polygons of Aceh (Fig. 1). Since we were interested in tsunami damage only and did not want to process excessive pixels of no damage, this band accommodated all of the damaged locations except for a few small locations further inland near Banda Aceh. The number of 30-m pixels was high, with 2.2 million of damage/nodamage in the small zone, and 8.4 million in the larger zone.

The coastline was developed from Shuttle Radar Topography Mission (SRTM, srtm.usgs.gov) 30-m data, with some adjustment from Indonesia coastline data where elevations were coded as 0 at the coast. The resulting linear coastline was then buffered by 10 km inland to clip out the study area.

Distance to shore was calculated using the linedist (GRID) command on the shoreline established previously. Output was a 30-m cell grid for the 10-km inland strip generated earlier (Fig. 2c). The exposure to ocean was created by conducting a search around each cell (1 km radius) and then counting the number of 30-m ocean cells within the radius (using GRID's focalsum). The highest numbers (values ranging up to 2821) were in the small peninsulas jutting into the sea, while the large zone more than 1 km from the coast had a value of 0 (Fig. 2d).

The natural view data, at 15-m resolution, were extracted for the study area from four scenes. Data were derived from bands 7, 4, and 2 and were collected in 1999–2000. We separated the Landsat data into 100 classes using an unsupervised classification technique in ERDAS (gis.leica-geosystems.com). We then reclassified into eight classes: water, wetland, built/barren, agriculture/low vegetation, low forest, forest, cloud, and cloud shadow (Fig. 2e). Built and barren were combined as representing very low vegetation classes; less than 2% of the small area and a slightly larger percentage of the large area were under cloud. The classified image was reprojected to 30-m cells. Cloud- and shadow-covered areas were eliminated from the predictive modeling. Though spot verification was achieved using higher resolution imagery (IKONOS and SPOT), and visual checking was widespread, no formal classification accuracy assessment or ground truth assessment was performed. Nonetheless, the above classes were general and distinct enough to instill confidence in the

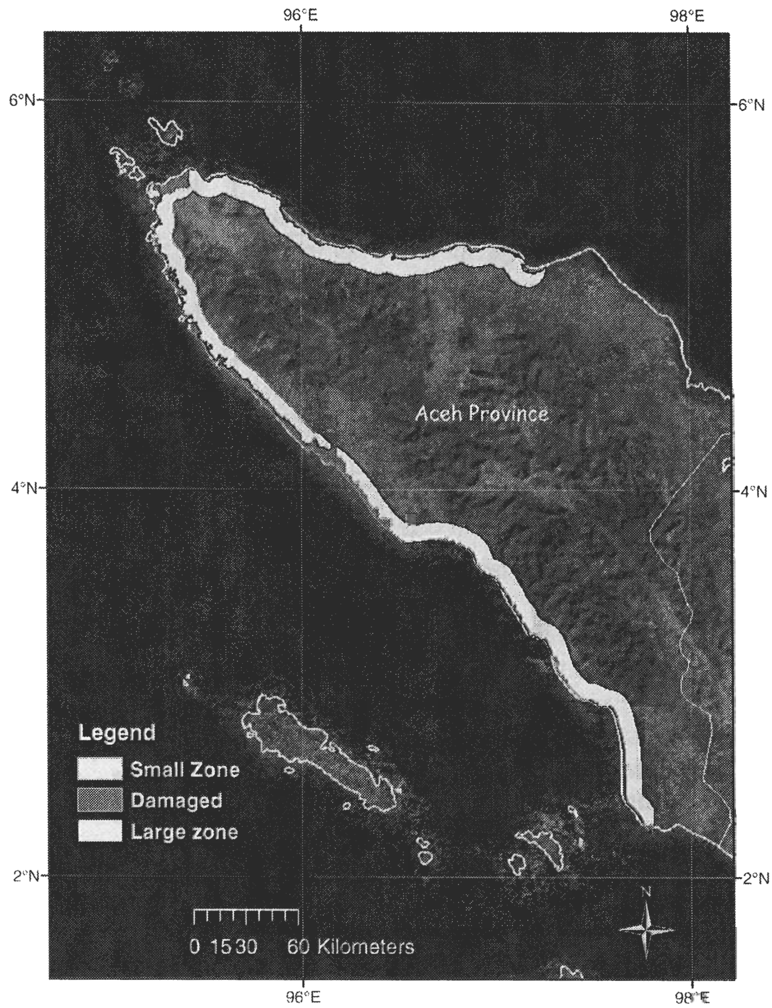

Fig. 1 Zone of analyses for large and small study areas around the coastline of Aceh

classification. No information was available on the land changes that occurred between 2000 and the time of the tsunami in 2004.

Assessment of forest vs. developed land

To compare the relative impact of damage to forest vs. developed land due to the tsunami, the classes were condensed to: (1) forested, from the low forest and forest classes, and (2) developed, from the built/barren and agriculture/low vegetation classes. The other classes were ignored for this cross tabulation analysis. Ratios of damaged to undamaged land areas were calculated for each class, and for the small and large study areas. We then compared these ratios for forested vs. developed regions.

Model of damage

The model was built for the small area using Random Forests (RF) package in R (R Development Core Team 2004; Breiman 2001; Prasad et al. 2006). RF is a new data-mining technique designed to produce very accurate predictions that do not overfit the data (Breiman 2001). RF uses bootstrap sampling of two-thirds of the data to construct multiple trees; each tree is also grown with a randomized subset of predictor variables (in our case 2 out of the 4 variables were selected for each perturbed tree). In RF, a very large number of trees (500 in our case) are grown (hence a 'forest' of trees) and averaged to yield powerful predictions closer to the true error of the estimated population rather than just the

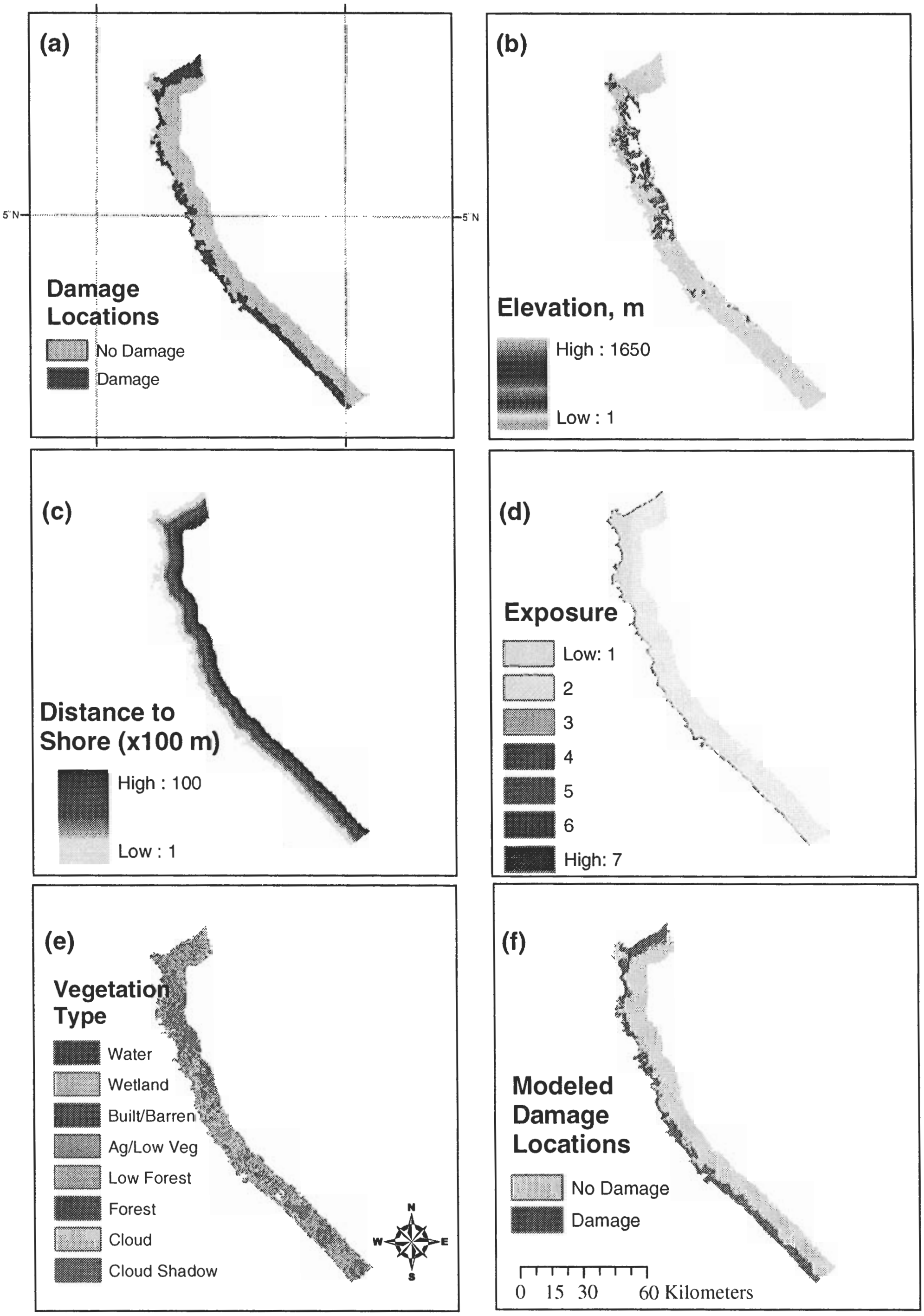

Fig. 2 Model input variables (for small zone): (**a**) damaged locations as remotely sensed by the U.S. Government, (**b**) elevation (units = meters), (**c**) distance to shore (units = 100-m intervals), (**d**) exposure (1 = no exposure, 10 = highest exposure; see text), (**e**) vegetation type, (**f**) modeled damage

training error. In all our statistics and mapping, we used the 'out-of-bag' outputs from RF, meaning the statistics were calculated from the third of the data not used to build the model, so that overfitting of the data is not a problem. We have thoroughly documented our procedure in Prasad et al. (2006).

The model was built for the small area only, as this zone had the most direct hit from the tsunami and represents the best location for validation as well. The resultant model was applied to both the small and large areas, to produce a vulnerability map for damage should they receive a direct hit from a large tsunami in the future. The total number of pixels for the larger area was divided up into 160 subunits and the saved 'forest' was applied to each pixel in a piecemeal fashion to each subunit before pasting them back together. This approach was essential due to memory limitations of RF when modeling millions of pixels of data. Predictor importance was evaluated by examining the mean decrease in accuracy based on random permutations of predictors using out-of-bag data, as output from the RF tool.

During the modeling process, we also evaluated two other data mining techniques that compete with RF for predictive accuracy. These are two adaptations of the boosting technique: See5 (http://www.rulequest.com) and Generalized Boosted Models (gbm) package in R (Ridgeway 2005). See5 uses an adaptive boosting technique for discovering patterns that delineate categories, assembling them into classifiers, and using them to make predictions (Schapire and Freund 1997). Gbm package implements boosting techniques using greedy function approximation and stochastic gradient boosting (Freidman 2001; Freidman 2002). After studying the outputs of both the boosting techniques and conducting several tests, we found they were inferior to RF for this study. We do not discuss the comparisons as it would fall outside the realm of this study.

The modeled outputs and damage maps were compared using the Map Comparison Kit 3.0 (http://www.riks.nl/mck, Hagen-Zanker et al. 2005). Kappa statistics including K_{histo} and K_{loc} were calculated, as was the percentage classified correctly by the model. Kappa is the product of K_{histo} and K_{loc}, where K_{histo} depends only on the total number of cells taken in by each category (i.e., only the histogram is considered, not the spatial distribution of the cells). K_{loc}, on the other hand, depends on the spatial distribution of the categories on the map (Pontius 2000; Hagen 2003). Kappa is the proportion of agreement after chance agreement—the percentage of agreement expected after randomly relocating all cells in the maps—has been removed (Monserud and Leemans 1992). It ranges from –1 to 1, with 1 being perfect agreement. We also calculated percent disagreement due to quantity and location according to Pontius (2000).

Results

Assessment of forest vs. developed land

Analysis of damage:undamage ratios indicate that relatively higher proportions of the developed land were damaged compared to forested lands (Table 1). The small region had twice as much forest land as developed land (957 km^2 vs. 471.5 km^2) but 20% less damaged area. Thus, developed land was 2.5 times more likely to incur damage than forested land (Table 1). For the large region, we calculate twice the likelihood of damage on developed lands relative to forested lands.

Model of damage, small region

According to U.S. Government estimates from the air around the time of the disaster, this region

Table 1 Area of land forested vs. developed with or without damage from the tsunami (km^2)

Land cover	Small zone		Large zone	
	No damage	Damage	No damage	Damage
Forested	957.0	162.4	2699.6	198.3
Developed	471.5	203.5	1784.5	264.4
Ratio—Damage:No Damage				
Forested	0.17		0.07	
Developed	0.43		0.15	
Ratio				
Developed: Forested	2.54		2.02	

had 450 km^2 or 22.5% percent damage in the 10-km strip along its 187 km of coastline. The RF model slightly underestimated the area affected at 21.7% (Table 2).

The RF model for damage recorded an out-of-bag error rate of 6.02%. The percentage disagreement due to quantity was 0.8%, while that due to location was 5.2% (Table 2). Most of the disagreement occurred approximately 2–4 km inland, outside the zones of most severe damage, and particularly in the northern regions that include Banda Aceh. The damage model on the smaller region classified 94% of the pixels correctly, with a cell-by-cell Kappa statistic of 0.824 (a result of K_{loc} of 0.844 and K_{histo} of 0.977) (Table 2). According to Monserud and Leemans (1992), the agreement between maps is "excellent" (Fig. 2a vs. Fig. 2f).

With respect to variable importance, elevation and distance to shore had almost identical contribution to the model according to the Mean Decrease in Accuracy statistic of the RF software (0.1647 vs. 0.1643), followed closely by vegetation, with 0.1589, and then exposure with 0.1354 (Table 2). The lesser importance for exposure can be attributed to the narrow zone (<1 km) of influence along the coastline (Fig. 2d).

Table 2 Statistics of areal estimates, coastal lengths, variable importance and classification accuracy using the RF software for the small zone and large zone

Variable	Small zone	Large zone
US government data		
Area (km^2)	2000	6645
Damaged area (km^2)	450	886
Damaged area (%)	22.5	13.3
Coastline length (km)	187	730
Modeled data		
Correctly classified (%)	93.9	Na
Classified damaged (km^2)	435	1632
Classified damaged (%)	21.7	24.6
Kappa statistic	0.824	Na
Klocation	0.844	Na
Khisto	0.977	Na
Kappa class	Excellent	Na
Disagreement, quantity (%)	0.8	
Disagreement, location (%)	5.2	
Variable importance		
Elevation	0.1647	Na
Distance to shore	0.1643	Na
Vegetation type	0.1589	Na
Exposure	0.1354	Na

Na = Not applicable

Model of damage vulnerability, large region

For the larger region, the total area of damage, as estimated by the U.S. Government, was 886 km^2 or 13.3% of the 10-km strip along the nearly 730 km of coastline (Table 2). Our model of vulnerability for this area classified a total of 24.6%, or 1632 km^2, as vulnerable should they receive a direct hit from another large tsunami (Table 2, Fig. 3). Of course, this is nearly twice the area than was actually damaged in the 2004 tsunami. Of the zone not in direct line with the tsunami, only 8.9% of the area and only 37.8% of the vulnerable locations were damaged. The model predicted vulnerable zones for portions of the northeastern and southwestern sections of the larger study area that were not actually damaged in the 2004 tsunami because they were not in a direct line with the tsunami. These areas would be vulnerable, however, should another tsunami hit with a direct path into these shorelines. Note that the maps presented (Figs. 1–3) are small representations of very detailed maps (30 m resolution) that could be used for local analysis and planning.

Discussion

In this paper we reiterate two points that have been made elsewhere but bear substantiation and repeating due to the enormity of the devastation to human life: (1) forested vegetation near the coasts can dampen the impacts of tsunamis and storm surges; and (2) areas most susceptible to such damage are reasonably predictable (and thus impacts can be largely mitigated).

For the small region, the zone of direct hit from the tsunami waves, we found a 2.5 times greater likelihood of damage in developed areas vs. forested areas. A similar result was found by Danielsen et al. (2005) for the Cuddalore District in Tamil Nadu, India, where treed areas within 1000 m of the coast suffered much less damage than non-treed areas. We attributed some of the difference in damage between forested and

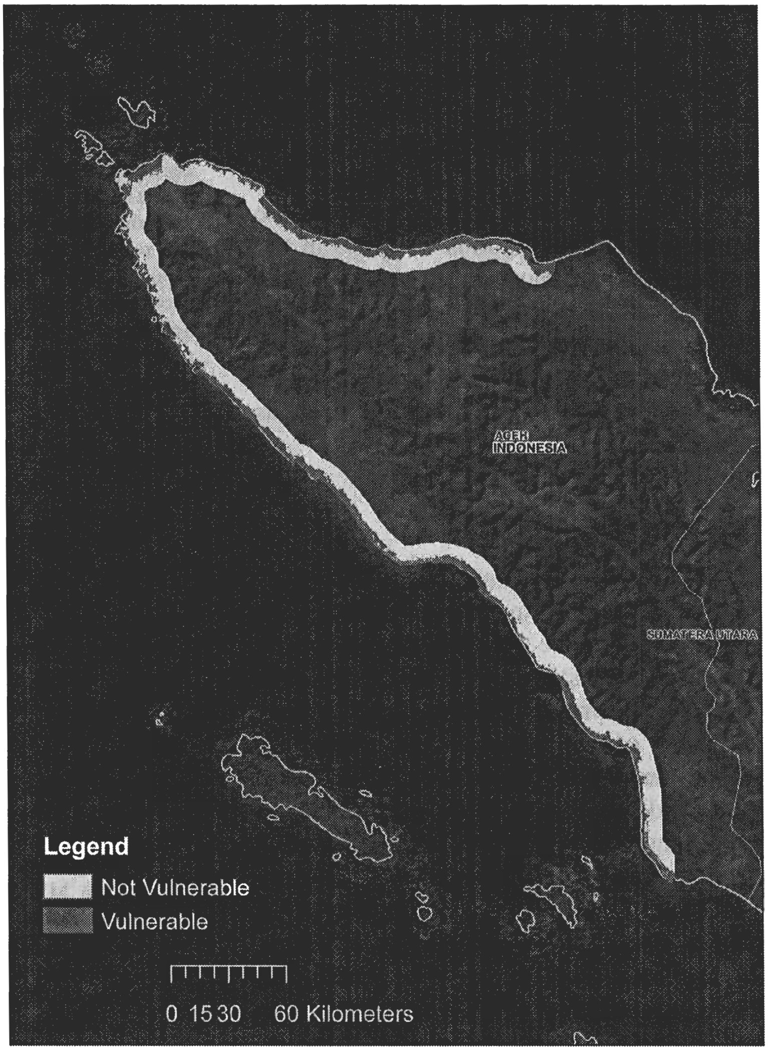

Fig. 3 Zone of tsunami vulnerability based on the model for actual damage to the small area and applied to the entire area

developed lands to the fact that developed lands tend to be in higher proportions at low elevations along the coast than farther inland or upland. However, much of the forest also is low-elevation, mangrove-type vegetation; we calculate that the land less than 20 m in elevation and less than 2 km from the coast was divided equally between forest and developed for the small region. Within this restricted elevation-distance zone, we also found that the ratio of damaged to undamaged land was 40% higher for developed than forested areas, providing further evidence of the protective nature of the forests. In recent decades, mangroves have been replaced, mostly by shrimp farms. According to a United Nations Food and Agriculture Organization report (2003), the area of mangroves was reduced by 26% between 1980 and 2000 for the five South-East Asian countries that were affected most by the tsunami.

Our models of predicting damage areas had "excellent" fit to the actual damaged areas, according to the Kappa statistics. Particularly for the small zone, which corresponded to the zone of direct tsunami hit, a reasonably accurate map of damage zones can be created, that is based primarily on distance to shore, elevation, and vegetation type (Fig. 2). The model was then extrapolated to a larger area, producing a map of vulnerability, at a resolution approaching 30 m (Fig. 3). These models may not be as mechanistically sophisticated as ones in which bathymetry and wave propagation through the seas are considered (e.g., Bhattacharjee 2005). Still, they are based on high-resolution empirical data that were not readily available previously. The data sources are now relatively easy and inexpensive to obtain over large areas of shoreline, as is the modeling capability. The RF software proved effective for this application as it allowed the use of both nominal and continuous variables in the classification. We are now in an era of data mining and predictive modeling that surpasses previous times (Iverson et al. 2004; Prasad et al. 2006; Elith et al. 2006), and these tools should be used for disaster preparation purposes such as demonstrated here.

Summary

We evaluated damage and built predictive models, at a resolution of 30 m, within 10 km of the coast for a small zone of 187 km along the western and northwestern coast, and a vulnerability map for a larger, 730-km coastline that encircles most of the province. In evaluating forested vs. developed zones, we found that developed land was much more susceptible to tsunami damage than forested land, 2.5 and 2 times greater likelihood, respectively, for the small and large zones. Though some portion of this result is due to relatively greater development in the near-coast areas, this result also provides further evidence of the protective power of coastal forests. We also created a RF model of damage for both zones based on elevation, distance to shore, vegetation type, and exposure to the ocean. The model for the small area correctly classified 94% of the pixels and had an "excellent" fit with actual damage, according to the Kappa statistic. The ranked importance of the variables overall in predicting damage were elevation = distance to shore > vegetation type >> exposure. The models can be applied elsewhere to assess vulnerabilities of landscape positions along coastlines. We recommend that managers responsible for coastal zones with some level of tsunami risk conduct similar modeling exercises to establish zones of vulnerability and consequent warning and protection defenses, e.g., tree belts, against these potentially devastating forces.

Acknowledgments This work was initiated while the senior author was on detail with the U.S. Agency for International Development in the aftermath of the tsunami. We are grateful to the many workers in USAID, the International Programs Office of the Forest Service, and other U.S., foreign, and United Nations agencies, universities, and corporations for their diligent and sincere efforts to save human lives and alleviate suffering. The impassioned sharing of data and expertise was remarkable. Special thanks to Rhonda D. Stewart and Dong Chaing for their great tutoring on data and procedures while at USAID. Thanks also to Mark Schwartz, John Peyrebrune, John Stanovich, and the reviewers/editors of this journal for improvements to the manuscript.

References

Barber B (2005) Tsunami relief. US Agency for International Development. Washington, D.C. 28 pp

Bhattacharjee Y (2005) In wake of disaster, scientists seek out clues to prevention. Science 307:22–23

Borrero J, Bu J, Saiang C, Uslu B, Freckman J, Gomer B, Okal E, Synolakis C (2003) Field survey and preliminary modeling of the Wewak Papua New Guinea Tsunami. Seismol Res Lett 74:393–405

Borrero JC, Legg MR, Synolakis CE (2004) Tsunami sources in the southern California bight. Geophys Res Lett 31(13):L13211

Breiman L (2001) Random forests. Mach Learn 45:5–32

Danielsen F, Sorensen MK, Olwig MF, Selvam V, Parish F, Burgess ND, Hiraishi T, Karunagaran VM, Rasmussen MS, Hansen LB, Quarto A, Suradiputra N (2005) Asian tsunami: a protective role for coastal vegetation. Science 310(Oct 28):643

Elith J, Graham CH, Anderson RP, Dudyk M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scahetti-Pereira R, Schapire RE, Sobero'n J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151

Food and Agriculture Organization (2003) State of the World's Forests. United Nations Food and Agriculture Organization, Rome

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232

Friedman JH (2002) Stochastic gradient boosting. Comput Stat Data Anal 38(4):367–378

Geist EL, Titov VV, Synolakis CE (2006) Tsunami: wave of change. Scientific American (January 2006):56–63

Hagen A (2003) Fuzzy set approach to assessing similarity of categorical maps. Int J Geogr Inform Sci 17:235–249

Hagen-Zanker A, Engelen G, Hurkens J, Vanout R, Uljee I (2005) Map Comparison Kit 2. User Manual. Research Institute for Knowledge Systems. Maastricht, The Netherlands. 120 pp

Iverson LR, Prasad AM, Liaw A (2004) New machine learning tools for predictive vegetation mapping after climate change: Bagging and Random Forest perform better than regression tree analysis. In: Smithers R (ed) Proceedings, UK-International Association for Landscape Ecology. Cirencester, UK, pp 317–320

Kulikov EA, Rabinovich AB, Thomson RE (2005) Estimation of tsunami risk for the coasts of Peru and Northern Chile. Nat Hazards 35(2):185–209

Monserud RA, Leemans R (1992) Comparing global vegetation maps with the Kappa statistic. Ecol Modell 62:275–293

Pontius RG (2000) Quantification error versus location error in comparison of categorical maps. Photogramm Eng Rem S 66:1011–1016

Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181–199

R Development Core Team (2004) R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. [online] URL: http://www.R-project.org

Ridgeway G (2005) Generalized boosted models: a guide to the gbm package. Available at http://www.i-pensieri.com/gregr/papers/gbm-vignette.pdf

Schapire R, Freund Y (1997) A decision theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55:119–139

Titov VV, Gonzalez FI (1997) Implementation and testing of the method of splitting tsunami (MOST) model. NOAA Technical Memorandum ERL PMEL-112. Pacific Marine Environmental Laboratory. Seattle, WA, 11 pp

Titov VV, Synolakis CE (1998) Numerical modeling of tidal wave runup. J Waterw Port C 124(4):157–171