Predicting soil depth using survival analysis models

Quentin Styc, Philippe Lagacherie

Predicting soil depth using survival analysis models

Quentin Styc[1] and Philippe Lagacherie[2]

[1] INRA, UMR LISAH, Montpellier, France

Soil depth is a very important soil property for agronomists and environmental scientists and has been included as such in the GlobalSoilMap specifications (Arrouays et al, 2014). The Digital Soil Mapping applications that deal with soil depth are all confronted with the problem of the "right-censored" input soil data. This means that a significant proportion of the observed sites that serve as input data for Digital Soil Mapping models are characterized by a maximal depth of observation that is smaller than the real soil depth. The information on soil depth at such sites is not an exact value but an inequality (soil depth ⩾ site observation depth).

Such situations are frequently encountered in survival analysis that aim to predict the expected duration of time until one or more events happen, such as death in biological organisms and failure in mechanical systems. This very active branch of statistics has produced a lot of statistical inference models for dealing with right-censored data. By doing an analogy between time at "depth" or "failure" and soil depth ("depth at soil end"), such models could be applied to the digital mapping of soil depth.

In this paper, we apply two survival analysis models, namely Cox Regression (Andersen and Gill, 1982) and Random Survival Forest (Ishwaran et al, 2008) for predicting soil depths in two study areas of southern France with contrasted extents and soil observation densities (Languedoc Roussillon Region,and Peyne Watershed). The soil depth prediction results provided by these two models are evaluated over independent validation sets and compared with classical DSM models using as input i) the sites with a known exact soil depth values only or ii) the whole set of sites with the replacement of the censored soil depth by a « common sense » estimation.

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes, a large sample study. Annals of Statistics 10, 1100-1120.

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca Santos, M.L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.-L., 2014. Chapter Three – GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties, in: Advances in Agronomy. pp. 93–134.

Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S., 2008. Random survival forests. Ann. Appl. Stat. 2, 841–860