



Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

A matching procedure to improve k-NN estimation of forest attribute maps

F. Baffetta^a, P. Corona^{b,*}, L. Fattorini^a^a Dipartimento di Economia Politica e Statistica (DEPS), Università di Siena, P.za S. Francesco 8, 53100 Siena, Italy^b Dipartimento per l'Innovazione nei Sistemi Biologici, Agroalimentari e Forestali (DIBAF), Università della Tuscia, via San Camillo de Lellis, 01100 Viterbo, Italy

ARTICLE INFO

Article history:

Available online 20 July 2011

Keywords:

Forest management planning and scenario analysis
 Design-based inference
 0-Inflated beta distribution
 Monte Carlo study

ABSTRACT

The integration of forest inventory and mapping has emerged as a major issue for assessing forest attributes and multiple environmental functions. Associations between remotely sensed data and the biophysical attributes of forest vegetation (standing wood volume, biomass increment, etc.) can be exploited to estimate the attribute values for sampled and non-sampled pixels, thus producing maps for the entire region of interest. Among the available procedures, the k-nearest neighbours (k-NN) technique is becoming popular, even for practical applications. However, the k-NN estimates at the pixel level tend to average towards the population mean and to have suppressed variance, since large values are usually underestimated and small values overestimated. This tendency may be detrimental for k-NN applications in forest resource management planning and scenario analysis where the representation of the spatial variability of each attribute of interest across the surveyed territory is fundamental. The present paper proposes a procedure to tackle such an issue by modifying k-NN estimates via a post-processing procedure of distribution matching. The empirical distribution function of the population values is estimated from the sample of ground data by using the 0-inflated beta distribution as the assisting model and the k-NN estimates are subsequently modified in such a way as to match the estimated distribution. The statistical properties of the distribution matching estimators for totals and averages are theoretically derived, while the performance of the distribution matching estimator at the pixel level are empirically evaluated by a simulation study.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Sound forest ecosystem management depends on accurate, complete, and concise information regarding the extent, condition and productivity of the natural resources. Estimation of these attributes is a common objective of forest monitoring and assessment programs at a variety of spatial scales (e.g., Kleinn, 2002; Köhl et al., 2006).

The integration of forest inventory and mapping has emerged as a major issue for assessing forest attributes and multiple environmental functions (Corona, 2010). Associations between remotely sensed data and the biophysical attributes of forest vegetation (standing wood volume, biomass increment, etc.) can be exploited to estimate the attribute values for sampled and non-sampled pixels, thus producing maps for the entire region of interest.

Several techniques can be exploited for coupling information from digital remotely sensed imagery and ground sample data. Among non-parametric approaches, the k-nearest neighbours (k-NN) technique is becoming popular (e.g. McRoberts and Tomppo, 2007; Chirici et al., 2008), even for practical applications.

The data necessary to perform k-NN estimation of forest attribute values are the numerical values from remotely sensed imagery, which play the role of auxiliary variables and must be known for all the pixels in the study area, and the values of the interest variable (i.e. the forest attribute), which instead are only known for a sample of pixels and in most situations are from forest inventories. Once remotely sensed and ground sample data are obtained, k-NN estimation of the interest variable for a specific pixel is made by a weighted average of the values of the interest variable in the nearest pixels (with respect to a fixed metric in the space of the auxiliary variables) among those included in the sample.

To support forest resource management it is important that the spatial pattern of each attribute be represented across the whole territory. Moreover, in management planning and scenario analysis it is important that the variability of each attribute throughout the study region be adequately represented by the estimated maps (Barth, 2007; Barth et al., 2008). In this perspective, a well-known shortcoming of k-NN estimation at the pixel level is that large values tend to be underestimated, while small values are overestimated. Accordingly, in most cases, the k-NN estimates tend to average towards the population mean with variance smaller than the actual population variance. This tendency is at once apparent by a visual check of the histograms of the k-NN estimates which usually show tails much shorter than those achieved in the

* Corresponding author. Tel.: +39 0761 357425; fax: +39 0761 357389.

E-mail address: piermaria.corona@unitus.it (P. Corona).