Europ. J. Agronomy 74 (2016) 173–184

Contents lists available at ScienceDirect

European Journal of Agronomy

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

Mapping crop diseases using survey data: The case of bacterial wilt in bananas in the East African highlands

H. Bouwmeester[a,*], G.B.M. Heuvelink[b,c], J.J. Stoorvogel[c]

[a] GeoSpace, Roseboomlaan 38, 6717 ZB Ede, The Netherlands
[b] ISRIC - World Soil Information, PO Box 353, 6700 AJ Wageningen, The Netherlands
[c] Wageningen University, Soil Geography and Landscape Group, PO Box 47, 6700 AA Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form
21 December 2015
Accepted 22 December 2015
Available online 5 January 2016

Keywords:
Agriculture
Geostatistics
Indicator regression kriging
Sampling design
Spatial aggregation

ABSTRACT

Globally, crop diseases result in significant losses in crop yields. To properly target interventions to control crop diseases, it is important to map diseases at a high resolution. However, many surveys of crop diseases pose challenges to mapping because available observations are only proxies of the actual disease, observations often are not normally distributed and because typically convenience sampling is applied, leading to spatially clustered observations and large areas without observations. This paper addresses these challenges by applying a geostatistical methodology for disease incidence mapping. The methodology is illustrated for the case of bacterial wilt of banana (BWB) caused by *Xanthomonas campestris pv. musacearum* in the East African highlands. In a survey using convenience sampling, 1350 banana producing farmers were asked to estimate the percentage yield loss due to bacterial wilt. To deal with the non-normal distribution of the data, the percentages were classified into two binary variables, indicating whether or not the disease occurred and whether or not the yield loss was severe. To improve the spatial prediction of disease incidence in areas with low sampling density, the target variables were correlated in a logistic regression to a range of environmental variables, for which maps were available. Subsequently, the residuals of the regression analysis were interpolated using simple kriging. Finally, the interpolated residuals were added to the regression predictions. This so-called indicator regression kriging approach resulted in continuous maps of disease incidence. Cross-validation showed that the method yields unbiased predictions and correctly assesses the prediction accuracy. The geostatistical mapping is also more accurate than conventional mapping, which uses the mean of observations within districts as the predicted value for all locations within the district, although the accuracy improvement is not very large. The maps were also spatially aggregated to district level to support regional decision-making. The analysis showed that the disease is widespread on banana farms throughout the study area and can locally reach severe levels.

1. Introduction

Despite years of agricultural research, diseases, pests, and weeds still seriously hamper agricultural productivity. In literature reviewed by Savary et al. (2012) direct yield losses caused by diseases, animals, and weeds are estimated to be responsible for losses ranging between 20% and 40% of global agricultural productivity. Large yield gaps, i.e., substantially lower actual yields compared to their potential (Van Ittersum et al., 2012; Spiertz, 2012), are particularly found in regions with low external inputs, such as Eastern Africa. Reducing the yield gap in Africa scores high on the political agenda, considering the serious problems around food security (Ozor et al., 2014) and the importance of agriculture for the national economies (agriculture employs the majority of the workforce and contributes significantly to the gross domestic product) (AfDB et al., 2015). Crop diseases are among the main causes of yield gaps (Van Ittersum et al., 2012; Wang et al., 2015). To allow extension officers and policy makers to act better informed and to target pests and diseases efficiently, it is necessary to map the distribution and extent of crop diseases. This supports the design and allocation of proper management measures and effective policies. Regional datasets on crop diseases collected through farm surveys provide valuable information to generate disease maps. However, these datasets have a number of characteristics that hamper mapping (Bouwmeester et al., 2012):

* Corresponding author.
E-mail address: h.bouwmeester@solum.nl (H. Bouwmeester).

http://dx.doi.org/10.1016/j.eja.2015.12.013
1161-0301/

- Surveys rarely focus on diseases directly and often make use of proxies based on symptoms or estimated yield loss. These proxies are subsequently interpreted as indicators of disease incidence. For example, Legg et al. (2006) reviewed the literature reporting on the spread of Cassava Mosaic Virus and show that a range of different methods to describe the virus incidence were used, ranging from ELISA-based diagnostics, nucleic acid-based techniques, but also the registration of whitefly, the vector of CMV, and disease symptoms. Brentu et al. (2012) used the percentage of affected citrus trees with Citrus Black Spot as a proxy for crop loss. Disease symptoms also do not distinguish between different causal agents and may be the result of multiple interacting microbes rather than a single one, which hampers attribution of disease incidence to a single cause, such as BWB.
- Data on crop diseases are often qualitative (e.g., low, medium, high disease severity) or binary (absent or present). For example, Zinga et al. (2013) scored the severity of Cassava Mosaic Disease from 1 (asymptomatic) to 5 (most severely diseased).
- If data are quantitative, they are rarely normally distributed and often present a large number of zeros. These data characteristics require specific statistical methodologies and tailored approaches.
- Data are mostly collected through a convenience sample (e.g., Thompson et al., 2011) as terrain accessibility is problematic and resources do not allow for sampling design optimization for mapping purposes. This often results in an uneven geographical distribution of points and/or clustered data.
- The data often present a high short-distance variability, as illustrated by a case study for palm oil by Rakib et al. (2014).

The above limitations of survey data are often the result of choices in the survey design while dealing with the trade-off between resource availability and data quality requirements. Limited resources may force surveyors to measure proxies of a disease and use convenience sampling. Nevertheless, despite the practical problems with data collection, several studies used survey data to map crop diseases in Eastern Africa. Tusheremereirwe et al. (2006) created maps of bacterial wilt in banana in Uganda, Night et al. (2011) created maps of cassava mosaic disease and cassava brown streak disease in Rwanda, and Legg et al. (2006) created maps for the same cassava diseases for East and Central Africa. Most studies calculate the arithmetic mean of the crop disease parameter in an area delineated by administrative boundaries. However, this conventional method does not address the limitations listed above, and does not quantify the uncertainty associated with the maps.

The objective of this study is to apply a geostatistical methodology to map disease patterns while dealing with the aforementioned limitations in survey design as adequately as possible. Similar approaches but with different limitations and applications are reported in Munar-Vivas et al. (2010), Park et al. (2012), Lamichhane et al. (2013), and Rees et al. (2014). The methodology is illustrated with a case study on bacterial wilt of banana (BWB) caused by *Xanthomonas campestris pv. musacearum* in the East African highlands. BWB is considered to be one of the major threats to banana yields in the region (Blomme et al., 2014; Biruma et al., 2007). The East African highlands is the largest banana producing and consuming region of Africa (FAO, 2012). Up to 60% of the caloric intake of the rural population consists of banana products and bananas are an important cash–crop (Abele et al., 2008). BWB affects all banana cultivars and causes rotting of the banana bunch, yellowing of the leaves and eventually dying of the mother plant. It causes immediate yield losses but also impacts the plant production cycle on the long term. BWB was introduced to Africa in the 1960s in Ethiopia, where it stayed and did not spread until it was reported in Uganda in 2001, from where it quickly spread to the Democratic Republic of Congo, Rwanda, Tanzania, Kenya, and Burundi (Tripathi et al., 2009). Blomme et al. (2013) identify regular surveillance, the supply of clean planting material and phytosanitary measures as crucial to control diseases such as BWB and maintain a healthy banana sector in Africa. Maps on disease distribution and damage are required to guide these management measures. We explore the use of indicator regression kriging for mapping and compare the results with conventional mapping methods. The performance of both mapping methods is evaluated using a visual interpretation and cross-validation.

2. Materials and methods

2.1. Study area

The study area (Fig. 1) is part of the East African highlands and includes Rwanda, Burundi and the main banana growing areas in Western Kenya, Southern Uganda and Northern Tanzania. The East African highlands are part of the African Rift System with Lake Victoria as the central basin. The area has a diverse agro-ecology, resulting from large variation in altitude (546–4610 m above sea level (m.a.s.l.); CGIAR-CSI, 2008), mean annual rainfall (492–2250 mm; Hijmans et al., 2005) and mean annual temperature (from 3.0 to 25.9 °C; Hijmans et al., 2005). A large part of the area is very suitable for growing crops such as banana, maize and cassava because of fertile soils and high rainfall. Agriculture in the region can be characterized as subsistence farming with complex mixed cropping. Areas above 2500 m.a.s.l. are excluded from the study area as these are deemed unsuitable for banana production. The study area is delimited and stratified using administrative boundaries. For Tanzania, Kenya and Uganda the boundaries correspond to district boundaries, for Rwanda to the prefecture boundaries and for Burundi to the provincial boundaries (derived from Hijmans et al., 2013). For simplicity we will refer to the administrative regions as districts in the remainder of this article.

2.2. Farm survey

The disease incidence dataset was collected by the Crisis Crop Control Program (Abele et al., 2007). This program carried out a large scale survey in banana and cassava growing areas in Burundi, Democratic Republic of Congo (DRC), Kenya, Rwanda, Tanzania and Uganda. A total of 2871 farms were surveyed between July 2006 and July 2007. The survey resulted in a database with information on household characteristics, crop yields and food security. BWB incidence was not measured directly in the field but a proxy related to yield loss was included. Farmers were asked to estimate the yield loss due to BWB as a percentage of their total banana yield. Farm locations were recorded using a handheld GPS. From the overall survey, 1350 farms were located within the study area and included banana cultivation. The spatial distribution of these farms (Fig. 1) clearly shows that sampling took place along the major routes using a convenience sample. Descriptive statistics of disease incidence are presented in Table 1. Almost a third of the selected farms were infected with BWB, with some experiencing a 100% yield loss (Fig. 2). However, large regional differences were found. In Uganda, almost 70% of the farmers experienced yield losses. For the other countries yield losses were much smaller, with average losses not exceeding 10% and with median scores of 0%. Locally, however, the losses were considerable and frequently exceeded 50%.

The survey included several of the commonly occurring limitations to disease mapping described in the Introduction: a convenience sample where a proxy of disease incidence in terms of yield loss was used and that presents a non-normal distribution and high short-distance variability. Despite some of the characteristics of the dataset, the conventional approach of describing the spatial