Earth Surf. Dynam. Discuss.,
doi:10.5194/esurf-2016-6-RC2, 2016

Earth Surface Dynamics
Discussions
Open Access

Interactive comment on "An introduction to learning algorithms and potential applications in geomorphometry and earth surface dynamics" *by* Andrew Valentine and Lara Kalnins

NS Anders (Referee)

n.s.anders@uva.nl

Received and published: 20 March 2016

This paper sums up the most widely used learning algorithms to extract information from geospatial data sources. The geospatial analysis and geomorphometry related research is becoming more and more data-driven and I encourage papers advocating making use of modern computational techniques in geomorphology. This paper is particularly interesting to those unfamiliar to learning algorithms. Below I have addressed some comments that could improve the paper.

General

* The paper explains well the most widely used methods in current literature with effec-

Printer-friendly version

Discussion paper

tive illustrations. Each topic is supported with one or few references to applications in the geosciences, mostly landslide susceptibility. However, to justify the part in the title concerning '...and potential applications in geomorphometry and earth science dynamics' I think the paper should cover a more in depth section on the potential applications of learning applications in the geosciences. Or perhaps remove 'potential' from the title and at least provide more references in the 3.1-3.7 subsections to current research that are applying these algorithms.

* Also in the subsections there are occasional statements such as 'it is claimed that' (P4L8) or 'literature describes a rule of thumb' (P6L9), which are not followed by references. I think it is required to provide references to support such statements in general.

* I suggest presenting an overview of strengths and weaknesses in different scenarios (data sets), or what learning algorithm could be appropriate for particular data sets or different scenarios, possibly as a table. For example when should scientists choose random forests over neural networks. There is probably not a single best method for each scenario, but I guess some algorithms are ruled out because of certain data types or data properties (maybe due to statistical limitations). This would also broaden the audience reading this paper.

* Although the English is much better than my own, there are a few singular/plural errors throughout the paper. I suggest to have the paper thoroughly checked.

Specific

P3L11: I would not say LiDAR 'image'. In remote sensing an image is often associated with spectral measurements captured by a sensor creating this image. The authors probably mean a derived grid of interpolated elevation values from a LiDAR point dataset, hence LiDAR DEM or DTM is more appropriate than 'image'. Also the abbreviation style LiDAR (with capital A) is more commonly used than LiDaR.

P4L33: It is unclear to me what exactly this 'data vector' means with respect to a grid and why a 100m x 100m grid with with 10 m cells (thus 10x10=100 cells) would have (11x11=) 121 dimensions.

P5L6-10: I wonder if dimension reduction by means of visualization is appropriate to mention in this section, as dimension reduction in the context of learning algorithms is primarily used to reduce complexity of calculations and/or computing. To me the way how results are visualized and whether dimensions or detail are lost is a different discussion.

P5L11-15: While vegetation of course has indirectly also geomorphological implications I'd like to, rather shamelessly, self-advertise our recent paper that is a more pure example of cluster analysis in (palaeo-glacial) geomorphology, see Smith et al (2016) in ESPL: http://onlinelibrary.wiley.com/doi/10.1002/esp.3828/abstract

P6L10: is 'performance' here meant in terms of computation time, or in terms of accuracy, or both.

P9L18: now LiDAR is abbreviated as LIDAR. I suggest selecting one style of abbreviation, preferably Lidar, LiDAR or LIDAR.

Section 3.3. I think PCA should be explained in more detail. It is for example not mentioned what the axes actually mean and that what information is derived from the point distribution along the axes. For example that the axes do likely resemble an average of multiple dimensions is not mentioned. For example in the example of a PCA of plant species distribution with respect abiotic factors such as soil type, ph, and water availability a single PCA axis could resemble both ph and water availability and could make interpretation of PCA more complex. When describing PCA I think such information is essential to mention.

Interactive comment on Earth Surf. Dynam. Discuss., doi:10.5194/esurf-2016-6, 2016.