Geomorphology 184 (2013) 156–159

Contents lists available at SciVerse ScienceDirect

Geomorphology

journal homepage: www.elsevier.com/locate/geomorph

Reply: Comparison of slope instability screening tools following a large storm event and application to forest management and policy

Kara A. Whittaker [a,*], Dan McShane [b,1]

[a] Washington Forest Law Center, 615 Second Ave., Suite 360, Seattle, WA 98104, USA
[b] Stratum Group, PO Box 2546, Bellingham, WA 98227, USA

ARTICLE INFO

Article history:
Received 9 November 2012
Accepted 13 November 2012
Available online 22 November 2012

Keywords:
Forestry
Landslide
Slope instability
Screening tools

ABSTRACT

A large storm event in southwest Washington State triggered over 2500 landslides and provided an opportunity to assess two slope stability screening tools. The statistical analysis conducted demonstrated that both screening tools are effective at predicting where landslides were likely to take place (Whittaker and McShane, 2012). Here we reply to two discussions of this article related to the development of the slope stability screening tools and the accuracy and scale of the spatial data used. Neither of the discussions address our statistical analysis or results. We provide greater detail on our sampling criteria and also elaborate on the policy and management implications of our findings and how they complement those of a separate investigation of landslides resulting from the same storm. The conclusions made in Whittaker and McShane (2012) stand as originally published unless future analysis indicates otherwise.

1. Introduction

Two discussions of our recent article “Comparison of slope instability screening tools following a large storm event and application to forest management and policy” (Whittaker and McShane, 2012) question the accuracy and scale of the spatial data we used (Lingley et al., 2013-this volume) and our discussion of the screening tool development (Shaw, 2013-this volume). Both discussions also allege shortcomings in the policy and management implications we identified, but neither of them address our statistical analysis or results demonstrating that both screening tools are effective at predicting where landslides were likely to take place. Both Shaw (2013-this volume) and Lingley et al. (2013-this volume) misrepresent or greatly misunderstood our discussion on policy considerations. Here, we address each of these issues and recommend additional analysis that can resolve any outstanding concerns or misunderstandings. Until such analysis is conducted, we argue that the conclusions made in Whittaker and McShane (2012) stand as originally published.

2. Research section discussion

2.1. Accuracy and scale of spatial data

The comments provided by Lingley et al. (2013-this volume) suggest that we did not recognize the scale differences and resolutions of the landslide initiation points and the two screening tools analyzed. They also rather specifically implied that we did not understand the reconnaissance level of the landslide initiation data set. This is simply not true, and we made a fair effort to present that indeed there are resolution and scale issues with the data and the screening tools. For example, we noted the source of the initiation point data and referenced that source: “Landslide initiation point data was gathered by the Washington Department of Natural Resources (WDNR) during reconnaissance flights across southwest Washington immediately following the December 2007 storm (WDNR, 2009).” In the discussion we stated: “Another probable source of Type I errors are GIS mapping artifacts, hence, the error rates reported here should be treated as estimates rather than absolute values.” In addition, we provided a lengthy discussion describing how the assumptions and resolutions of the screening tools can introduce errors in the identification of landslide sites. For instance, we acknowledged “the overall type I error rates we observed probably would have been lower if we had utilized different methods for mapping landslides”, and “Small mapping errors in landslide initiation points are more likely to lead to classification errors on higher resolution maps if classification is based on the hazard level of single pixel.” All of the above caveats were appropriately disclosed in our article. Despite the poor resolution of the landslide data points and screening tools, our statistical analysis suggests that the screening tools developed by the WDNR work well in identifying potentially unstable slopes.

Lingley et al. (2013-this volume) also contend that our analysis was conducted at a finer scale than the intended scales of the screening tools. This was not our intent, but rather was a byproduct of the data available to us and the nature of the analysis we wished to conduct. In contrast to Lingley et al.'s (2013-this volume) assertions, WDNR staff did not discuss this scale limitation with us nor did they communicate the intent and quality of the data in writing, verbally, or in their report

* Corresponding author. Tel.: +1 206 223 4088x5; fax: +1 206 223 4280.
E-mail addresses: kwhittaker@wflc.org (K.A. Whittaker), mcshanedan@gmail.com (D. McShane).
[1] Tel.: +1 360 714 9409.

0169-555X/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.geomorph.2012.11.008

(Sarikhan et al., 2008). On the contrary, Sarikhan et al. (2008) reported that "landslide location was accurately determined using vegetation and other visible clues on the orthophotos". Knowing WDNR was refining the landslide initiation point file over time, we repeatedly confirmed with them that we had the most recent and accurate information available before conducting our analysis. On one occasion (via email), WDNR staff explained that the landslide initiation point file had been "significantly cleaned up and made a lot more accurate (at a 1:24,000 scale)". This was the *only* mention of scale to us by WDNR. Thus, we felt confident that we were using the best available source of landslide spatial data available to us at the time of our analysis. We did not spatially rectify the landslide initiation points using high-resolution orthoimagery because we conducted our analysis before the post-storm orthophotos became available. Nor did we verify the landslide locations in the field, because the vast majority of them occurred on private timberlands inaccessible to the public. Given our finding that the two screening tools provided statistically valid predictions of potentially unstable slopes as they were designed to do, we did not see the utility in spatially rectifying the landslide data or seeking permission to access private timberlands and repeating our analysis.

In their critique of the scale of our analysis, Lingley et al. (2013-this volume) made an inapplicable and incorrect reference to the National Map Accuracy Standards (United States Geological Survey, 1999). They pointed out that "any data points generated from the landslide observations are also reconnaissance-level and should not be applied to an analysis other than those using approximately 100,000-scale or coarser (United States Geological Survey, 1999)." First, we did not generate any data points in our analysis. Second, the USGS (1999) reference is not applicable to landslide data points. The USGS map accuracy standards tolerate up to 10% error for "well-defined points" that have been rigorously field surveyed such as property boundaries, road intersections, and building corners or center points (USGS, 1999). In contrast, landscape initiation points do not meet the criteria of appropriate locations for testing mapping accuracy, as they are "features not identifiable upon the ground within close limits…even though their positions may be scaled closely upon the map" (USGS, 1999). This reference does not support Lingley et al.'s position that our analysis was conducted at an inappropriate scale.

Lingley et al. (2013-this volume) incorrectly reported that we stated that HAZONE was higher resolution because it was vector data. Rather, our statement that "Small mapping errors in a landslide initiation points are more likely to lead to [hazard] classification errors on higher resolution maps if classification is based on the hazard level of a single pixel" referred to the greater likelihood of mapping errors with the higher resolution SLPSTAB raster data than with the lower resolution HAZONE vector data. Despite this difference, neither screening tool showed a statistically greater ability to predict landslide locations than the other.

2.2. SLPSTAB development and assumptions

Uncovering the methods used to develop the SLPSTAB screening tool was a challenge. Our only source of information was the WDNR through their website and multiple public disclosure requests for any information on this topic. We were not informed of the details of SLPSTAB development disclosed by Shaw (2013-this volume), nor were we provided Ms. Vaugeois's unpublished document cited by Shaw (2013-this volume). We were unsuccessful in our attempts to locate and contact Ms. Shaw to inquire about this process. We appreciate now knowing more about the model development process through this discussion.

Shaw (2013-this volume) pointed out that "SLPSTAB was not intended to be used in comparing model predictions of landslide potential with deep-seated or road-related landslide initiation points". We do not disagree, but the intent of the screening tool was to identify landforms with shallow landslide potential. In our analysis of the SLPSTAB and HAZONE screening tools, we did not discriminate among landslide types because this information was not included in the landslide initiation point attribute data provided by WDNR. Had we had access to these data, we doubt our results would have differed for several reasons. First, data reported in Stewart et al. (unpublished results; Table 5-5) indicate that 98% ($N=1429$) of landslides measured fit the definition of shallow rapid landslides (WDNR, 2004). Of these, 23% ($N=331$) were road-related landslides and 1% ($N=15$) were deep-seated landslides. While road-related landslides may result from road-related factors (i.e., culvert blockage, sidecast road fill), they also result from factors unrelated to road structures (i.e., slope gradient, convergence, hydrology) that can be identified with screening tools. Because we were unable to distinguish between these types of landslide triggers, we retained all road-related landslides in our sample. Second, tests of shallow landslide screening tool models by others included both road-related and deep-seated landslides (Montgomery et al., 1998; Shaw and Vaugeois, 1999). Further, SHALSTAB predicted landslides associated with either roads or harvest units equally well, and the location of road-related landslides was topographically driven (Montgomery et al., 1998). Lastly, shallow landslides and deep-seated landslides are often not mutually exclusive events (Shaw and Vaugeois, 1999; WDNR, 2004), though purely deep-seated landslides could have been appropriately removed from our sample had we been privy to such information.

2.3. Other concerns

Lingley et al. (2013-this volume) felt our discussion of other factors influencing slope stability was inadequate, namely, bedrock hydrogeology, known unstable geologic units, and the rain-on-snow event associated with the December 2007 storm. The purpose of our analysis was to assess the effectiveness of slope stability screening tools; a detailed analysis of the underlying geologic conditions and landslides was beyond the scope of our analysis. However, we did describe the geology and topography of all three watersheds in our Methods (Section 2.2. Geological context). In addition, in our discussion we discussed the underlying geologic units, the relationship between bedrock, soil porosity, and subsurface water concentration, and the potential for geologic conditions to cause type I errors. We mentioned the use of known unstable geologic units in the development of both screening tools (Methods, Section 2.3. Slope instability screening tools) and in our discussion of the relationship between type I errors and screening tool resolution and parameters. We did not specifically mention rain-on-snow in association with the December 2007 storm, but we recognized the heavy precipitation, hurricane-force winds, and significant flooding associated with the storm. The storm event has been more thoroughly addressed by others cited in our Introduction (Mote et al., 2007; Reiter, 2008; Sarikhan et al., 2008; Turner et al., 2010).

2.4. Verification of findings

It is unclear why the WDNR Division of Geology and Earth Resources is not more supportive of the fact that the screening tools that *they* developed have been demonstrated by independent authors as being very effective. We encourage the WDNR to repeat our analysis using the most accurate, spatially rectified, and field-truthed data available (from Stewart et al., unpublished results) at the scale and level of accuracy they deem sufficient, and we would gladly lend our support in such efforts once the WDNR makes that data publicly available. We expect such an analysis would further validate the findings in our original article, and given the potential error sources we discussed may very well show WDNR's models to be even more accurate than our analysis showed.

3. Policy section discussion

Lingley et al. (2013-this volume) claimed it was inappropriate for us to cite an unpublished paper (Stewart et al., unpublished results).