ETWG Soils - Topic - Disaggregation

Participants

- Mike Wieczorek
- Travis Nauman
- Norm Bliss
- Sharon Waltman

Disaggregation Overview

The term 'disaggregation' has been commonly used in the literature to describe different approaches to improve the thematic and spatial resolution of conventional soil surveys (CSM). The root of the need for disaggregation is that many soil surveys are structured around map units concepts that include several soils. These map units consist of polygon delineations that map groups of soils that can have a variety of properties (Fig. 1). In Fig. 1 there are three different map units that include multiple soils (PLF, GdE, and DrF) that the authors disaggregate using different GIS environmental rasters (DEM and derivatives, imagery, geology, land cover, and others) and machine learning (Nauman & Thompson, 2014).

![Figure 1. Example of a soil survey 'disaggregation' in West Virginia, USA. The figure is out of a 2014 article in Geoderma (Nauman and Thompson, 2014).](image)

Evolution of Methods
As computing possibilities expanded in the late 1990s into the 2000s, researchers started attempting to get more out of soil surveys (Bui et al., 1999; Bui & Moran, 2001; De Bruin et al. 1999; McBratney, 1998; Zhu et al., 1996). Concurrently, the 'digital soil mapping' field emerged and started developing more tools that could be used to modify soil maps (Scull et al., 2003; McBratney et al., 2003). Some of the earlier attempts at disaggregation were aimed at capturing expert knowledge of the soil surveyor in a more formal GIS setting (Qi & Zhu, 2003; Shi et al., 2004; Zhu, 1997; Zhu et al., 1997; Thompson et al., 2010). Many studies looked at the application of machine learning approaches with a large number of studies focusing on classification tree type applications (Bui and Moran, 2001; Hansen et al., 2009; Häring et al. 2012; Moran and Bui, 2002; Nauman & Thompson, 2014; Odgers et al., 2014; Subburayalu et al., 2014; Wei et al., 2010; among others). Most of these studies use attribute information about how the soils vary within map unit delineations to help train classification tree learning models, which can be a complicated process, but several studies skipped this training step in order to streamline the disaggregation workflow (Odgers et al., 2014; Wei et al., 2010). These studies randomly allocate training pixels weighted according to the proportion they represent in the original soil survey map unit description, and rely on a classification tree ensembles to pick out patterns in order to converge on a correct classification result (DSMART algorithm) (Odgers et al., 2014; Wei et al., 2010). The reliance on machine learning in these studies when additional training information is available begs the question of whether accuracy could be improved when training is more directed (e.g. Nauman & Thompson, 2014). Comparisons of these models are needed to better understand this trade off. The advantage of the DSMART pure machine learning approach is its ease in application across large datasets; all that is needed is the original soil map, the list of soils and their proportions present in each map unit and a set of environmental rasters to use in modeling (Odgers et al., 2014; Chaney et al., 2014). Chaney et al. (2014) used DSMART to disaggregate the gSSURGO dataset for the entire lower 48 U.S. states using high performance computing, which is currently being written up for peer review. Ultimately, disaggregated soil maps can be used to get better maps of soil properties that can be used in many applications (Nauman et al., 2012; Odgers et al., 2015).

### Research papers:


