



## **Lots of legacy soil data are available, but which data do we need to collect for regional land use analysis?**

Chantal Hendriks (1), Jetse Stoorvogel (1), Lieven Claessens (1,2)

(1) Soil Geography and Landscape Group, Wageningen University, Wageningen, the Netherlands, (2) International Crops Research Institute for the Semi-Arid Tropics, Nairobi, Kenya

In the past, soil surveying techniques were mainly developed for qualitative regional land use analysis (RLUA) like land evaluation and land use planning. Conventional soil survey techniques usually describe soil types according to a soil classification scheme (e.g. Soil Taxonomy and World Reference Base). These soil surveys met the requirements of qualitative land evaluation and land use planning. However, during the last decades there is an increased need for quantitative RLUA resulting in an increased demand for quantitative soil data. The rapid developments in computing technology and the availability of auxiliary information (e.g. remote sensing and digital elevation models) allowed for the development of new soil surveying techniques like digital soil mapping. These new soil surveying techniques aim to produce continuous maps of quantitative functional soil properties. However, RLUA nowadays requires soil data that include a description of the spatial variability of the entire pedon in which correlations between soil properties are retained. Current surveying techniques do not fully fulfil these requirements resulting in a gap between the supply and demand of soil data in RLUA. The gap is caused by the fact that legacy soil data are collected for different purposes and inherently have different assumptions on e.g., soil variability. In this study, some of these assumptions are tested and verified using primary soil data collected during a recent field survey in Machakos and Makueni County (Kenya). Subsequently an ongoing RLUA, the Global Yield Gap Atlas (GYGA) project, is taken as a case study to evaluate the effect of different sources of soil data on the results of the RLUA. The results of the study show that various assumptions underlying the soil survey hamper the quality requirements of soil data for the specific objectives of the RLUA. To give a few examples: mapping soil properties individually ignores correlations between them, soil properties differed significantly between natural and agricultural land, discrete soil mapping units described by a representative soil profile showed internal variability. None of the legacy datasets fitted the requirements of the RLUA. However, resources to collect additional primary soil data are limited. Evaluating legacy data allows us to identify the soil data that we need to collect. Legacy data lack information on e.g. soil management and effective rooting depth, while these data are often required for RLUA. This results in the use of assumptions, estimations and simplifications in a RLUA. The choice of legacy data has a profound effect on the results of a RLUA. The GYGA case study shows for example that different sources of soil input data can lead to differences in simulated water-limited maize yields of up to 3 ton/ha.