3rd 2017 NASL

3rd North American Symposium on Landslides

June 4-8 2017, Roanoke, Virginia, USA

Effects of Inventory Bias on Landslide Susceptibility Calculations.

Stanley, T.A., thomas.a.stanley@nasa.gov,
Universities Space Research Association, Columbia, Maryland, 21046
Goddard Earth Sciences Technology and Research, Columbia, Maryland, 21046
Hydrological Sciences Laboratory, Goddard Space Flight Center, Greenbelt, MD, 20771

Kirschbaum, D.B., dalia.b.kirschbaum@nasa.gov
Hydrological Sciences Laboratory, Goddard Space Flight Center, Greenbelt, MD, 20771

ABSTRACT: Many landslide inventories are known to be biased, especially inventories for large regions such as Oregon's SLIDO or NASA's Global Landslide Catalog. These biases must affect the results of empirically derived susceptibility models to some degree. We evaluated the strength of the susceptibility model distortion from postulated biases by truncating an unbiased inventory. We generated a synthetic inventory from an existing landslide susceptibility map of Oregon, then removed landslides from this inventory to simulate the effects of reporting biases likely to affect inventories in this region, namely population and infrastructure effects. Logistic regression models were fitted to the modified inventories. Then the process of biasing a susceptibility model was repeated with SLIDO data. We evaluated each susceptibility model with qualitative and quantitative methods. Results suggest that the effects of landslide inventory bias on empirical models should not be ignored, even if those models are, in some cases, useful. We suggest fitting models in well-documented areas and extrapolating across the study region as a possible approach to modelling landslide susceptibility with heavily biased inventories.

INTRODUCTION

The spatial distribution of landslide activity is often described by empirically derived landslide susceptibility maps. Empirical methods are frequently used, due to the "objective" nature of statistical and machine-learning tools. However, the quality of the landslide inventory used to calibrate landslide susceptibility models can have a significant impact on the accuracy of the resulting map (Galli et al., 2008).

Both random and systematic errors in landslide inventories may affect the quality of landslide susceptibility maps (Steger et al., 2016a; Steger et al., 2016b). Landslide inventory bias can take several forms. Some biases may be introduced when an inventory is created from information available from a specific point in time, often in response to a single catastrophic event. Event inventories may be dominated by a specific movement type or triggering mechanism, even if other behavior is more common. Spatial biases may also be imposed by the size or direction of a single triggering event, with effects on the distribution of some explanatory variables. For example, aspect might be biased by the direction of storm winds (Bucknam et al., 2001).

Remote sensing of landslides is affected by many factors, including cloud cover, forest cover, and illumination (Guzzetti et al., 2012). Remotely sensed inventories may also suffer from land-cover biases (Steger et al., 2016a), such as the removal of old landslide scars by agricultural disturbance. The apparent landslide size distribution may be affected by surface cover or image resolution (Bucknam et al., 2001).

Archival inventories are known to have serious temporal biases (Guzzetti and Tonelli, 2004; Taylor et al., 2015). In addition, media reports and similar sources of information tend to focus on damages to populated areas and infrastructure. As a result, archival inventories often have a strong spatial bias toward developed areas (Guzzetti and Tonelli, 2004; Kirschbaum et al., 2015; Taylor et al., 2015).

Given the variety of biases that affect landslide inventories, empirical hazard models may be degraded by the presence of systematic error in the training data. Oregon, a large state with extensive landslide records, is an ideal testbed to determine the effects of inventory bias on landslide susceptibility calculations.

OREGON LANDSLIDE INVENTORIES

Terrain, seismicity, climate, and development vary widely across Oregon. These factors influence both occurrence and reporting of landslides, as evidenced by multiple landslide inventories. Oregon's population is largely concentrated in the Willamette valley. Frequent precipitation falls on western Oregon, due to the influence of the Pacific Ocean and the presence of major mountain ranges. In contrast, much of eastern Oregon consists of shrublands and desert. Tectonism has produced many volcanic peaks and soils, and the possibility of earthquakes and volcanism remains significant.

The Oregon Department of Geology and Mineral Industries (DOGAMI) produces a Statewide Landslide Information Database of Oregon (SLIDO) from geologic studies by several authoritative sources (Burns, 2014). SLIDO contains tens of thousands of historic and prehistoric landslides, stored as points, polygons, or scarp polylines. Nevertheless, it is not considered a complete record of recent landslide activity. Nor can SLIDO be considered an unbiased sample of Oregon landslides, because the scale, scope, and focus of early studies varies widely throughout the state (Burns, 2014). Most SLIDO entries are not associated with a specific date, but a broad date range is often available.

The Global Landslide Catalog (GLC) is not limited to Oregon, but it does contain hundreds of entries for the Pacific Northwest. The GLC includes landslides reported in news media, academic works, and other sources (Kirschbaum et al., 2010). All entries in the GLC are associated with a specific date, because the database was created to study the role of precipitation in triggering landslides.

As part of a NASA project for the National Climate Assessment (NCA), the GLC, SLIDO, and other sources of information were combined into a single Pacific Northwest Landslide Inventory (PNLI) for the purpose of identifying trends in landslide occurrence (Kirschbaum et al., 2016). As in the GLC, all entries used in this study were associated with a specific date, which limited the inventory to 3,366 landslide events in Oregon. The majority of these data points were derived from records maintained by the Oregon Department of Transportation (ODOT), resulting in a significant geographic bias near highways in the PNLI (Figure 1). Since this reporting bias may influence the conclusions of any research based on the PNLI, we analyzed the effects of the most obvious biases of this inventory on a commonly used empirical model.

METHODS

Synthetic landslide inventories

In order to evaluate the effects of reporting bias, it would be helpful to compare the effects of modeling with a biased dataset against the results arising from the use of unbiased data. Since no such landslide inventory exists for the whole of Oregon, we created a synthetic inventory as a proxy. Although the true locations of all historic landslides are rarely known, the spatial probability of occurrence is estimated by susceptibility maps such as the Landslide susceptibility overview map of Oregon (Burns et al., 2016). We selected 336,700 points at random from all locations in Oregon, but with the restriction that the distribution of landslide susceptibility at the random points match the distribution of landslide susceptibility in known historic landslides. These points formed the unbiased synthetic landslide inventory (Figure 1).

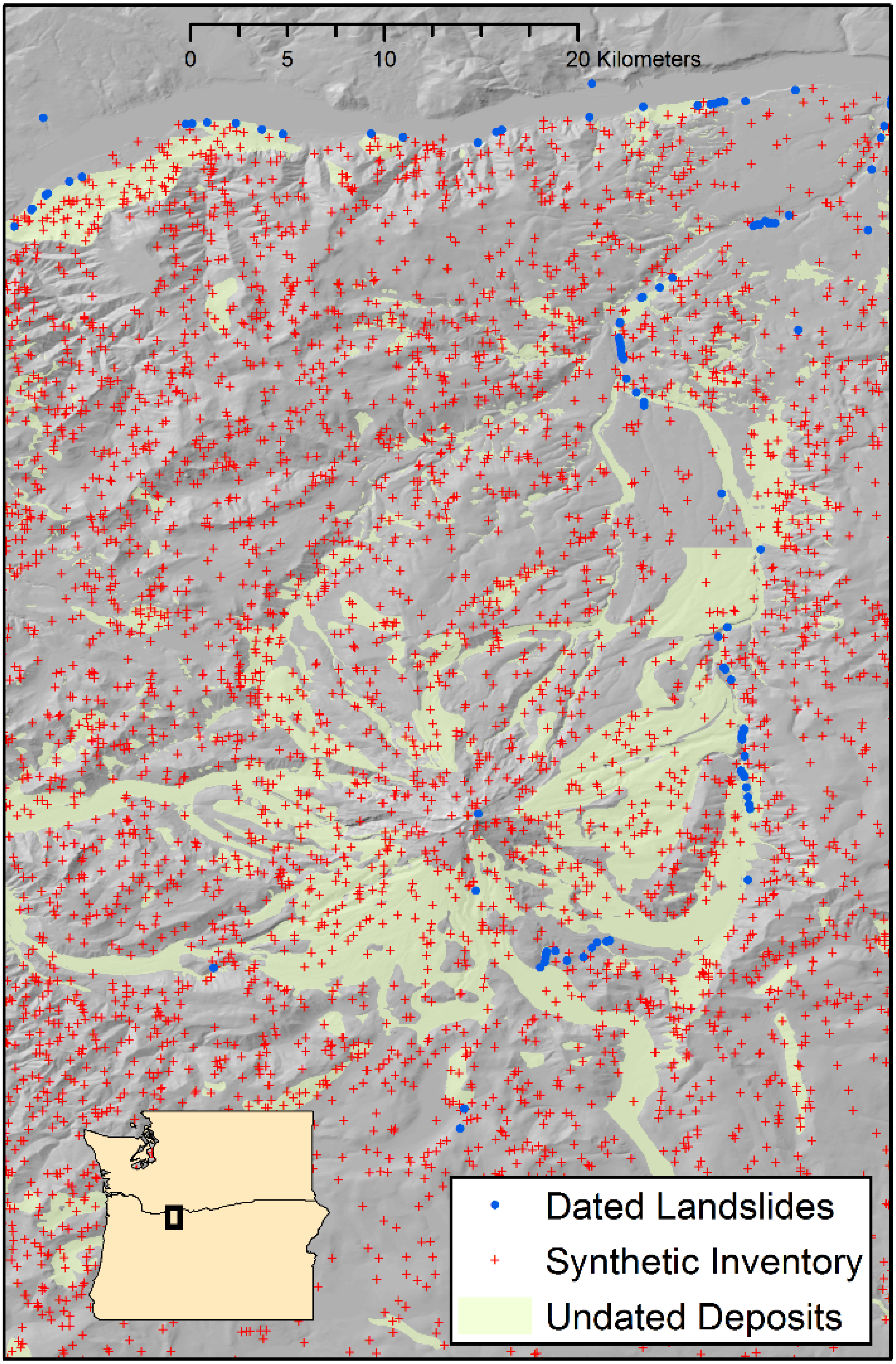

Figure 1. Landslide inventories near Mount Hood. Relatively few landslides were recorded with dates (blue), and most of these were located adjacent to roads. The landslide susceptibility overview map of Oregon was derived from the many landslides reported without exact dates (green). In turn, this map was used to generate a synthetic landslide inventory (red).

In order to simulate the effects of specific biases, the distribution of values within each biasing factor was determined. First, we calculated in ArcGIS 10.2 (ESRI, 2013) the distance from each point in the synthetic inventory to the nearest highway stored in the Global Roads Open Access Dataset (CIESIN, 2013). Second, the approximate population at each location was derived from LandScan data (Bright et al., 2015). Then, both steps were repeated for the PNLI. Compared to the synthetic inventory, the PNLI shows strong biases, with over 90% of all reported landslides within 1 kilometer of a major highway (Figure 2). Over 50% of the PNLI was reported in populated areas (Figure 3). Both of these factors may represent a mixture of reporting bias and slope instability due to human activity. However, the role of human actions, such as oversteepening slopes by excavation, is at least partially reflected in the 32.8-ft^2 elevation grid used to generate the susceptibility map and its derivative, the synthetic landslide inventory. Thus, reporting bias is the more important cause of the difference between distributions. The initiation points derived from undated landslide deposits shows distributions intermediate between the other inventories.

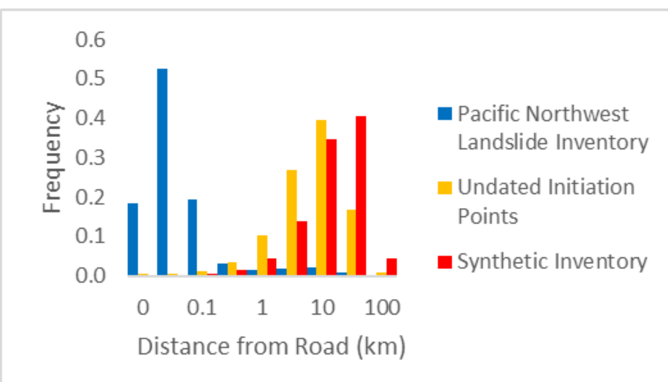

Figure 2. Distance to road from PNLI points (blue), undated initiation points (yellow), and unbiased synthetic inventory points (red).

Next, we created new versions of the synthetic landslide inventory by selecting data points to match the biased distributions observed in the PNLI. To create an inventory with a roads bias, the unbiased synthetic inventory was divided into bins, each representing a multiple of 1,000 meters of distance from a road. Data points were selected at random from each bin to create a sample of equal size to the same bin in the PNLI. The resulting biased inventory contained 3,366 points, the same as the PNLI. This process was repeated with the population variable to create a biased inventory with the same population distribution as the PNLI. Finally, both road and population biases were combined to form a fourth synthetic inventory.

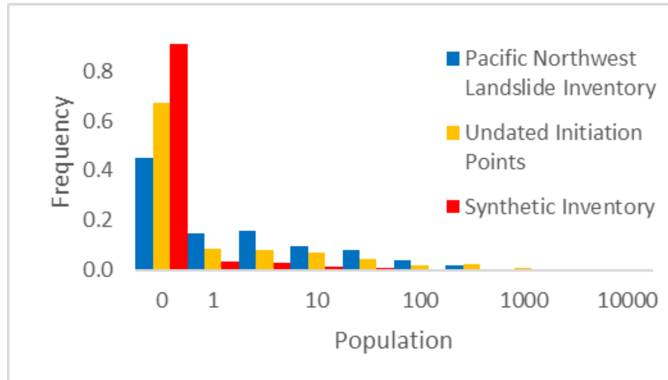

Figure 3. Estimated population at PNLI points (blue), undated initiation points (yellow), and unbiased synthetic inventory points (red).

Biasing the historical record

In addition to examining the effects of biasing a hypothetically complete inventory, we performed the same procedure on a relatively complete inventory: the undated polygons from SLIDO. This inventory contains 41,029 polygons, in a wide variety of sizes and types (Figure 1). In order to avoid fitting an empirical model on the basis of large runout zones, initiation points for these landslides were approximated by selecting the vertex with the highest elevation from each polygon. As with the synthetic inventory, the statewide database was then truncated to match the roads (Figure 2) and population (Figure 3) biases observed in the dated PNLI. The undated initiation points showed a distribution of attributes intermediate between the dated events and the synthetic inventory, so the subsequent truncation was less extreme.

Logistic regression

Numerous methods for calculating landslide susceptibility exist. Logistic regression has been applied for this purpose for over a decade (Atkinson and Massari, 1998), and its use has become common worldwide (Chau and Chan, 2005; Gorsevski et al., 2006; Ninu Krishnan et al., 2014; Brenning et al., 2015; Steger et al., 2016b). We selected this method to evaluate the effects of biased inventories on a standard approach to landslide susceptibility mapping. Logistic regression is a flexible technique that can combine continuous and categorical predictors (independent variables) into a statistical estimate of the probability of a response (dependent variable). In landslide susceptibility studies, the response is the presence of a landslide at a given location, and the model output is the spatial probability of occurrence. A critical assumption of the method is that the landslide inventory used for model fitting represents the accurate distribution of landslides. This assumption may be incorrect for a variety of reasons, including systematic error in the inventory.

Previous landslide susceptibility researchers have used a wide variety of predictors to predict the occurrence of landslides. However, some variables may be more strongly associated with inventory bias than true landslide occurrence, thus degrading the performance of a susceptibility map (Steger et al., 2016b). In order to avoid this problem, we omitted variables such as land use and road presence. The presence of burned areas was considered, but was not found to be statistically significant. In addition, the size and frequency of wildfires appears to be associated with mean annual precipitation, which was included in the model. Slope and geology were used to generate the DOGAMI map (Burns et al., 2016), and these factors have been used to make many landslide susceptibility maps. Because both tectonic and climatic influences have a role in landslide occurrence, distance to fault and mean annual precipitation were analyzed. Slope was generated from the National Elevation Dataset (U.S. Geological Survey, 2016) at a resolution of 30 meters. All other variables (Table 1) were rasterized to match this resolution.

Table 1. List of independent variables

Predictor	Source
Slope	(U.S. Geological Survey, 2016)
Distance to nearest fault	(Smith and Roe, 2015)
General geologic unit	(Smith and Roe, 2015)
Mean annual precipitation	(Oregon State University, 2016)

These variables were combined into a landslide susceptibility map for each version of the synthetic landslide inventory. In place of known locations without landslides, points were randomly assigned to this category, with the restriction that no "landslide points" be included. One fifth of each combined landslide/random point dataset was randomly selected and saved to validate the empirical fit of the other subsets. The logistic regression models were fitted and evaluated in R, a statistical software environment (Pebesma and Bivand, 2005; Sing, 2005; Bivand et al., 2015; Hijmans, 2015; R Core Team, 2015). Then the models were applied across Oregon to produce a series of landslide susceptibility maps with 1-arcsecond resolution.

Binning

Landslide susceptibility models with continuous outputs, such as the probabilities from logistic regression, are usually categorized into zones for the convenience of the user. Although the binning process is rarely the focus of research, many techniques for doing so have been implemented. Four or five bins are typically used (Das et al., 2010). We binned the continuous output of logistic regression into four categories: low (probability<0.07), moderate (0.07-0.17), high (0.17-0.25), and very high (>0.25).

RESULTS

Each version of the synthetic landslide inventory produced a distinct susceptibility map (Figure 4). The size of each susceptibility category was consistent, due in large part to the use of identical methods and predictor datasets (Table 2). Fitting a model to the synthetic inventories identified few locations as very high susceptibility, and a high proportion as low susceptibility. This may be caused by the low ratio of landslide points to random points, which was based on the area previously delineated as historic landslides (Burns et al., 2016). The undated landslide initiation points from SLIDO produced similar results (Figure 5). Table 2 also lists the proportions of each category shown in the Landslide Susceptibility Overview of Oregon, but these cannot be compared directly, due to the use of a different modeling approach, binning method, and data selection.

Table 2. Size of susceptibility categories by training dataset

Inventory	Low	Moderate	High	Very high
Synthetic Inventory				
Unbiased	80%	17%	2%	0%
Population Bias	81%	18%	1%	0%
Roads Bias	81%	16%	2%	0%
Both Biases	80%	19%	1%	0%
Undated Initiation Points (from SLIDO)				
Unbiased	82%	13%	3%	2%
Population Bias	82%	12%	3%	3%
Roads Bias	82%	14%	3%	2%
Both Biases	83%	12%	3%	3%
PNLI	82%	15%	2%	1%
Landslide Susceptibility Overview of Oregon				
SLIDO	*37%*	*28%*	*30%*	*5%*

Receiver operating characteristic analysis (ROC) showed little difference between biases (Table 3). The ROC is often used to evaluate landslide susceptibility models (Steger et al., 2016b), due to its insensitivity to class size. It is constructed by calculating the true positive rate (TPR) and false positive rate (FPR) for each possible threshold, then plotting all TPR-FPR pairs. The area under this curve (AUC) ranges from 0 to 1, with higher values indicating better model performance. Table 3 shows the AUC calculated for the fifth of each catalog that was not used to train the model. The PNLI performed better according to this metric, but this fact does not necessarily mean that the PNLI-based classification of terrain was more successful.

Table 3. AUC by validation dataset

Synthetic Inventory	AUC
Unbiased	0.74
Population Bias	0.71
Roads Bias	0.73
Both Biases	0.69
Undated Initiation Points	AUC
Unbiased	0.81
Population Bias	0.86
Roads Bias	0.82
Both Biases	0.85
PNLI	0.77

Visual comparison of the landslide susceptibility maps revealed much more (Figure 4). The unbiased synthetic inventory produced a map (Figure 4a) with a geographically broad distribution of highly and very highly susceptible locations. In contrast, the other maps, especially the population-biased map (Figure 4b), concentrated landslide susceptibility in the Coast Range. This is probably due to the association between climate and the pattern of human settlement in Oregon. The unbiased susceptibility map shows fewer artifacts associated with geologic units. The other models were fitted to inventories with constrained geographical distributions, with the result that large areas were identified as low susceptibility. All of the biased logistic regression models overemphasize geology, unlike the Landslide Susceptibility Overview Map of Oregon (Figure 4f). Population bias also appears to have increased the effect of faults in the Willamette Valley, which probably does not reflect a truly increased susceptibility to landslides. The substantial differences between these maps were not reflected in the validation statistics (Table 3); this discrepancy has been noted in previous studies (Steger et al., 2015; Steger et al., 2016b).

The distribution of susceptibility in maps derived from the undated initiation points was similar. The unbiased map concentrated susceptibility in the (Figure 5a). Population bias raised susceptibility (Figure 5b). Roads bias (Figure 5c). The combined bias (Figure 5d).

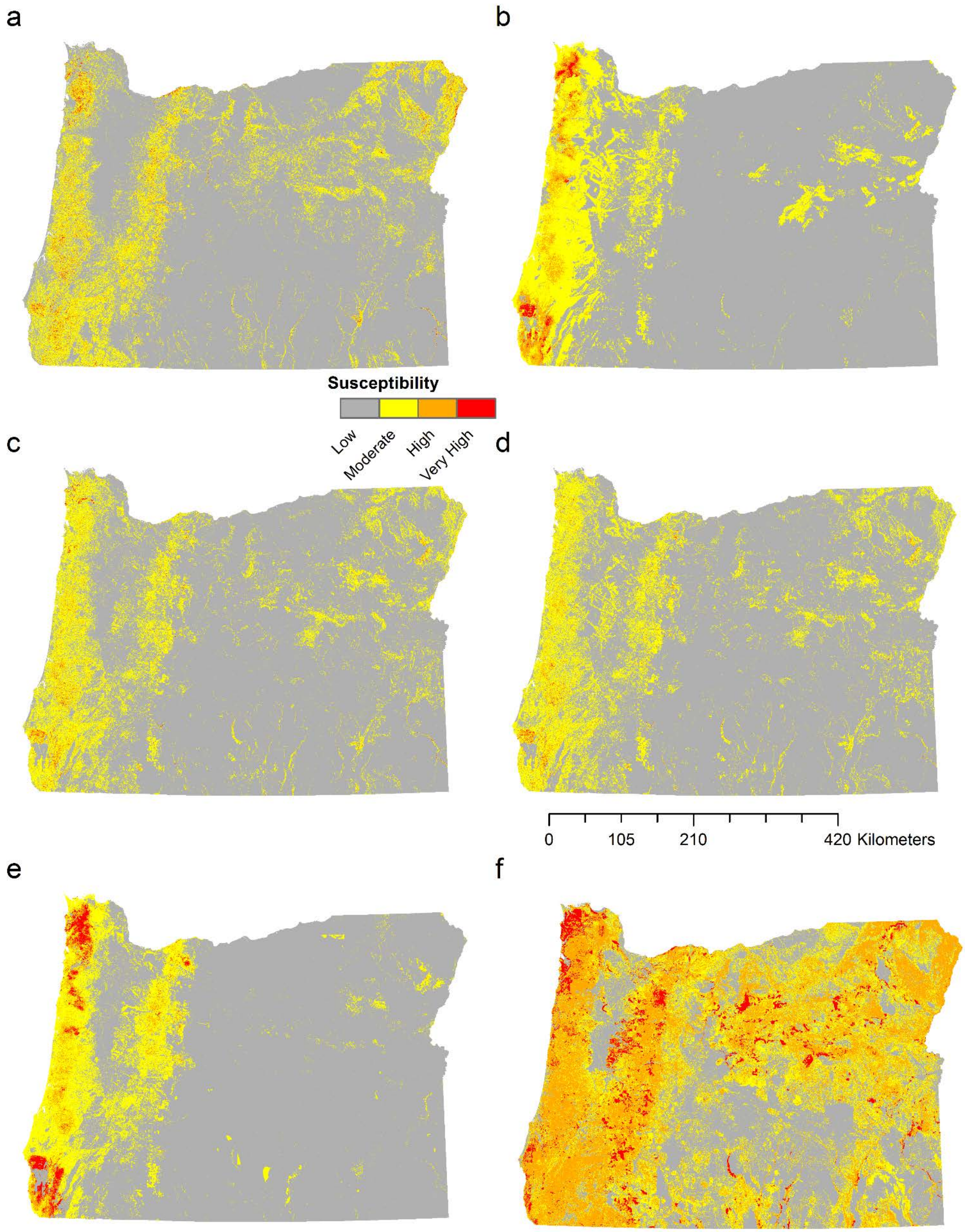

Figure 4. Landslide susceptibility calculated from various inventories: a. the synthetic inventory with no added biases; b. the synthetic inventory with a population bias added; c. the synthetic inventory with a roads bias added; d. the synthetic inventory with both roads and population biases; e. the PNLI; f. the Landslide Susceptibility Overview of Oregon; note that the categories in f have different definitions and are not directly comparable.

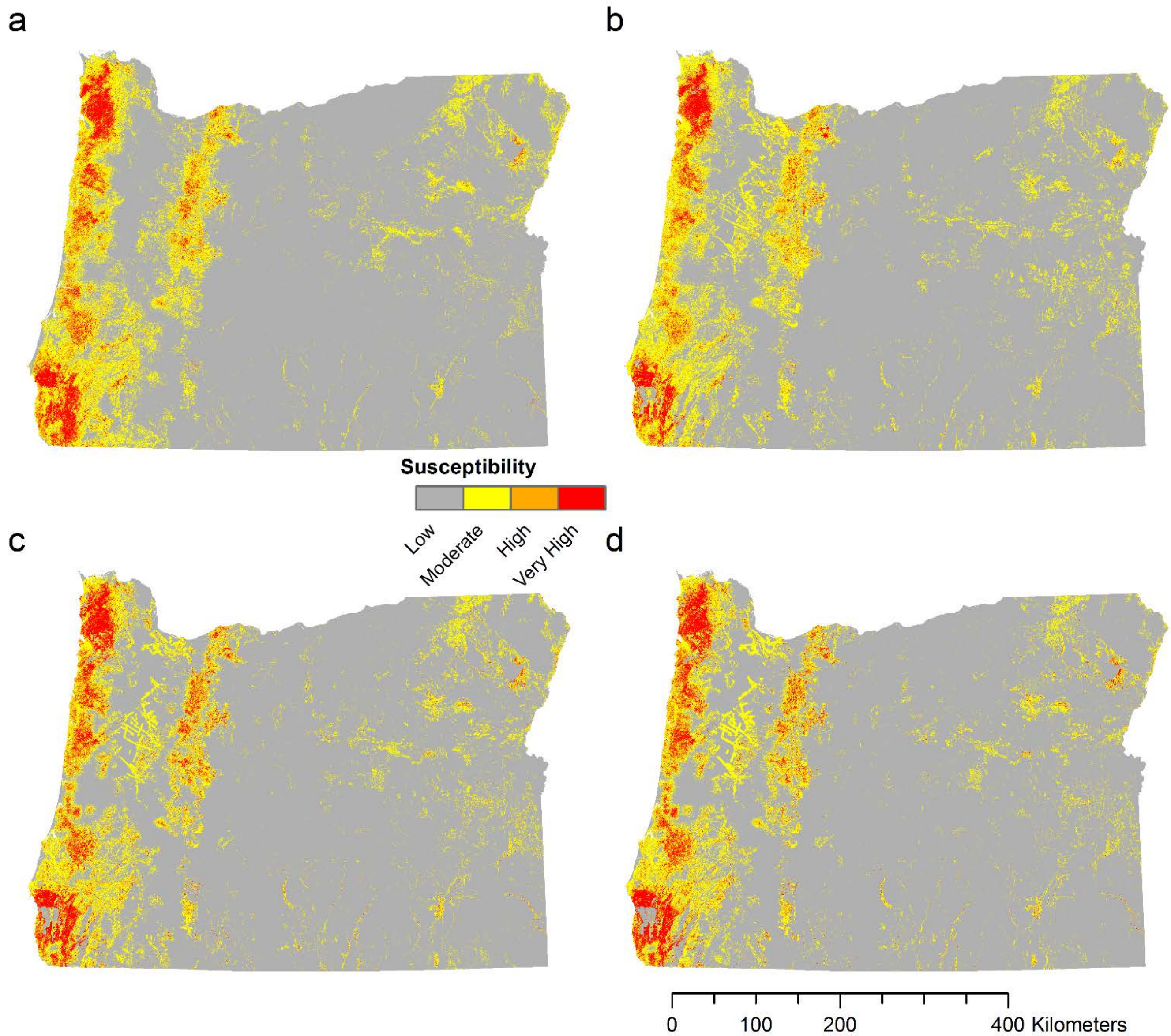

Figure 5. Landslide susceptibility calculated from various inventories: a. the undated initiation points with no added biases; b. the same inventory with a population bias added; c. the inventory with a roads bias added d. the inventory with both roads and population biases.

All maps shared some common attributes. The concentration of susceptibility in the west, and the low susceptibility of central and southeastern Oregon are due not only to reporting biases in the inventories, but also to the geography of the critical landslide triggering factors, seismicity and precipitation. Although roads were excluded from this study as a potentially biasing factor, road construction in Oregon may have destabilized some slopes (Montgomery, 1994).

Despite these similarities, the introduction of even one reporting bias into a landslide inventory can drastically alter the output of empirical models. When landslide training data is confined to lowland areas but non-landslide training data is ubiquitous, regression coefficients are likely to underestimate the effects of steep slope and similar variables. To avoid this problem, every landslide susceptibility map would ideally be derived from a complete and unbiased landslide inventory. In practice, this is only possible for small study areas. If empirical landslide susceptibility models are to be employed for large areas, a method for mitigating the effects of inventory bias must be applied.

BIAS MITIGATION

Landslide inventory bias clearly has an effect on these statistical models, potentially degrading the utility of the resulting susceptibility maps. Multiple techniques for mitigating this problem can be used. The first method is simply to rely upon prior knowledge, either as a physically-based model or as a heuristic model derived from expert opinion. However, these methods fail to take advantage of the important information contained in landslide inventories. Furthermore, physical models typically require calibration and thus may be influenced by any biases in the calibration inventory.

A second method is to explicitly include the biasing factors in an empirical model in such a way that the relationship between landslide occurrence and other predictors is estimated correctly. Multilevel statistical models have been used for this purpose at the local scale (Steger et al., 2016a), but it is unclear whether these models are appropriate for use over very large areas. A third method would be to determine the probability of landslide detection prior to empirical modeling. A correction could then be made to the training dataset, with the result that the trained model would not require any bias correction. However, it would be difficult to determine the non-detection probability for a complex, multisource inventory such as the PNLI.

A fourth method is simply to train the empirical model with data from areas that are well-documented, then apply the model regionally or globally. Limitations to this method include the possibility that some conditions, such as specific geologic units, will not exist within the training areas, and the possibility that landslide behavior is anomalous in training areas (e.g. construction projects alter drainage, which alters landslide susceptibility). While these concerns cannot be ignored, we believe that spatial extrapolation offered the best chance for empirical modeling of landslide susceptibility with a large and complex inventory. In order to fit a model to well-documented areas, the state of Oregon was divided into areas within 1 kilometer of any major highway, and more distant areas. The areas proximal to roads were then treated as a training dataset, while the distal areas were treated as a validation dataset. This division placed over 90% of the synthetic landslides in the training dataset, along with sections of every general geologic unit. Nevertheless, the training areas accounted for only 3% of Oregon, which should greatly reduce the influence of false negatives on the model fitting process. Figure 6 shows that the population density is much higher within the training area, which suggests that the population bias may also be addressed by this method. Areas near roads are typically flatter (Figure 7), but a wide range of values is still available to fit the model. Little difference can be seen in precipitation distributions, although there appear to be fewer roads in dry areas (Figure 8).

A logistic regression model was fitted to the training dataset, which consisted of all 3,128 synthetic landslides within the training area and 59,534 random points. The model was then applied to the entire state of Oregon. ROC analysis was performed on a validation dataset consisting of all 215 landslides located outside of the training area and 4,082 random points. The AUC was 0.69, which represents the likely outcome when evaluating with an incomplete inventory such as the dated portion of the PNLI. Evaluating with a more complete inventory, such as the unbiased synthetic

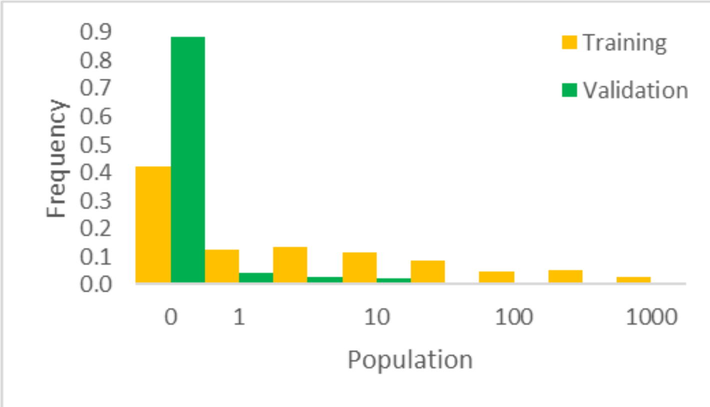

Figure 6. Estimated population within 1 km (yellow) and beyond 1 km (green) from major highways.

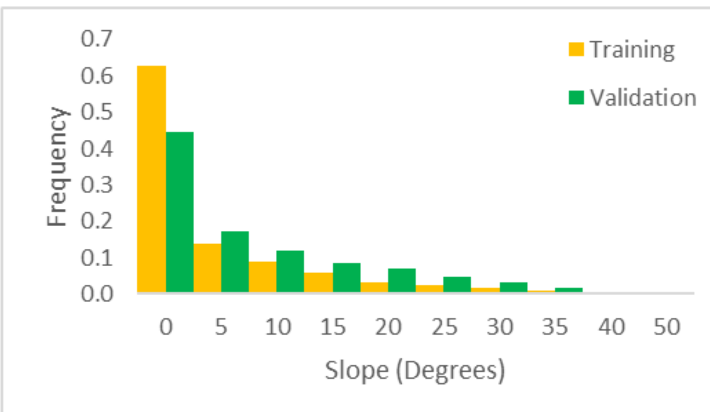

Figure 7. Slope distribution within 1 km (yellow) and beyond 1 km (green) from major highways.

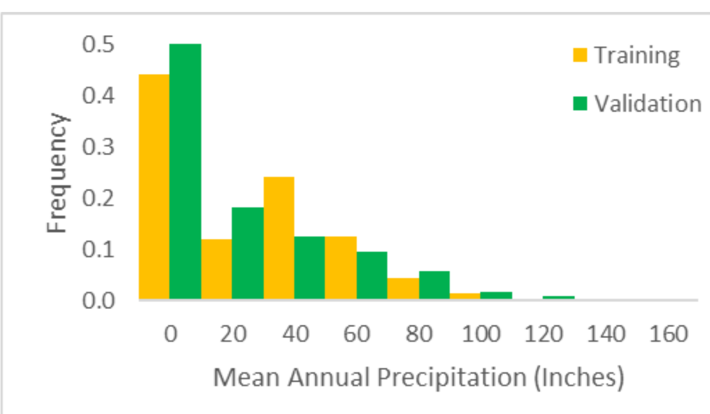

Figure 8. Distribution of mean annual precipitation near (yellow) and far (green) from major highways.

dataset, produced an AUC of 0.73. (329,640 landslides and 4,529,928 random points were located away from roads.)

This method produced a landslide susceptibility map with a plausible geographic distribution (Figure 9a). Very highly susceptible locations are most common in the western third of the state, where topography, seismicity and precipitation are more intense. Nevertheless, much of eastern Oregon is also susceptible, due to rugged terrain. The geologic unit “batholith rocks” was associated with low susceptibility, forming distinctive patches near Baker City. This unit was not spatially extensive enough to intersect many near-highway landslides, so the true landslide susceptibility may be higher. Other geologic units do not appear to exert such a strong influence. Unlike the results shown in Figure 4b-e, the population bias in this training dataset has not resulted in a concentration of nearly all highly susceptible pixels in the coast range. Instead, the bias-mitigated susceptibility resembles the “true” distribution of susceptibility (Figure 4f) from which the synthetic inventory was derived, with the highest susceptibility in the Coast Range and other concentrations in the Cascades and northeastern Oregon.

We repeated this process with 2,246 undated landslide initiation points, of which 1,824 were located near a highway, and 42,674 random points. The AUC was 0.82 when validated with the biased version of the inventory, and 0.81 with the full inventory. Susceptibility was distributed in a pattern similar to that derived from the synthetic inventory, but was higher overall (Figure 9b).

CONCLUSIONS

Experiments with synthetic and historic landslide inventories revealed that reporting biases can affect the accuracy of empirically derived susceptibility models. Population had a greater biasing effect than distance to road. We controlled other potential biases, such as linguistic or economic factors, by focusing on a single country. Future research would be required to address all the sources of bias in the GLC and similar landslide inventories. A simple strategy for working with biased inventories was tested. Spatial extrapolation performed well and is recommended as a first approach to modeling

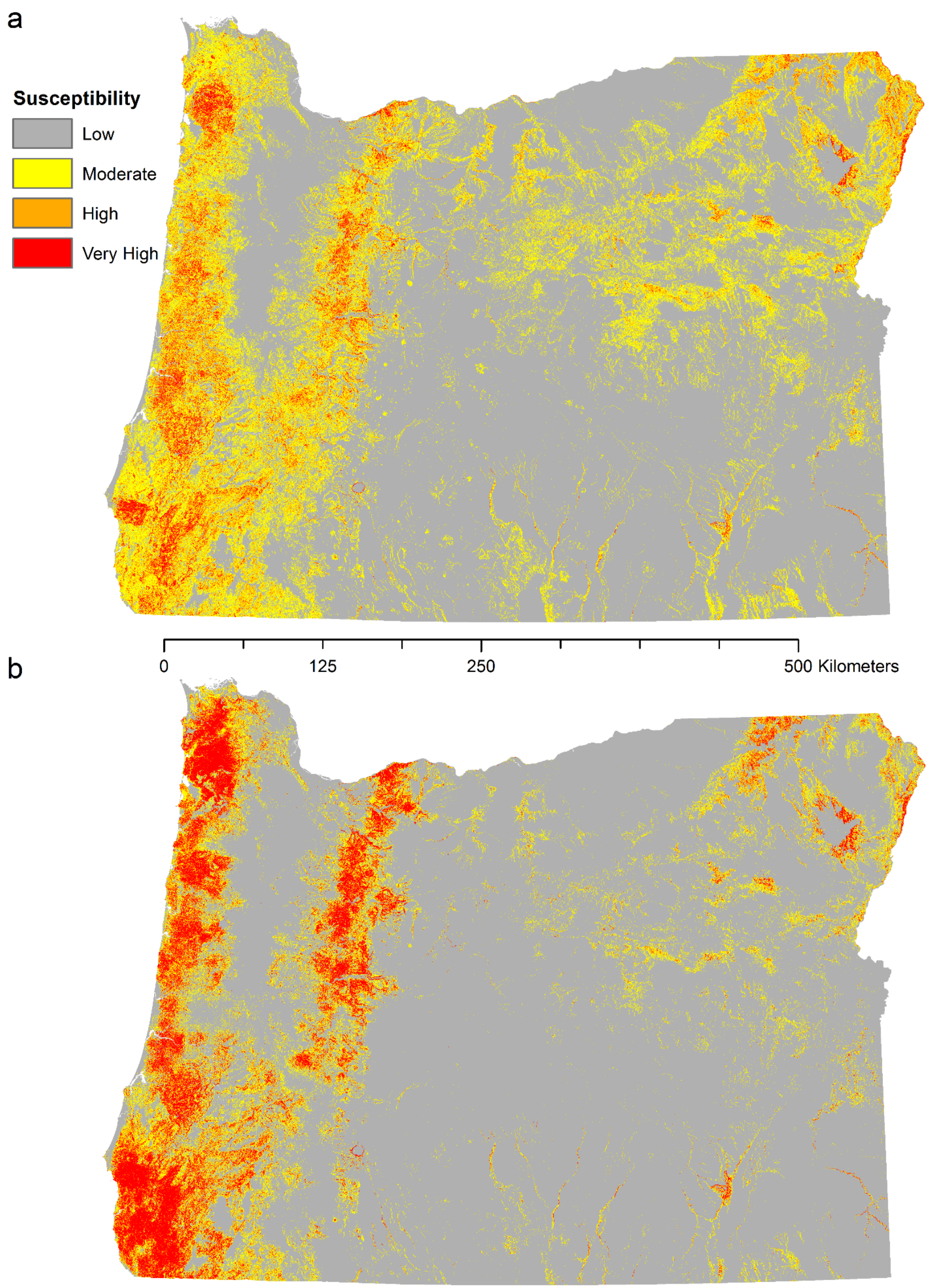

Figure 9. a) Landslide susceptibility maps calculated from the synthetic inventory with both road and population biases, where the training data consisted of points within 1 kilometer of a major highway. b) The same treatment was applied to an inventory of undated landslide initiation points.

landslide susceptibility with heavily biased inventories. Ultimately, the creation of new, comprehensive inventories may simplify the task of landslide susceptibility mapping over large areas.

ACKNOWLEDGMENTS

This research was funded by the NASA NNH14ZDA001N-INCA: Climate Indicators and Data Products for Future National Climate Assessments. This work would not have been possible without the data provided by DOGAMI, ODOT, and many others. We also thank Jordan Psaltakis for assembling the PNLI.

REFERENCES

Atkinson, P.M., and Massari, R., 1998, Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy: Computers & Geosciences, v. 24, no. 4, p. 373–385, doi: 10.1016/S0098-3004(97)00117-9.

Bivand, R., Keitt, T., and Rowlingson, B., 2015, rgdal: Bindings for the Geospatial Data Abstraction Library:

Bright, E.A., Rose, A.N., and Urban, M.L., 2015, LandScan 2014:

Bucknam, R.C., Coe, J.A., Chavarría, M.M., Godt, J.W., Tarr, A.C., Bradley, L.-A., Rafferty, S., Hancock, D., Dart, R.L., and Johnson, M.L., 2001, Landslides Triggered by Hurricane Mitch in Guatemala — Inventory and Discussion.:

Burns, W.J., 2014, Statewide landslide information database for Oregon, release 3.2:

Burns, W.J., Mickelson, K.A., and Madin, I.P., 2016, Landslide susceptibility overview map of Oregon.:

Center for International Earth Science Information Network, and Information Technology Outreach Services, 2013, Global Roads Open Access Data Set, Version 1:

Chau, K.T., and Chan, J.E., 2005, Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island: Landslides, v. 2, no. 4, p. 280–290, doi: 10.1007/s10346-005-0024-x.

Das, I., Sahoo, S., van Westen, C., Stein, A., and Hack, R., 2010, Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India): Geomorphology, v. 114, no. 4, p. 627–637, doi: 10.1016/j.geomorph.2009.09.023.

ESRI, 2013, ArcGIS Desktop:

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., and Reichenbach, P., 2008, Comparing landslide inventory maps: Geomorphology, v. 94, no. 3–4, p. 268–289.

Gorsevski, P. V, Gessler, P.E., Foltz, R.B., and Elliot, W.J., 2006, Spatial Prediction of Landslide Hazard Using Logistic Regression and ROC Analysis: Transactions in GIS, v. 10, no. 3, p. 395–415, doi: 10.1111/j.1467-9671.2006.01004.x.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K.T., 2012, Landslide inventory maps: New tools for an old problem: Earth-Science Reviews, v. 112, no. 1–2, p. 42–66, doi: 10.1016/j.earscirev.2012.02.001.

Guzzetti, F., and Tonelli, G., 2004, Information system on hydrological and geomorphological catastrophes in Italy (SICI): a tool for managing landslide and flood hazards: Natural Hazards and Earth System Science, v. 4, no. 2, p. 213–232, doi: 10.5194/nhess-4-213-2004.

Hijmans, R.J., 2015, raster: Geographic Data Analysis and Modeling. R package version 2.4-15.:

Kirschbaum, D.B., Adler, R.F., Hong, Y., Hill, S., and Lerner-Lam, A., 2010, A global landslide catalog for hazard applications: method, results, and limitations: Natural Hazards, v. 52, p. 561–575, doi: 10.1007/s11069-009-9401-4.

Kirschbaum, D., Psaltakis, J., and Stanley, T., 2016, Spatiotemporal properties of landslides in the pacific northwest, *in* GSA Annual Meeting,.

Kirschbaum, D.B., Stanley, T., and Zhou, Y., 2015, Spatial and temporal analysis of a global landslide catalog: Geomorphology, v. 249, no. Geohazard Databases: Concepts, Development, Applications, p. 4–15, doi: 10.1016/j.geomorph.2015.03.016.

Montgomery, D.R., 1994, Road surface drainage, channel initiation, and slope instability: Water Resources Research, v. 30, no. 6, p. 1925–1932, doi: 10.1029/94WR00538.

Ninu Krishnan, M. V., Pratheesh, P., Rejith, P.G., Hamza, V., and Vijith, H., 2014, Determining the Suitability of Two Different Statistical Techniques in Shallow Landslide (Debris Flow) Initiation Susceptibility Assessment in the Western Ghats: Environmental Research, Engineering and Management, v. 4, no. 70, p. 27–39, doi: 10.5755/j01.erem.70.4.8510.

Oregon State University, 2016, 1971-2000 Annual Average Precipitation by State:

Pebesma, E.J., and Bivand, R.S., 2005, Classes and methods for spatial data in R: R News, v. 5, no. 2, doi: 10.1159/000323281.

R Core Team, 2015, R: A language and environment for statistical computing:

Sing, T., 2005, Visualizing Classifier Performance in R: Bioinformatics, v. 21, no. 20, p. 7881.

Smith, R.L., and Roe, W.P., 2015, Oregon Geologic Data Compilation, release 6:

Steger, S., Bell, R., Petschko, H., and Glade, T., 2015, Evaluating the Effect of Modelling Methods and Landslide Inventories Used for Statistical Susceptibility Modelling, *in* Engineering Geology for Society and Territory - Volume 2, Springer International Publishing, Cham, p. 201–204.

Steger, S., Brenning, A., Bell, R., and Glade, T., 2016a, The impact of systematically incomplete and positionally inaccurate landslide inventories on statistical landslide susceptibility models, *in* Geophysical Research Abstracts,.

Steger, S., Brenning, A., Bell, R., Petschko, H., and Glade, T., 2016b, Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps: Geomorphology, v. 262, p. 8–23, doi: 10.1016/j.geomorph.2016.03.015.

Taylor, F.E., Malamud, B.D., Freeborough, K., and Demeritt, D., 2015, Enriching Great Britain’s National Landslide

Database by searching newspaper archives: Geomorphology, v. 249, p. 52–68, doi: 10.1016/j.geomorph.2015.05.019.
U.S. Geological Survey, 2016, The National Map: 3DEP products and services: The National Map, 3D Elevation Program Web page.