landslides using a spectral clustering approach similar to that of *Shi and Malik* [2000], but with the advantage that the objective function is physically based. Using a slope stability model such as that in section 2, we determine the contribution of each soil column to the driving force of a potential landslide that includes that cell, the resistance on the base of each soil column, and the resistance between each soil column and its neighbors. We encode the driving and basal resisting forces contributed by each column in the vertices of the graph, and the potential lateral resisting forces that can develop between columns in the edges of the graph. Each vertex thus corresponds to a discretized soil column (i.e., a grid cell), and each edge corresponds to a boundary between columns (Figure 1). We define the objective function as the ratio of all the resistances acting on a cluster of cells and the total driving force contributed by those cells. We develop a spectral method to find clusters (i.e., landslides) that minimize this function, guiding the search for the unstable areas of the landscape.

3.2. Spectral Clustering Derivation

The landscape and all its spatial attributes are discretized into a regular grid (i.e., with square grid cells). The attributes of the landscape relevant to shallow landslides are represented as an undirected weighted graph $G = (V,E)$, where the vertices $v \in V$ of the graph represent the vertical soil columns corresponding to the original discretized grid cells. An edge $e_{ij} \in E$ is formed between every pair of neighboring vertices v_i and v_j and represents the forces (which are encoded in the edge weights) acting between two adjacent columns (Figure 3a). The discretization of the landscape into soil columns with four sides results in each vertex having four neighbors. The edges of G are thus associated with the four lateral faces of the soil columns, on which the forces acting between adjacent grid cells are defined. This graph G can be represented as a weighted adjacency matrix **W**, with entries w_{ij} set to 0 when v_i and v_j are not connected in G and to the magnitude of the resistive forces on edge e_{ij} otherwise. G is four regular as each vertex has four neighbors (upslope, downslope, left, and right). For a grid of size n by m, **W** will be of size nm by nm but will contain mostly zero values, and thus will be extremely sparse [*Cormen et al.*, 2001]. This is important computationally, as the use of sparse matrices introduces little memory overhead in comparison to the original elevation grid.

The vertices of this graph are annotated with the forces contributed by each grid cell: the driving forces f and the resistive forces b due to friction and cohesion acting on the base of the column (Figure 3a). Similarly, the edges of the graph are annotated with the forces acting between a grid cell and its neighbors: these resistive forces, which arise from root strength and earth pressure, are mostly positive, but may be negative to represent the active "push" from upslope neighbors. Spectral clustering methods associate a scalar weight with each node [*Von Luxburg*, 2007], and thus, we associate each grid cell with the magnitude of its driving force, rather than magnitude and direction. The total driving force is thus the arithmetic sum rather than the vector sum, resulting in an over estimation of the force (Figure 2a). As a result, some stable areas may be unnecessarily examined by the search algorithm. However, when candidate landslides are examined by the algorithm, the FS is correctly computed using the vector sum (Figure 2a). The driving and resistive forces are encoded in two nm by nm matrices **F** and **R**. The force matrix **F** contains the driving forces associated with the vertices of G along the diagonal (Figure 3b):

$$\mathbf{F_{ii}} = F_{d_i}, \tag{7}$$

where F_{d_i}, is the force contributed by vertex v_i. The resistance matrix **R** contains the resistive forces associated with both the vertices (basal resistance) and the edges (lateral resistance) of G (Figure 3c). The diagonal of **R** is

$$\mathbf{R_{ii}} = R_{b_i} + \sum_{j \neq i}^{m^* n} w_{ij}, \tag{8}$$

where R_{b_i} is the basal resistance contributed by the soil column represented by vertex v_i and the w_{ij} term is the sum of all the resistances between v_i and its neighbors. This definition differs from the general formulation of the Laplacian matrix typically used in spectral clustering methods [e.g., *Shi and Malik* [2000]], as it also includes the R_{b_i}, terms (which are associated with the vertices of G). A central assumption of spectral clustering approaches is that the relationship between neighbors is symmetric (i.e., the graph is undirected) [*Von Luxburg*, 2007]. The pixel similarity used by *Shi and Malik* [2000] is a symmetrical measure. In their graph representation, v_i similar to v_j implies v_j similar to v_i and that $w_{ij} = w_{ji}$, resulting in an undirected graph. This is not the case in this application, as the resistive forces between grid cells are not always symmetrical. In particular, the active and passive earth pressures across a vertical boundary separating an