distribution of sizes and locations of the predicted and observed landslides rather than on prediction of exact location of individual slides. Nevertheless, Figure 8 suggests that estimating field parameters (e.g., soil depth, root strength, and pore pressure) to a few percent of their actual value can lead to satisfactory approximate results, highlighting the importance of characterizing these parameters in the field as accurately as possible.

5.3. Performance

In the synthetic tests we planted landslides on a landscape, applied a search algorithm to find them, and recovered their approximate size and location, irrespective of the shape of the low-strength patch. However, in many cases all the search methods found different landslides within the same low-strength patch. In absence of any other information, each of these predictions is equally likely. While the set of possible landslides is tightly constrained to a ± 1 grid cell from the boundary of the low-strength patches, making these tests nontrivial, multiple configurations of grid cells can fail implying that there is not a single true answer. This prevents a strict quantitative comparison between the search methods and the planted landslides. Nevertheless, the tests demonstrate that our search algorithm always found the landslides and that the predictions had similar size and shape to the low-strength patches.

Application of the search algorithm to a real landslide at CB-1 is a somewhat stricter test because (1) there is a ground truth (i.e., we know which of the possible landslides actually failed) and (2) all the landslide-relevant parameters vary continuously across the landscape rather than changing abruptly as in the synthetic landscape. The search algorithm applied to a real landscape at CB-1 predicted failure at the same location and of similar size to the landslide that occurred during the November 1996 storm (Figure 7c), without any parameter tuning (all parameters are defined by field measurements). The least stable predicted landslide precisely straddles the area of high pore pressure and thicker soils, while the observed head scarp intersects the pore pressure peak (Figure 7c). As a result, it has an area that is 9 m^2 larger than the observed failure, extending approximately 2 m further upslope. This is likely due to the small-scale parameter variations that exist in natural landscapes which cannot be captured even in a site as well studied as CB-1. In particular, we note that the spatial arrangement of the piezometer nests used to measure the pore pressure values (maps in *Montgomery et al*. [2009]) will inevitably result in the highest pressure values being extended upslope when a continuous field is interpolated, because no other piezometers were in the immediate vicinity. Moreover, the use of a raster framework will inevitably result in slight differences between predictions and observations that could become more significant with coarser resolutions.

Any failure at this site is likely to be centered around the maximal values of pore pressure and soil depth, so that location alone is not a definitive test. However, considering the wide distributions of the possible unstable configurations, their size, shape, and location (Figure 8), the similarity of the least stable predicted landslide to the observed is very strong. In contrast, the predictions resulting from the alternative exhaustive search methods deviate considerably from the observations (Table 2 and Figure 8).

The rectangular search finds a landslide with a FS that is lower than both that for our predicted landslide and that for the observed landslide. This is likely due to the fact that, in situations where the grid is oriented slope parallel, rectangles will minimize perimeter length for a given area. Furthermore, the calculated FS of the observed landslide is also affected by grid discretization. The reason that our search identifies a rounder landslide (despite the higher perimeter to area ratio for this shape) is very likely related to the balance between global and local components in the optimization, resulting in a smooth eigenvector surface (Figure 7a) guiding the search algorithm. This points to the fact that while the mathematical properties of Laplacian matrices are extremely useful (see overview in *Mohar* [1991]), the link between those matrices and real-world clustering problems such as ours is somewhat heuristic.

From a methodological point of view this application (searching for potential landslides) is particularly well suited to spectral clustering. The lack of a rigorous definition of what is a good measure of similarity in the input data results in a number of spectral clustering algorithm variants, and no agreement on which one is best (see review in *Von Luxburg* [2007]). In contrast, we were able to take a physically based model and use a well-defined Factor of Safety formulation to define the objective function for the spectral clustering. While we ultimately use similar linear-algebraic techniques as in other spectral clustering methods [e.g., *Shi and Malik*, 2000], we suggest that their justification is in this case much stronger.