Effects of inventory bias on landslide susceptibility calculations

Thomas Stanley, USRA/GESTAR/NASA GSFC
Dr. Dalia B. Kirschbaum, NASA GSFC

“Multi-temporal landslide information is essential to new approaches for the generation of quantitative landslide probability maps.”

-Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview
Cees J. van Westen, Enrique Castellanos, Sekhar L. Kuriakose
Pacific Northwest Inventories

Oregon
http://www.oregongeology.org/slido

Washington
https://fortress.wa.gov/dnr/protectionis/geology/?Theme=natural_hazards
Landslide susceptibility workflow

Landslide susceptibility. A quantitative or qualitative assessment of the classification, volume (or area), and spatial distribution of landslides which exist or potentially may occur in an area. Susceptibility may also include a description of the velocity and intensity of the existing or potential landsliding. Although it is expected that landsliding will occur more frequently in the most susceptible areas, in the susceptibility analysis, time frame is explicitly not taken into account.

Landslide susceptibility analysis of lifeline routes in the Oregon Coast Range

- Predictors: PGA, PGV, Slope, Precipitation (PRISM 30-year MAP)
- Empirical susceptibility model fitted with landslide polygons and points
- Random assignment of non-landslides
- Classification of probability into 5 bins

Landslide susceptibility overview map of Oregon

- Predictors: Slope, Geologic Unit
- Empirical susceptibility model fitted with landslide polygons
- Validated with historic landslide points
- Classification into 4 bins
- Very high susceptibility category consists solely of SLIDO polygons

Inventory bias

Not only can inventory bias alter the results of landslide susceptibility calculations, it may even **improve** the validation statistics, giving false confidence in the map.

# How it works

<table>
<thead>
<tr>
<th>Location</th>
<th>Mean slope</th>
<th>Forest cover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cathy’s Condo</td>
<td>30</td>
<td>0%</td>
</tr>
<tr>
<td>Millie’s Meadow</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Bob’s Bluff</td>
<td>30</td>
<td>50%</td>
</tr>
</tbody>
</table>

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|---------|
| (Intercept) | -23.57     | 79460.0 | 0       | 1       |
| slope    | 1.57       | 3746.0  | 0       | 1       |
| forest   | 0.00       | 224800.0 | 0    | 1       |
Revised susceptibility workflow

In order to train an accurate landslide susceptibility model, we must correct for the numerous false negatives implied by our biased inventory.
1 Compile Pacific Northwest Landslide Inventory (PNLI)

- 7,454 landslides with a known date, with 3,373 in Oregon and 4,081 in Washington.
- Year of occurrence was known for an additional 7,967 landslides.
- Year of occurrence was not known for 58,780 landslides. These were not used to generate the initial landslide susceptibility map.
PNLI reporting biases

- A synthetic inventory was generated for this analysis from the Landslide susceptibility overview map of Oregon by randomly creating landslide initiation points with the same frequency of occurrence documented for each category (low-very high).

The synthetic inventory was generated for this analysis from the Landslide susceptibility overview map of Oregon.
Simulate PNLI reporting biases

a) Start with two relatively unbiased inventories for Oregon

b) Truncate each inventory to match biases (roads, population, roads + population)

c) Fit logistic regression models

d) Compare biased susceptibility maps to results for initial inventories

Synthetic inventory derived from:
2 Mitigate bias

• While the problem could be solved through the use of a predefined method, the large inventory available for the Pacific Northwest would not inform such a model.

• In order to reduce the influence of false negatives, only areas located within 1 kilometer of a major highway were used to fit the model. Landslides in more remote areas were used as a validation dataset.
3 Fit logistic regression model

Distance to fault
Mean Annual Precipitation
Slope
Geologic Unit


4 Validate results

Biased
- Overemphasis on distance to fault
- Underestimate in specific geologic units

Bias-mitigated
- Coast Range identified as highly susceptible
Discussion

• While we believe that the PNLI is concentrated near highways and cities due to reporting bias, some of this effect is probably due to anthropogenic disturbance.

• Therefore, susceptibility may have been overestimated in the areas away from highways (validation zone). The estimates near highways should have represented anthropogenic effects correctly.
Conclusions

• Reporting bias can have a strong effect on the fitting of empirical landslide models.

• Although many strategies for bias mitigation could be employed, the simplest approach delivers generally plausible results that are most reliable in the most critical locations: along major highways and rail lines.
Next steps

• Map susceptibility across the Pacific Northwest, with separate models for each landslide type

• Identify trends in landslide-triggering precipitation

• Apply lessons learned across the USA
Acknowledgements
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.

Recommended reading
• 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.