Failure may also occur in a progressive fashion, both at the scale of individual roots [*Schwarz et al*., 2010] and at the hillslope scale [*Petley et al*., 2005]. Alternative stability models that account for internal deformation explicitly [e.g., *Griffiths and Lane*, 1999; *Itasca, F. L. A. C.*, 2000; *Borja and White*, 2010] or implicitly [e.g., *Lehmann and Or*, 2012; *Ruette et al*., 2013] can capture some of these effects. However, since the mechanics of progressive failure remain poorly understood, we opt to use a limit-equilibrium framework in which the stability problem is well defined, resulting in a suite of robust approximations [e.g., *Fellenius*, 1927; *Bishop*, 1955; *Spencer*, 1967; *Hovland*, 1977]. In its current form, the algorithm could be used to represent progressive failure with an iterative approach in which the unstable shapes are removed or redistributed, in a manner similar to *Lehmann and Or* [2012].

While the search algorithm can use any three-dimensional slope stability model within the limit-equilibrium framework, specific models make different assumptions. The model MD-STAB [*Milledge et al*., 2015] used here assumes that the central block is rigid with failure by shear on a surface parallel to the ground surface at a prescribed depth. Additionally, the model assumes (1) that failure occurs in drained conditions under steady slope-parallel groundwater flow without suction in the unsaturated zone; (2) that the soil is normally consolidated with isotropic frictional properties and its density is independent of moisture content; (3) that the cross-slope boundaries are vertical, and that earth pressure on these boundaries is in an at-rest condition; (4) that earth pressure on the upslope and downslope margins is characterized by active and passive conditions, respectively; and (5) that intercolumn shear forces do not exceed the shear strength within the unstable block.

5.2. Multiple Overlapping Predictions

Applying this search algorithm even to very small landscapes introduces a new problem: the search identifies multiple overlapping clusters of unstable cells (e.g., Figure 7), but any cell can only fail once. This problem is not specific to our search algorithm but to any search that calculates the stability of many candidate clusters (e.g., an exhaustive search). In our search algorithm overlapping predictions may arise when each edge of the contour tree is traversed (Figure 4c), when contour tree edges merge (i.e., within the same eigenvector), and when combining results from many eigenvectors or from overlapping spatial windows. Here we propose to retain the least stable shape (FS_{min}), which performed best in our tests and is most consistent with the optimization; and also test an alternate end-member method that retains the unstable shape with FS closest to unity (the FS_{max} method). Selecting which model output is most likely to occur in reality is not a well-studied problem, particularly given the levels of approximation that go into all models. In our tests we found that the FS_{min} method performed consistently well, but that while most of the 475 unstable clusters found in the CB-1 test overlap with the observed landslide (Figure 7b), the choice of which overlapping prediction to select can have an impact on predicted landslide location and size. For example, the FS_{max} method predicted a larger and more elongated landslide, located further downslope (gray line in Figure 7b). While retaining all the possible unstable configurations in a catchment-scale application is impractical, retaining the size distribution of all landslides found at each location would not require significant additional computational costs and would provide probabilistic constraints on the possible outcomes. However, sampling from this distribution is a nontrivial problem. Figure 8 shows that the observed landslide did not correspond to the modes of the size, location, and shape distributions, suggesting the need for a weighted sampling based on prior knowledge. This weighting could be based on the FS, as suggested by *Stark and Guzzetti* [2009], whereby the least stable overlapping shape would be assigned a higher probability of being selected than more stable shapes. Alternatively, as more accurate and diverse landslide data sets become available, data-driven methods could be used to make an informed choice on which of the overlapping predictions is more likely under a specific set of conditions (e.g., using observed landslide size and shape data for a particular region to constrain which landslide to pick from the modeled distribution).

The presence of many overlapping predictions does indicate that many outcomes are possible for any scenario and suggests that subtle variations in local conditions not captured by the model could determine which outcome is ultimately realized. In most natural landscapes the details of the local conditions are not known, and may be unknowable at a scale relevant to shallow landslides. This implies that data quality may place an unsurpassable constraint on our ability to forecast the exact size of a landslide at a specific location and suggests that instead calibration and testing of the model should be based on the frequency

©2015. American Geophysical Union. All Rights Reserved.