

Automatic stratification for an agricultural area frame using remote sensing data

Stephanie Zimmer*

Iowa State University, Ames, Iowa, USA, sazimme2@iastate.edu

Jae Kwang Kim

Iowa State University, Ames, Iowa, USA, jkim@iastate.edu

Sarah Nusser

Iowa State University, Ames, Iowa, USA, nusser@iastate.edu

The National Agricultural Statistics Service (NASS) is responsible for conducting monthly and annual surveys and preparing official USDA data and estimates of production, supply, prices, and other information necessary to maintain orderly agricultural operations. One survey is the June Area Survey which utilizes an area sampling frame to collect data used to supply direct estimates of acreage and measures of sampling coverage. The population consists of all land in the USA, except Alaska, which is stratified for sampling. Currently, the stratification is labor intensive because segments on the land are hand drawn and stratified by individuals. Our goal is to make this process automatic by using permanently defined segments and auxiliary data from remote sensing, the Cropland Data Layer which classifies satellite imagery into crop types and non-agriculture categories. Using this data, we propose treating stratum identifier as a missing data value and use an expectation-maximization (EM) algorithm to assign the stratum indicator. Each segment is comprised of many pixels and segments of similar composition will be grouped together using the algorithm.

Key Words: EM algorithm, Multivariate stratification, Missing data