Applied Geography 68 (2016) 84–94

Contents lists available at ScienceDirect

Applied Geography

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

Contextualizing local-scale point sample data using global-scale spatial datasets: Lessons learnt from the analysis of large-scale land acquisitions

Sandra Eckert*, Markus Giger, Peter Messerli

Centre for Development and Environment, University of Bern, Hallerstrasse 10, CH-3012 Bern, Switzerland

ARTICLE INFO

Article history:
Received 1 September 2015
Received in revised form
16 December 2015
Accepted 23 January 2016
Available online 11 February 2016

Keywords:
Geospatial analysis
Geospatial accuracy
Geographic context
Scale
Large-scale land acquisitions

ABSTRACT

This paper examines how the geospatial accuracy of samples and sample size influence conclusions from geospatial analyses. It does so using the example of a study investigating the global phenomenon of large-scale land acquisitions and the socio-ecological characteristics of the areas they target. First, we analysed land deal datasets of varying geospatial accuracy and varying sizes and compared the results in terms of land cover, population density, and two indicators for agricultural potential: yield gap and availability of uncultivated land that is suitable for rainfed agriculture. We found that an increase in geospatial accuracy led to a substantial and greater change in conclusions about the land cover types targeted than an increase in sample size, suggesting that using a sample of higher geospatial accuracy does more to improve results than using a larger sample. The same finding emerged for population density, yield gap, and the availability of uncultivated land suitable for rainfed agriculture. Furthermore, the statistical median proved to be more consistent than the mean when comparing the descriptive statistics for datasets of different geospatial accuracy. Second, we analysed effects of geospatial accuracy on estimations regarding the potential for advancing agricultural development in target contexts. Our results show that the target contexts of the majority of land deals in our sample whose geolocation is known with a high level of accuracy contain smaller amounts of suitable, but uncultivated land than regional- and national-scale averages suggest. Consequently, the more target contexts vary within a country, the more detailed the spatial scale of analysis has to be in order to draw meaningful conclusions about the phenomena under investigation. We therefore advise against using national-scale statistics to approximate or characterize phenomena that have a local-scale impact, particularly if key indicators vary widely within a country.

1. Introduction

Geospatial information technology is ideally suited for rigorous data collection and analysis. It provides users with a set of tools for assessing and analysing aggregated data (such as land concession inventories or socio-economic statistics) at different scales, and for investigating their topology and spatial characteristics, their non-spatial attributes, and any combinations of the two (Burrough & McDonnell, 1998). However, analyses and policy evidence generated by means of a geographic information system (GIS) are only as accurate as the original data inputted into the GIS. The main characteristics that determine the quality of geospatial data are accuracy, precision, consistency, and completeness (Veregin, 1998). The quality of spatial and non-spatial data may be compromised by uncertainties stemming from a variety of sources, including measurement errors, inadequate content definitions, low spatial resolution, and insufficient samples (Burrough & McDonnell, 1998; Heuvelink, 1998; Jones, 1997). If spatial data inputted into a GIS operation are unsuitable or flawed, this will automatically affect the output, potentially to the point where it becomes too unreliable for drawing broader conclusions (Heuvelink, 1998, 2002; Zhang & Goodchild, 2002).

Besides uncertainty related to spatial data themselves, additional errors and uncertainty may be introduced during data processing (Rae, Rothley, & Dragivevic, 2007). Further, two methodological challenges need to be considered that are relevant

* Corresponding author.
E-mail addresses: sandra.eckert@cde.unibe.ch (S. Eckert), markus.giger@cde.unibe.ch (M. Giger), peter.messerli@cde.unibe.ch (P. Messerli).

http://dx.doi.org/10.1016/j.apgeog.2016.01.008

to any geospatial analysis: (1) analysts' choice of a zoning scheme and/or geographic scale of areal units, also known as modifiable areal unit problem (MAUP); and (2) analysts' choice of how to geographically delineate and define a specific context area (whose location and extent may change over time), introduced by Kwan (2012) as the uncertain geographic context problem (UGCoP). Both problems influence the outcome of the analysis in terms of how accurately the characteristics attributed to the places analysed describe the actual reality on the ground.

All of these methodological challenges arose in our own recent geospatial analysis examining large-scale land acquisitions and the socio-ecological characteristics of the areas they target (Messerli, Giger, Dwyer, Breu, & Eckert, 2014). In the present paper we will therefore use that study as an example to substantiate and illustrate our methodological reflections. Large-scale land acquisitions are a phenomenon that can be observed all over the world, and therefore on a global scale; but their implications on the ground may vary from case to case, necessitating analysis on a local scale. Our 2014 study aimed at a global-scale analysis of the local-scale effects that land deals have on the ground, in the geographic contexts they target. Accordingly, it required thematic spatial datasets that offered global coverage – at the expense of accuracy and level of detail. The study also required large and detailed samples of large-scale land acquisitions, in which each land deal had to be geo-located, preferably with known geospatial accuracy and high reliability. Based on these requirements, we decided to use global raster datasets of land cover, population density, and two indicators of agricultural potential. Data on land deals were retrieved from the largest and most detailed global database on large-scale land acquisitions, the Land Matrix database, taking into account their inherent scales and accuracy levels.

With a view to the present study, we then formulated the following two methodological research questions: How are our conclusions about the geographic contexts targeted by large-scale land acquisitions influenced by our sample's geospatial accuracy? And how are they affected by the sample's size, that is, the number of land deals we include in our analysis? To achieve the most accurate results, should researchers aim to collect as many samples as possible of acceptable accuracy, or should they focus on a smaller sample of high geospatial accuracy? To address these questions, we analysed land deal datasets of varying geospatial accuracy and varying sizes and compared the results. More specifically, to address the first question, we tested the use of different areal extents of target contexts – that is, different buffer sizes – representing different degrees of geospatial accuracy, and examined how this affected our conclusions about the target contexts' socio-ecological characteristics. To address the second question, we retained a certain buffer size while testing the use of differently sized samples – that is, including more or fewer land deals – and examined how that affected our resulting conclusions. In this paper, we not only present findings specific to our example study of large-scale land acquisitions and their target contexts, but also look beyond the example to draw general conclusions about the implications of geospatial uncertainty and sample size for geospatial data analysis. The paper ends with recommendations for future research in land science and geography that involves spatially explicit information.

2. Materials and methods

2.1. Land Matrix data

Data on large-scale land acquisitions were taken from the global Land Matrix database, which is operated by the Land Matrix Partnership (Anseeuw, Lay, Messerli, Giger, & Taylor, 2013), a joint initiative of several research and development organizations.[1] The Land Matrix Partnership has been collecting data on large-scale land acquisitions since 2009. It focuses on land deals that: (i) involve the sale, lease, or concession of land; (ii) entail a transfer of user rights in land from smallholders and communities to commercial users; (iii) cover an area greater than 200 ha; and (iv) have been announced or concluded since the year 2000 (Anseeuw et al. 2012, 2013). Our example of a geospatial analysis focused on international large-scale land deals for agricultural purposes and aimed at an improved understanding of the geographic contexts they target. For that reason, we looked exclusively at land deals that serve agricultural purposes and involve international investors. Moreover, we looked only at deals which had not failed, that is, for which the negotiation status was reported as "intended" (expression of interest, or contract under negotiation) or "concluded" (contract signed, or subject to oral agreement). This resulted in an initial sample of 892 land deals. The positional accuracy of each deal was verified and, wherever possible, improved by consulting the local partner organizations who had collected the data, as well as print and online media and recent Google Earth satellite imagery. Next, we grouped the deals into three classes based on their level of geospatial accuracy, namely: *high* (139 deals with an accuracy < 10 km); *intermediate* (408 deals with an accuracy of 10–100 km); and *low* (345 deals with only the host country known). Each level of spatial accuracy corresponds to a geospatial scale of analysis: high accuracy enables analysis at the local scale, intermediate accuracy enables analysis at a regional (i.e. subnational) scale, and low accuracy corresponds to the national scale. The dataset used in this study was exported on 7 April 2013.

2.2. Geospatial datasets

We analysed the target contexts of large-scale land acquisitions in terms of (1) *land cover*, (2) *population density*, and (3) two indicators for *agricultural potential* – namely yield gap and availability of uncultivated land that is suitable for rainfed agriculture. *Land cover* was analysed using GlobCover 2009. This dataset has a spatial resolution of 300 m and was processed from Medium Resolution Imaging Spectrometer (MERIS) full resolution data collected between 1 January and 31 December 2009. Land cover was classified according to the classification system of the United Nations Food and Agriculture Organization (FAO) (Arino, 2010). *Population density* was analysed using the latest edition of the Oak Ridge National Laboratory (ORNL) LandScanTM database; displaying an approximate resolution of 1 km at the equator, this product offered the highest available resolution for global population distribution data (ORNL, 2013). We analysed yield gaps using the latest freely accessible dataset of the International Institute for Applied Systems Analysis (IIASA); this dataset was created based on the global agro-ecological zoning (GAEZ) method (Fischer, Hiznyik, Priler, Shana, & van Vethuizen, 2002) and indicates the gap between actual rainfed yields and potential yields for five major crops, taking into account local agro-ecological conditions. The IIASA dataset has a spatial resolution of approximately 10 km. Unfortunately, we had to use the dataset for the year 2000, as no newer version was available at the time we conducted our analysis (IIASA, 2010).

Our spatial dataset on the availability of uncultivated land

[1] The Land Matrix initiative is coordinated by five main partners: ILC (International Land Coalition), CIRAD (Centre de coopération internationale en recherche agronomique pour le développement), CDE (Centre for Development and Environment, University of Bern), GIGA (German Institute of Global and Area Studies), and GIZ (Deutsche Gesellschaft für Internationale Zusammenarbeit GmbH).