G Model
EURAGR-25556; No. of Pages 17
ARTICLE IN PRESS
Europ. J. Agronomy xxx (2016) xxx–xxx
Contents lists available at ScienceDirect
European Journal of Agronomy
journal homepage: www.elsevier.com/locate/eja

Review of yield gap explaining factors and opportunities for alternative data collection approaches

Eskender Beza[a,b,*], João Vasco Silva[b], Lammert Kooistra[a], Pytrik Reidsma[b]

[a] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands
[b] Department of Plant Sciences, Group Plant Production Systems, Wageningen University, P.O. Box 430, 6700 AK Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form 25 April 2016
Accepted 30 June 2016
Available online xxx

Keywords:
Yield variability
Potential yield
Actual yield
Benchmarking
Crowdsourcing
Data collection

ABSTRACT

Yield gap analysis is gaining increased scientific attention, as estimating and explaining yield gaps shows the potential for sustainable intensification of agricultural systems. Explaining yield gaps requires detailed information about the biophysical environment, crop management as well as farm(er) characteristics and socio-economic conditions in which farmers operate. However, these types of data are not always available, mostly because they are costly to collect. The main objective of this research is to assess data availability and data collection approaches for yield gap analysis, and to summarize the yield gap explaining factors identified by previous studies. For this purpose, a review of yield gap studies (50 agronomic-based peer-reviewed articles) was performed to identify the most commonly considered and explaining factors of the yield gap. Besides a global comparison, differences between regions, crops and methods were analysed as well.

The results show that management and edaphic factors are more often considered to explain the yield gap compared to farm(er) characteristics and socio-economic factors. However, when considered, both farm(er) characteristics and socio-economic factors often explain the yield gap. Fertilization and soil fertility factors are the most often considered management and edaphic factors. In the fertilization group, factors related to quantity (e.g. N fertilizer quantity) are more often considered compared to factors related to timing (e.g. N fertilizer timing). However, when considered, timing explained the yield gap more often.

Explaining factors vary among regions and crops. For example, while soil fertility is considered relatively much both in Africa and Asia, it is often explaining in Africa, but not in Asia. Agronomic methods like crop growth simulation models are often used for yield gap analysis, but are limited in the type and number of factors that can be included. Qualitative methods based on expert knowledge can include the largest range of factors.

Although the data included in yield gap analysis also depends on the objective, knowledge of explaining factors, and methods applied, data availability is a major limiting factor. Bottom-up data collection approaches (e.g. crowdsourcing) involving agricultural communities can provide alternatives to overcome this limitation and improve yield gap analysis.

1. Introduction

Sustainable intensification of agricultural systems, including the closure of existing yield gaps on currently available agricultural land, has been pointed as a possible pathway to meet the future food demand (Cassman, 1999). The concept of 'yield gap' is based on production ecological principles and can be estimated as the difference between a benchmark (e.g. climatic potential or water-limited yield) and the actual yield (van Ittersum and Rabbinge, 1997). This concept is particularly important because it indicates the biophysical potential available to improve agricultural production in a specific location (van Ittersum et al., 2013).

Yield gap analysis provides the foundation for identifying the most important crop, soil and management factors limiting current farm yields (van Ittersum et al., 2013; Lobell et al., 2009; Tittonell et al., 2008; Lobell et al., 2005). Information on the magnitude of the yield gap, and associated explaining factors, is important for efficiently targeting efforts to increase crop production in a particular

* Corresponding author at: Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands.
E-mail address: eskender.beza@wur.nl (E. Beza).

http://dx.doi.org/10.1016/j.eja.2016.06.016
1161-0301/

Please cite this article in press as: Beza , E., et al., Review of yield gap explaining factors and opportunities for alternative data collection approaches. Eur. J. Agron. (2016), http://dx.doi.org/10.1016/j.eja.2016.06.016

farming system (Affholder et al., 2013). For example, a yield gap analysis for cassava in Cambodia revealed that soil nutrients, short crop duration and weed infestation explained much of the yield gap in the area and these factors had to be improved to increase cassava yield (Sopheap et al., 2012). A number of yield gap analysis studies have been conducted for different crops in different agro-ecological conditions (van Ittersum and Cassman, 2013) and the results of these studies showed that the magnitude and factors that cause the yield gap vary among locations (e.g. Affholder et al., 2013).

Many studies have examined yield gaps at the scale of the region or agro-climatic zone, using aggregated data on crop yields and explaining factors (e.g. Mueller et al., 2012; Neumann et al., 2010). These type of studies are useful to compare different regions in relative terms using harmonised data (van Ittersum et al., 2013). However, in order to further understand yield gaps, more local studies are needed to bring the role of farm and farmer characteristics into the picture, as well as more local biophysical and socio-economic conditions (Silva et al., this issue).

The interactions between different activities at farm level, together with resource constraints faced by individual farmers, likely explain why inputs are not optimally allocated across the farm and hence why yield gaps persist (e.g. Tittonell et al., 2008). Therefore, yield gap analysis at farm and farming system level can contribute to better understand whether or not yield gaps can be closed and if so, under which production, economic and environmental conditions (Giller et al., 2006). A major drawback of this type of analysis is the high data standards required which typically refer to (a) large sample size, (b) fine resolution and (c) great level of detail. Clearly, obtaining information about biophysical characteristics and crop and farm management for individual agricultural activities within a farm, as well as farm and farmer's characteristics and socio-economic conditions for a large number of farms is costly and time-consuming. Nowadays, the proliferation of computing devices like different types of mobile phones equipped with sensors (e.g. GPS), and other similar technologies makes it possible to implement effective and low-cost "bottom-up" data collection approaches such as crowdsourcing (Ferster and Coops, 2013). These innovative methodologies facilitate the collection of relatively large amounts of information directly from local communities (Herrick et al., 2013; Pratihast et al., 2012).

The main objective of this research was to review the yield gap explaining factors identified by previous studies, in order to assess data availability and suggest improved data collection approaches for yield gap analysis. To address this specific objective, the following steps were undertaken: (1) to provide an overview of factors considered and explaining yield gaps; (2) to identify most commonly considered and often explaining factors of the yield gap at the global, regional and crop levels; (3) to investigate if there are regional similarities or differences in the factors which are commonly considered and explaining yield gaps; (4) to identify the most common data sources for the different factors considered for yield gap analysis; (5) to evaluate to which extent innovative data acquisition methods (e.g. crowdsourcing) are relevant for improving data availability.

2. Methodology

2.1. Literature search and study selection

A detailed literature search was carried out as starting point for this review. The selection of papers was made through specific searches for peer-reviewed articles on yield gap analysis in agronomic journals with key words "yield gap", "potential yield", "yield variability", "water-limited yield" and "yield gap variability". The initial focus was on a special issue released by Field Crops Research on yield gap analysis (van Ittersum and Cassman, 2013). In addition, whenever peer-reviewed articles related to yield gap analysis were found in the reference list of an already reviewed article, they were analysed and included for our study. However, priority was given to articles which explained yield gaps and/or yield variability rather than only estimating the yield gap. The review was not completely systematic, as using a keywords-based approach resulted in a large amount of papers that were not directly relevant for this review, as they did not explain yield gaps. Although some relevant papers may be missing due to this, the selected papers provide a good basis to reach our objectives.

2.2. Review of studies and construction of database

A database was created using MS-Excel 2010 in order to store the information from the selected articles. The database consists of five different tables, namely: "yield gap", "determining factors", "considered factors", "explaining factors" and "validation table". Each of the tables was organised in such a way that information about the five main categories climate, edaphic, management, farm characteristics and socio-economic factors were stored separately. All of the tables were linked with unique identifiers (IDs) to facilitate information retrieval.

Specific information about the study locations including the continent, country, administrative region and site names and their respective coordinates were compiled in the "yield gap" table. When the coordinates of the study locations were not provided, the names of the study locations were used as a geographic reference and Google Maps was used to obtain the approximate coordinates of the study locations. In addition, information about the level at which the yield gap was estimated and explained (e.g. farm, field, regional or global level), resolution of data collection and the types of crops grown were also compiled. In this table, we also included the years in which the yield gap analysis was performed, the data sources used to estimate both actual and benchmarking yields as well as the methods used to estimate the benchmarking yield (e.g. name of crop model) and the term(s) used to indicate the benchmarking yield (e.g. potential yield, attainable yield, water-limited yield or economic yield). For studies that explained the yield gap, the explanatory methods used to explain the yield gap/yield variability (e.g. boundary-line, linear regression) were included in the database as well. Finally, the purpose of the different methods (for e.g. to explain yield gap or yield variability) used within each paper were recorded.

For each of the methods used in a specific paper, the dependent variable (Y) and the independent variables (X) were identified and included in the database. The independent variables were included in the "considered factors" table of the database. Out of these "considered factors", the ones which explained part of the yield gap and/or yield variability according to the criteria set by the specific paper were included in the "explaining factors" table.

In order to determine the number of records (entries) per study, the following criteria were used: number of crops considered, number of locations, years in which the yield gap analysis was performed, and methods used to estimate the benchmarking yield and to explain the yield gap. One record is a unique combination of location × crop × year × benchmark yield estimation method × yield gap explanatory method. A total of 270 records with unique identifiers (IDs) were included into the database. For studies which explicitly provided the actual yield (Ya) and the benchmarking yield (Yp or Yw), the magnitude of the yield gap (%) was calculated as the difference between the benchmark yield and the actual yield divided by benchmark yield times 100%. For studies which didn't provide the values explicitly, we didn't calculate the percentage of the yield gap and it was left blank in the database.