Detecting the linkage between arable land use and poverty using machine learning methods at global perspective

Citation for published version:
Tian, F, Wu, B, Zeng, H, Watmough, GR, Zhang, M & Li, Y 2022, 'Detecting the linkage between arable land use and poverty using machine learning methods at global perspective', *Geography and Sustainability*. https://doi.org/10.1016/j.geosus.2022.01.001

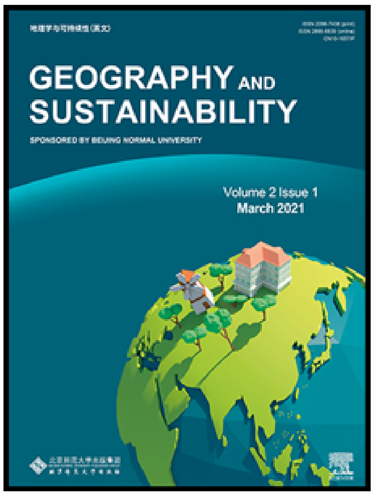
Digital Object Identifier (DOI):
10.1016/j.geosus.2022.01.001

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Version created as part of publication process; publisher's layout; not normally made publicly available

Published In:
Geography and Sustainability

Journal Pre-proof



Detecting the linkage between arable land use and poverty using machine learning methods at global perspective

Fuyou Tian , Bingfang Wu , Hongwei Zeng , Gary R Watmough , Miao Zhang , Yurui Li

PII: S2666-6839(22)00008-6
DOI: https://doi.org/10.1016/j.geosus.2022.01.001
Reference: GEOSUS 75

To appear in: *Geography and Sustainability*

Received date: 7 October 2021
Revised date: 12 January 2022
Accepted date: 13 January 2022

Please cite this article as: Fuyou Tian , Bingfang Wu , Hongwei Zeng , Gary R Watmough , Miao Zhang , Yurui Li , Detecting the linkage between arable land use and poverty using machine learning methods at global perspective, *Geography and Sustainability* (2022), doi: https://doi.org/10.1016/j.geosus.2022.01.001

Highlights:

1. We investigated the arable land use and poverty links from a global perspective.

2. Arable land use was assessed from three aspects with support of big geodata.

3. Non-parametric machine learning methods was applied to explore the linkage.

4. We find that RAPY, RPCPA, fertilizer consumption was highly related with poverty.

Detecting the linkage between arable land use and poverty using machine learning methods at global perspective

Fuyou Tian[1], Bingfang Wu[1,2,*] wubf@aircas.ac.cn, Hongwei Zeng[1,2], Gary R Watmough[3], Miao Zhang[1], Yurui Li[4,5]

[1]State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101

[2]University of Chinese Academy of Sciences, Beijing 100049

[3]School of Geosciences, University of Edinburgh, EH8 9XP Edinburgh, United Kingdom

[4]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China

[5]Key Laboratory of Regional Sustainable Development Modelling, Chinese Academy of Sciences, Beijing, 100101, China

[*]Correspondence Author: Prof Bingfang Wu

Abstract

Eradicating extreme poverty is one of the UN's primary sustainable development goals (SDG). Arable land is related to eradicating poverty (SDG1) and hunger (SDG2). However, the linkage between arable land use and poverty reduction is ambiguous and has seldom been investigated globally. Six indicators of agricultural inputs, crop intensification and extensification were used to explore the relationship between arable land use and poverty. Non-parametric machine learning methods were used to analyze the linkage between agriculture and poverty at the global scale, including the classification and regression tree (CART) and random forest models. We found that the yield gap, fertilizer consumption and potential cropland ratio in protected areas correlated with poverty. Developing countries usually had a ratio of actual to potential yield less than 0.33 and fertilizer consumption less than 7.31 kg/ha. Crop extensification, intensification and agricultural inputs were related to poverty at the global level.

Keywords

Arable land use; Poverty; Machine learning; Yield gap; Random forest

1. Introduction

As of 2015, there were 700 million people in the world living in extreme poverty (Steele et al., 2017). Eradicating extreme poverty (SDG1) is the first goal of the UN's Sustainable Development Goals (SDGs) (United Nations, 2015; Jean et al., 2016). The poor usually rely on environmental resources such as water, cropland and forests to ensure their basic livelihoods and subsistence (Dasgupta et al., 2005; Watmough et al., 2016; Tian et al., 2020). Most of them are engaged in agriculture to support themselves (Kassie et al., 2011; McArthur and McCord, 2017; Benfica et al., 2019; Tomich et al., 2019). In Asia, it is estimated that each 1.00% increase in crop productivity could decrease the poverty rate by 0.48% (Thirtle et al., 2003; Pingali, 2012).

However, some studies argue that agriculture exhibits poor performance in poverty reduction (Ellis, 2005; Diao et al., 2010; Harris and Orr, 2014; Dawson et al., 2016). First, global agricultural production has substantially increased (Pingali, 2012) since Asia's Green Revolution began in the 1960s (Evenson and Gollin, 2003), leading to an increase in food production, which addresses SDG2 (no hunger). A surge in the food supply results in decreased international food prices and reduces crop income and wages (Diao et al., 2010; Harris and Orr, 2014) in rural areas, adversely affecting SDG1. Additionally, agriculture development may raise considerable environmental issues, such as agro-ecological degradation (Maxwell and

Slater, 2004; Barrow, 2012) and increased greenhouse gas emissions (Burney et al., 2010). Therefore, the role of agricultural development and enhanced crop productivity in eradicating poverty has become controversial (Diao et al., 2010; Harris and Orr, 2014).

The reasons for these views are complex. Most studies are limited to a single country (Minten and Barrett, 2008; Kassie et al., 2011) or individual households (Harris and Orr, 2014; Leonardo et al., 2018). Few studies have examined the relationship between arable land use and poverty for different poverty levels. Thus, this research investigates the linkage between arable land use and poverty from a global perspective to provide an additional resource for realizing the UN's 2030 SDGs.

Arable lands resources are essential for agricultural production and food security globally (Alene and Coulibaly, 2009). The potential links between poverty and arable land use indicators in previous research (Table 1). Firstly, previous studies in Africa and India have mentioned that the amount of cultivated area is closely related to the poverty rate (Harris and Orr, 2014; Leonardo et al., 2018). Findings from Harris and Orr (2014) suggested that having more than 4.5 hectares of arable land per five-member household can ensure sufficient income in Africa and India. Protected areas, as one type of forbidden area for expanding, have received much attention. Brandon et al. (2005) and Duan et al. (2017) suggest that protected areas may negatively impact neighbouring residents and have proposed solutions to coordinate biodiversity conservation and human development.

Many studies have discussed the relationship between crop yield and poverty. The crop yield gap is closely related to poverty in rural areas of sub-Saharan Africa (Sumberg, 2012; Dzanku et al., 2015; Tian and Yu, 2019). McArthur et al. (2017) also estimated that a half-ton growth in crop yields produces a 14% to 19% increase in GDP per capita and a 4.6% to 5.6% decrease in the agricultural labour share five years later.

Agricultural inputs are important factors for crop yield, such as irrigation, fertilizer and field management. Irrigation and fertilizer are vital agricultural inputs, and many studies have focused on their relationship with poverty (Huang et al., 2006; Hanjra et al., 2009; Pingali, 2012). In sub-Saharan Africa (Hanjra et al., 2009; Burney and Naylor, 2012), China (Huang et al., 2006) and Ethiopia (Liverpool and Winter-Nelson, 2010; Adela et al., 2019), irrigation enhancement was found to be important in reducing poverty. However, a study in Malawi found that small households did not benefit greatly from increased fertilizer consumption (Ricker-Gilbert and Jayne, 2012).

Table 1. Links between arable land use indicates and poverty in previous research

Indicators	Region	Conclusion	Reference
Crop area	African and Indian	Having more than 4.5 hectares of arable land for a five-member household with 5 members can ensure adequate enough income in Africa and India	(Harris and Orr, 2014)
Protected area	Mexico	Protected areas may have a negative impact on neighbouring residents	(Brandon et al., 2005)
Protected area	Sichuan, China	The protected area had an adverse impact on the wealth of local residents	(Duan and Wen, 2017)
Crop yield gap	sub-Saharan Africa	The crop yield gap is closely related to poverty	(Sumberg, 2012; Dzanku et al., 2015; Tian and Yu, 2019)
Irrigation	China	Irrigation could support poverty alleviation.	(Huang et al., 2006)
Irrigation	sub-Saharan Africa	Irrigation should be considered as one of the pathways to break the poverty trap	(Hanjra et al., 2009; Burney and Naylor, 2012)
Fertilizer	Ethiopia	Fertilizer consumption can help to reduce the poverty rate	(Liverpool and Winter-Nelson, 2010; Adela et al., 2019)
Fertilizer	Malawi	Small households did not benefit much greatly from increased fertilizer consumption	(Ricker-Gilbert and Jayne, 2012)

Overall, previous research only investigates single indicator for one country or region. We will investigate the links between poverty and arable land from a global perspective with multiple indicators to see whether the links at the regional scale are applicable globally. This paper attempts to answer the following questions:

1) what are the links between arable land use and poverty on a global scale?

2) wgloballyrences in the linkages between arable land and poverty for different levels of poverty?

3) how can thpoverty levelsting poverty in the UN's 2030 SDGs be achieved from the perspective of arable land use?

2. Materials and methods

2.1 Poverty data

Open access poverty rate data were acquired from the World Bank and the Central Intelligence Agency (CIA) for 161 countries. Data from 2016 are shown in Fig. 1. The poverty rate for 147 countries was obtained from the World Bank. In cases where data were absent for 2016, the average poverty rates for the preceding five years (2010-2015) were used. The poverty rate is estimated with the data from the CIA for another ten countries (https://www.cia.gov/the-world-factbook/countries). The poverty rates of the remaining four countries, Somalia[1], Saudi Arabia[2], Oman[3] and Qatar[4], were obtained from news reports or webpages, which are shown in the footnotes. In 2016, the poverty rates of 161 countries ranged from 0 to 82.5%.

To ensure the accuracy of the fused poverty data, the poverty rate is validated with the infant mortality rate (defined as the number of children that died before age one for every 1,000 live births), which is a proxy for poverty (de Sherbinin, 2008; Barbier and Hochard, 2018). The infant mortality rate for a country was weighed according to the georeferenced infant mortality rate and population density following the formula below.

[1] Poverty rate derived from https://www.borgenmagazine.com/10-facts-poverty-in-somalia/
[2] Poverty rate derived from http://english.alarabiya.net/en/business/economy/2013/11/03/Kingdom-has-tenth-lowest-poverty-rate-worldwide-says-World-Bank.html
[3] Poverty rate derived from http://timesofoman.com/article/78972
[4] Poverty rate derived from https://en.wikipedia.org/wiki/Economy_of_Qatar

$$\text{IMR}_{country} = \frac{\sum \text{IMR}_{pixel} \times population_{pixel}}{\sum population_{pixel}}$$

where:

$\text{IMR}_{country}$ is the average infant mortality rate;

IMR_{pixel} is the georeferenced infant mortality in 2015 obtained from the *Global Subnational Infant Mortality Rates Version 2* (Creator: Center for International Earth Science Information Network - CIESIN - Columbia University, 2019; Publisher: NASA Socioeconomic Data and Applications Center (SEDAC); https://doi.org/10.7927/H4PN93JJ) grid dataset;

$population_{pixel}$ is the population for each pixel in the country in 2015, obtained from the *Gridded Population of the World, Version 4* (GPWv4): Population density, Revision 11(Creator: Center for International Earth Science Information Network - CIESIN - Columbia University, 2016; Publisher: NASA Socioeconomic Data and Applications Center (SEDAC); https://doi.org/10.7927/H49C6VHW).

As a result, the poverty rate was highly related to the infant mortality rate, with a correlation coefficient of 0.53, as shown in Fig. 1c.

To avoid the impact of having a variety of data sources, the poverty rate is divided into four classes according to natural breaks (Chen et al., 2013), which minimizes each class's average deviation from the class while maximizing each class's deviation from the means of the other groups. Four intervals that were adopted are: 0–4.9% (developed), >4.9%–16.0% (somewhat poor), >16.0%–38.6% (moderately poor) and >38.6%–82.5% (developing).

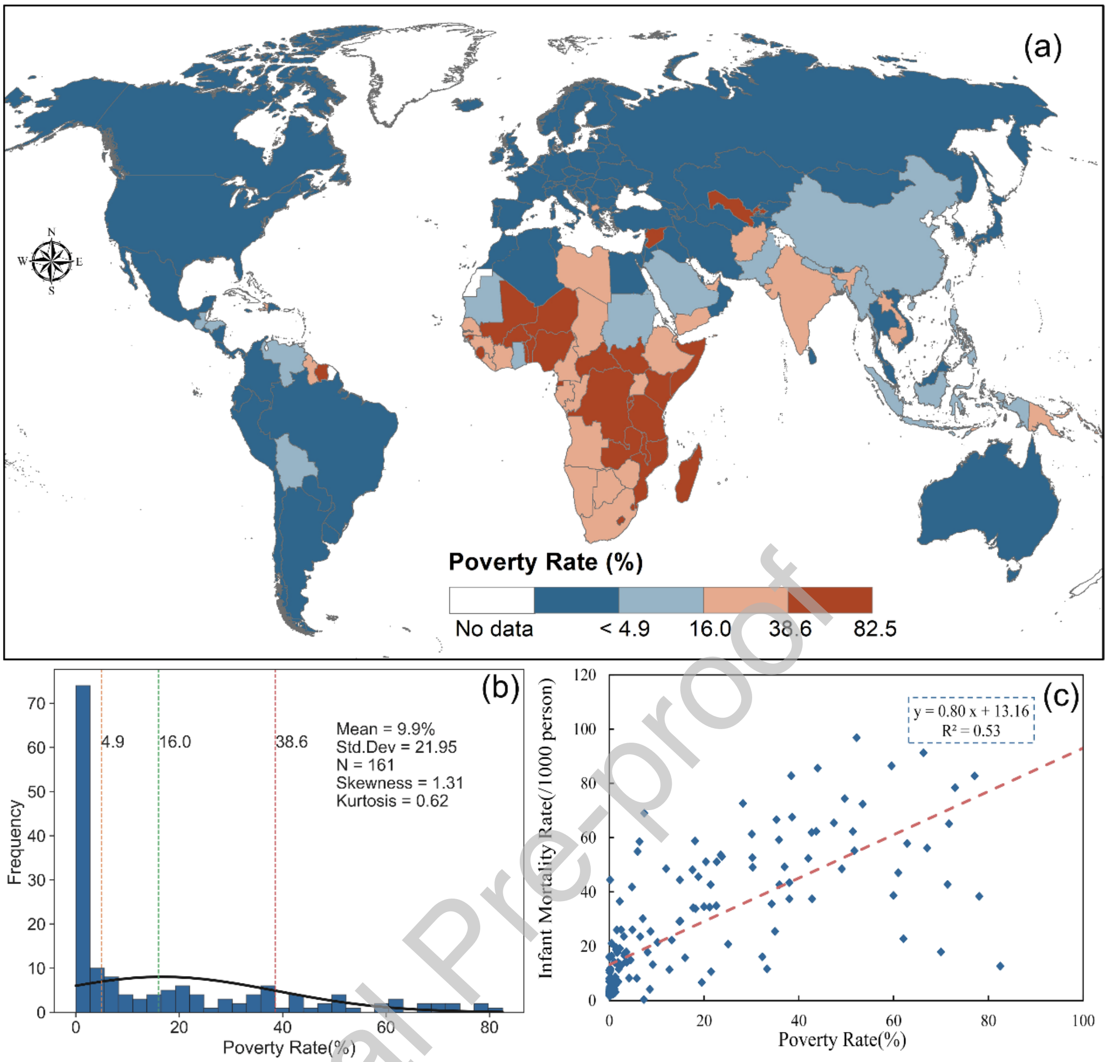


Fig. 1. (a) the global poverty rate in percentage; (b) the distribution of the poverty rate for each country; and (c) the validation of the poverty rate with the global infant mortality rate. The blanks in (a) indicate that no data was available and in (b) that poverty does not follow the normal distribution, as the skewness was 1.31 and kurtosis was 0.62.

2.2 Arable land use assessment

Based on previous research, arable land use was assessed from the perspective of food production. Crop yields and areas are two factors that influence crop production, and both are affected by arable land use. Crop intensification, crop extensification and agricultural input are closely related to crop yields and areas under cultivation. Crop extensification affects the cropped areas, and crop intensification determines crop yields. We will assess arable land use using these three perspectives.

By summarizing indicators from previous research and factoring in available data, we selected: cropland area per capita; the ratio of actual to potential cropland; the ratio of potential cropland in protected area to estimate cropland extensification; the ratio of actual to potential yield to reflect crop intensification; irrigation percentage and fertilizer use to reveal the level of agricultural inputs. The conceptual relationship between food production, arable land use and the poverty rate is shown in Fig. 2a.

2.2.1 Crop extensification

To explore the links between poverty and crop extensification at the global scale, the cropland area per capita indicator, the ratio of actual to potential acreage of cultivated land and the ratio of potential cropland in protected areas (RPCPA) were used.

The cropland area per capita in 2016 was collected from the World Bank (https://databank.worldbank.org/home.aspx), which represents the natural endowment of arable cropland per capita. The ratio of actual cropland to potential cropland (RAPC) indicates the extent of cropland development. The actual cropped land area is calculated by multiplying the population size by the cropped land area per capita. The potential acreage of arable land is calculated according to crop suitability assessed using the GAEZ model, which was jointly developed by the Food and Agriculture Organization of the United Nations (FAO) and the International Institute for Applied Systems Analysis (IIASA). Considering soil suitability and terrain suitability with different agricultural input levels, the GAEZ model evaluated the suitability of land for specific crop types. Maximum suitability is calculated for 17 major crops, including wheat, maize, wetland/dryland rice, soybean, alfalfa, banana, barley, buckwheat, rape, sorghum, sunflower, Phaseolus beans, carrot, cabbage, tomato, and sweet potato (IIASA and FAO, 2012), with high input levels and rain-fed irrigation. Grids with suitability values greater than 2,500 are considered potential cropland.

The RPCPA is the ratio of potential cropland area in protected areas. The area of protected land can have both negative and positive effects on poverty reduction. Therefore, we adopted the RPCPA to explore the relationship between arable land use and poverty reduction. For our analysis, the global protected area (PA) data were downloaded from the World Database on Protected Areas (WDPA). The WDPA is compiled by the United Nations Environment World Conservation Monitoring Center (UNEP-WCMC) and provides the latest and most comprehensive data on global land and marine protected areas and is updated monthly by governments, non-governmental organizations, landowners and communities. These data were combined with potentially arable land obtained from the GAEZ model, which enabled the calculation of the RPCPA.

2.1.2 Crop intensification

Crop intensification indicates how much potential crop productivity has been developed during the cultivation process. Besides increasing cropland area, enhancing crop intensification to increase yields is an effective way to increase crop production. In this study, the ratio of actual to potential yield (RAPY) is used as a direct measure of cropland utilization efficiency and crop intensification. Due to the diversity of crop management conditions, including fertilizer, pest control, sowing harvest (Mauser et al., 2015) and the number of harvests per year (Wu et al., 2018; Jiang et al., 2021), the actual yield is usually less than the potential yield as it can be constrained by environmental conditions. In the GAEZ model, the potential yield was estimated by considering the stress of radiation, temperature, water and soil. In contrast, the actual yield was obtained by down-scaling production from FAO's statistical data in 2005 (IIASA and FAO, 2012). Then, the yield and production gaps were estimated by RAPY, which is available on the GAEZ website (IIASA and FAO, 2012). RAPY for the high input main crop dataset in 2005 was adopted and seven classes were created: <10%, 10%–25%, >25%–40%, >40%–55%, >55%–70%, >70%–85%, and >85%. RAPY in 2016 was calculated according to the following formula:

$$RAPY_{2016} = RAPY_{2005} * (Production_{2016}/Production_{2005}),$$

where:

$RAPY_{2005}$ is the RAPY at the national scale;

$Production_{2016}$ and $Production_{2005}$ are the FAO statistical data of crop production in 2016 and 2005 at the national scale.

The national $RAPY_{2005}$ data were estimated by the weight of each class proportion and median of each class range.

2.1.3 Agricultural inputs

In this study, we explore the linkage between agricultural inputs and poverty on a global scale. The fertilization condition for each country was found in the World Bank databank as fertilizer consumption (kg/hectare of arable land) in 2016. The irrigation percentage was estimated by the ratio of actual irrigated area to potential irrigated area. The actual and potential irrigated area was obtained from the FAO AQUASTAT database (http://www.fao.org/aquastat/en/).

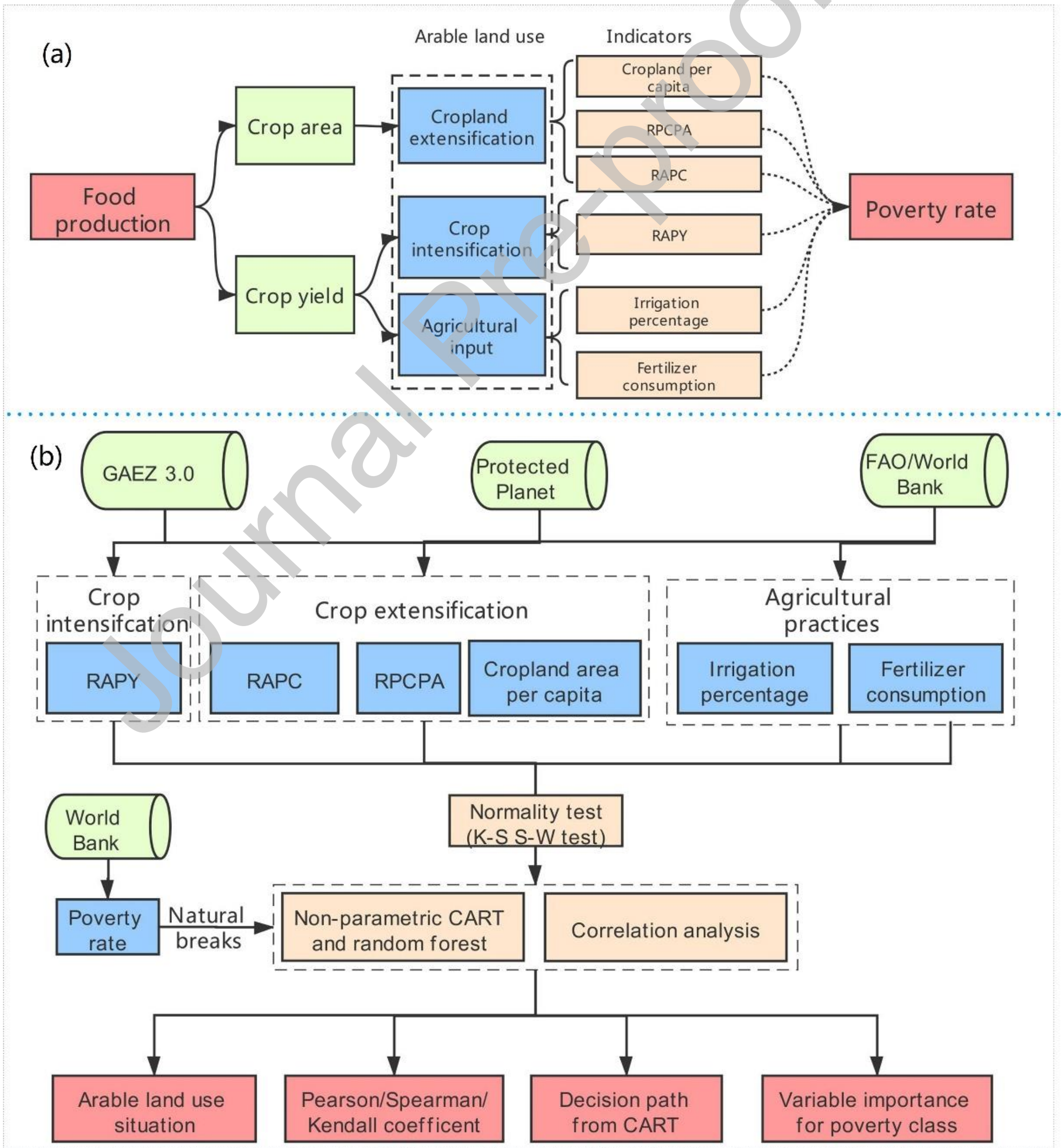


Fig. 2. (a) the conceptual framework between food production and arable land use

assessment (b) workflow used to explore the relationships between poverty and arable land use.
Note: RAPC is the ratio of actual to potential cropland; RAPY is the ratio of actual to potential yield; and RPCPA is the ratio of potential cropland in protected areas.

2.3. Non-parametric Machine Learning method

The internal linkage between arable land use and poverty is complex. Previous studies have assumed that variables follow a specific distribution or try to transform raw variables by using mathematical operations, such as log transformation, to ensure the normality of the variables (Huang et al., 2006; Minten and Barrett, 2008; Kassie et al., 2011; Burney and Naylor, 2012). Few data agree with the standard normal distribution in the real world. This study tries a non-parametric machine learning method to reveal the complex links between different types of data, such as clustering and classification techniques (Han et al., 2011), which has been successfully used to explore the poverty-environment relationship (Watmough et al., 2016; Watmough et al., 2019). This method has no special requirement for data distribution, which ensures a non-bias exploration of the linkage between poverty and arable land use, even though the relationship is non-linear (Han et al., 2011; Müller et al., 2013; Watmough et al., 2019).

Fig. 2b shows the workflow used to link poverty and arable land use. Six indicators were used to explore this relationship as shown in Table 2, along with the hypothesized links they have with poverty. To account for the non-normality in the data and to better understand the linkage between poverty and arable land use, a non-parametric classification and regression tree (CART) (Breiman, 2017) was used to estimate poverty levels using these six indicators. CART enables a fit for a non-linear relationship due to its hierarchies and repeated use of each variable (Han et al., 2011; Watmough et al., 2019), which can enhance the understanding of the links between poverty and arable land use.

Table 2. Indicators used and the hypothesized links they have with poverty

Aspects	Indicators	Definition	Data source	Hypothesized linkage with poverty
Crop intensification	RAPY[1]	$\frac{\text{Actual yield}}{\text{Potential yield}}$	GAEZ[2] & FAO[3]	High RAPY associated with a low poverty rate
Crop extensification	RAPC[4]	$\frac{\text{Actual cropland area}}{\text{Potential cropland area}}$	GAEZ & FAO	High RAPC is associated with a low poverty rate
	Cropland per capita	$\frac{\text{Total cropland area}}{\text{Population}}$	FAO	High cropland per capita associated with a low poverty rate
	RPCPA[5]	$\frac{\text{Potential cropland area in a p}}{\text{Protected area}}$	WDPA[6] & GAEZ	High RPCPA associated with a high poverty rate
Agricultural Inputs	Irrigation percentage	$\frac{\text{Actual irrigation area}}{\text{Potential irrigation area}}$	FAO	High irrigation percentage associated with a low poverty rate
	Fertilizer consumption	$\frac{\text{Total fertilizer consumption}}{\text{Total cropland area}}$	World Bank database	High fertilizer consumption associated with low poverty rate

[1] RAPY: ratio of actual to potential yield; [2] GAEZ: global agro-ecological zones; [3] FAO: Food and Agriculture Organization; [4] RAPC: ratio of actual to potential cropland; [5] RPCPA: ratio of potential cropland in protected areas; [6] WDPA: World Database on Protected Areas.

CART is prone to overfitting owing to complex rules. Therefore, to avoid this problem, an ensemble approach to CART, called a random forest, was used to avoid overfitting. Random forests use a process of randomly selecting variables and samples for model training (Breiman, 2001). During model training, the samples are selected with a bootstrap method. In each iteration of the model, two-thirds of the data are used in model training, while the remaining one-third are withheld and used in model testing (known as out-of-bag or OOB testing). This process is repeated hundreds or thousands of times, and different permutations of variables are then used as input (Breiman, 2001). In repeating the procedure, the variable importance can be estimated according to the change in accuracy of the model when a particular variable is left out. When a variable is absent in model training, accuracy is reduced based on the importance of the variable. The result from random forest considers the vote from each tree's prediction using a boosting algorithm (Watmough et al., 2019).

We conducted CART and random forest in *R 3.5.1* using the "*tree*" (Breiman, 2017) and "*random forest*" (Breiman, 2001) packages, respectively. In total, 104 samples were used to train the models after excluding the records with missing data. For CART, a 10-fold cross-validation approach was used to optimize the cost-complexity parameter with the *cv.tree* function, and the tree was pruned with the *prune.misclass* function. Finally, we employed the *draw.tree* function in the *maptree* package (https://www.rdocumentation.org/packages/maptree) to visualize the pruned tree. To further explore the relationship between the poverty rate and arable land use indicators, a non-parametric random forest regression was conducted using the "*random forest*" package. In random forest, the number of trees (*ntree*) was set to 1,000 because the accuracy converges when the number of trees is larger than 400 (Fig. 6b), and the increase in this parameter ensures the robustness of the variable importance metrics (Genuer et al., 2008).

To better understand the indicators and samples, the relationship between variables is explored with the Kendall tau rank correlation coefficient (r_k) and Spearman's rank correlation coefficient ($\mathrm{r_s}$). Spearman's rank correlation coefficient measures how well the relationship between two variables can be described by a monotonic function (Myers et al., 2013), while the Kendall tau rank correlation coefficient measures the portion of ranks that match between two datasets (Kendall, 1938). Both Spearman's rank and Kendall tau are non-parametric hypothesis tests for rank correlation or statistical dependence, which is appropriate for non-normality data in this research. Although the Pearson coefficient measures the linear correlation between two variables (parametric method) and is suitable for only the variables following a normal distribution, we also calculated it to show the difference in correlation detected by parametric and non-parametric methods.

3. Results

3.1 Agricultural intensification and poverty

The distribution of RAPY is shown in Fig. 3a, which depict the high RAPY in North America, East Asia, Australia and South America and low values in most of Africa and Eastern Europe. Globally, RAPY averages 0.45 with a median value of 0.41 (Fig. 4a). Ireland (1.11), the Netherlands (1.06), Finland (1.03), Jamaica (1.02) and Australia (1.01) had the highest RAPY, while Sudan (0.08), Botswana (0.08), Turkmenistan (0.13), Burkina Faso (0.14) and Uganda (0.15) had the lowest RAPY values.

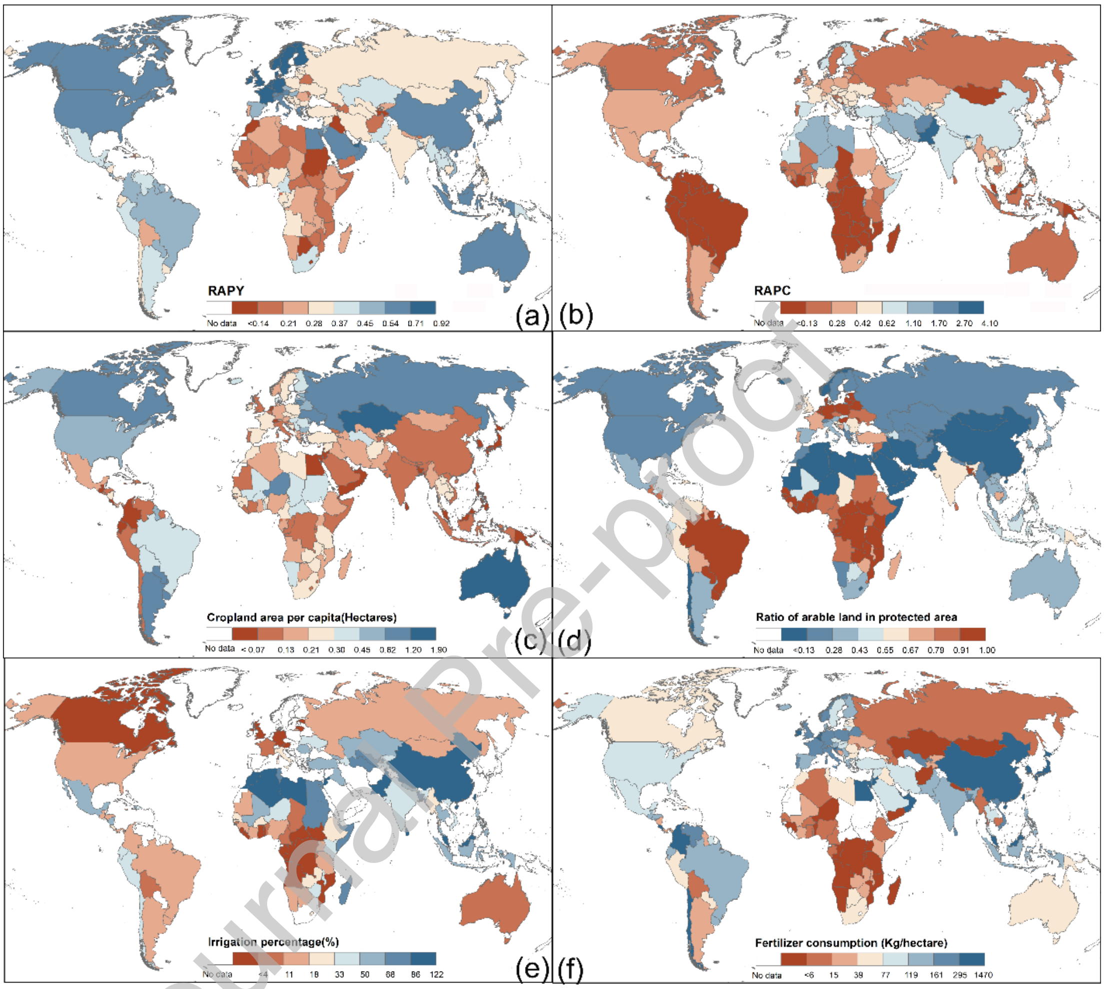


Fig. 3. Maps of: (a) the ratio of actual to potential yield; (b) the ratio of cropland to potential cropland area; (c) cropland area per capita; (d) the ratio of potential cropland in protected areas; (e) irrigation percentage; and (f) fertilizer consumption

Fig. 4 provides statistics for each poverty level, indicating that RAPY was low in developing countries and high in developed countries. For developed countries, the mean RAPY was 0.46 and 0.22 in developing countries. For developed, somewhat poor, moderate poor to developing countries, the RAPY is decreased monotonically, which generally support the hypothesis in Table 2 that high RAPY is associated with a low poverty rate. This link is confirmed by the relationships between poverty rate and

RAPY shown in Table 3 indicating that RAPY was significant and negatively correlated with poverty with r_p of −0.45, r_k of −0.37 and r_s of −0.54.

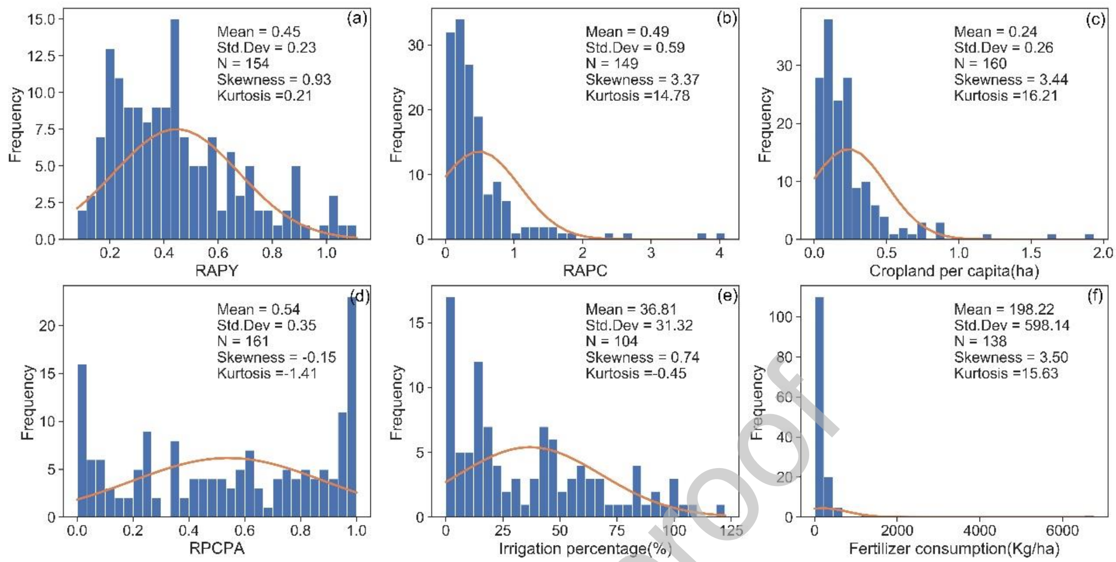


Fig. 4. Statistics for (a) RAPY (ratio of actual to potential yield); (b) RAPC (ratio of actual to potential cropland); (c) Cropland per capita; (d) RPCPA (potential cropland to protected areas); (e) Irrigation percentage; (f) Fertilizer consumption
Note: Mean value, standard deviation (Std. Dev), sample number (*N*), skewness and kurtosis are labelled. The line is the best fitted normal distribution.

3.2 Agricultural extensification and poverty

RAPC was comparatively high in northern Africa and central, eastern and southern Asia (Fig. 3b). The global mean value of RAPC was 0.49, and the median value was 0.33 (Fig. 4b). Pakistan (4.08), Bhutan (3.69), Cyprus (2.66), Afghanistan (2.42), and Comoros (1.90) had the highest RAPC, while Colombia (0.02), Gabon (0.02), the Democratic Republic of Congo (0.02), Papua New Guinea (0.01) and Suriname (0.01) had the lowest RAPC. The ratio of cropland to potential cropland area was 0.43 in developing countries and 0.59 in somewhat developing countries, which was the highest. RAPC was also negatively related to poverty, with a significant r_k of −0.19 and r_s of −0.26 (Table 3). Generally, the link between poverty and RAPC does not conform to the hypothesis in Table 2 with high RAPC not necessarily resulting in low poverty rates.

The mean cropland per capita at the global level was 0.24 ha, and the median value was 0.17 ha (Fig. 4c). Australia (1.93 ha), Kazakhstan (1.68 ha), Canada (1.22 ha), Argentina (0.90 ha) and the Russian Federation (0.85 ha) had the most cropland per capita (Fig. 3c), while Djibouti, the United Arab Emirates, Qatar, Oman and the Maldives had the least cropland per capita (less than 0.10 ha). The cropland per capita was highest in developed countries (0.29 ha/capita), followed by developing countries (0.22 ha/capita), moderately poor countries (0.18 ha/capita) and somewhat poor countries (0.12 ha/capita). So, its link with poverty is very weak, as depicted in Table 3 and did not confirm our hypothesis in Table 2.

The proportion of potential cropland in protected areas was high in central Africa, South America and most of Europe (Fig. 3d). The mean global value was 0.54 and the median was 0.58 (Fig. 4d). Luxembourg, Gambia, Benin, Moldova and Belarus had all protected areas identified as potential cropland, while there was no potential cropland in protected areas of Djibouti, United Arab Emirates, Qatar, Oman, the Maldives, Republic of Yemen, Saudi Arabia and Libya.

The mean proportion of cropland in protected areas of developing countries was 0.73 indicating that 73% of protected areas could be used as cropland, while the value was 0.42 in somewhat poor countries. This proportion increased to 0.49 in developed countries. The relationship between poverty and RPCPA seems to follow the Kuznets curve, which hypothesizes that environmental degradation first increases and then decreases with economic development as an inverted U-shape (Stern, 2004; Özokcu and Özdemir, 2017). Overall, the RPCPA was positively related to poverty, with r_k of 0.19 and r_s of 0.14, as well as r_p of 0.29.

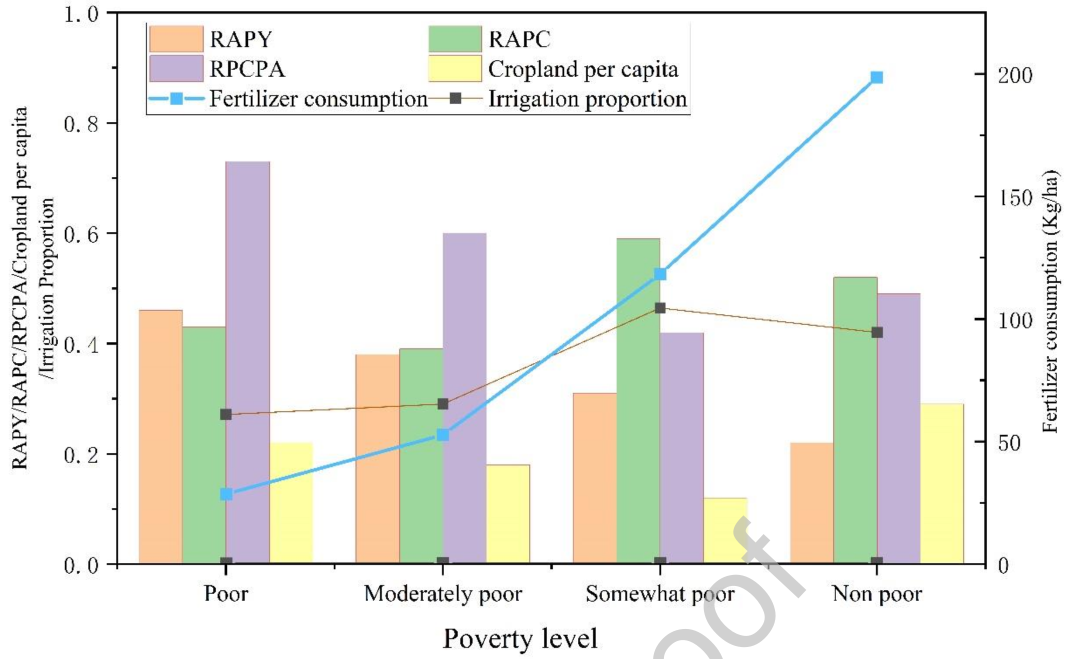


Fig. 5. Mean value of independent variables at each poverty level. The unit of fertilizer consumption was kg/ha.
Note: RAPY is the ratio of actual to potential yield; RAPC is the ratio of actual to potential cropland; and RPCPA is the ratio of potential cropland in protected areas.

3.3 Agricultural input and poverty

As depicted in Fig. 3e, the irrigation percentage was relatively high in northern Africa and eastern and central Asia. Irrigation is already utilized in all potential irrigation areas in Jordan (122%), Malaysia (107.3%), Sri Lanka (105.3%), Libya (100.0%) and China (99.8%). The values above 100% indicate that the irrigated area was larger than the potential irrigation area. Notably, the irrigation percentage was estimated from the ratio of irrigation area to potential irrigation area, which was estimated from the FAO. Thus the relative value of irrigation percentage is more meaningful than the absolute value. The irrigation percentages in the Central African Republic, the Democratic Republic of Congo, Estonia, Liberia and Gabon were less than 1%.

The proportion of irrigation was highest in somewhat poor countries (46.5%) and lowest in poor countries (27.1%), as shown in Fig. 5. Irrigation percentage was significantly related to poverty assuming a linear relationship, with a significant r_p of

−0.20 (Table 3).

Globally, mean fertilizer use per hectare was 140 kg/ha (World Bank), with a median of 112.05 kg/ha (Fig. 4f). As shown in Fig. 3F, there were several high values in China and the Republic of Korea and low values in Africa and the Russian Federation. In Qatar, fertilizer use per hectare was 6,755 kg in 2016, the highest followed by Malaysia (1,723 kg), Ireland (1,247 kg), Colombia (659 kg), Egypt and the Arab Republic of Syria (649 kg). On the other hand, Niger (0.4 kg), Gambia (1.2 kg), Guinea (1.6 kg), the Democratic Republic of Congo (1.8 kg) and Uganda (1.9 kg) used the least fertilizer.

The fertilizer use in developed countries was 198.6 followed by somewhat poor countries, moderately poor counties and poor countries, and only 28.59 kg/ha in poor countries (Fig. 5). Notably, fertilizer consumption was negatively related to the poverty rate, with a significant r_k of −0.38 and r_s of −0.53 (Table 3), indicating that increased fertilizer consumption per hectare usually coincided with a low poverty rate, as we assumed in Table 2.

Table 3. Pearson (R_p), Spearman (R_s) and Kendall tau (R_k) correlation coefficients between poverty and the ratio of actual to potential yield (RAPY), ratio of cropland to potential cropland area (RAPC), ratio of potential cropland in protected areas (RPCPA), cropland area per capita, (e) irrigation percentage, and fertilizer consumption.

Correlation coefficient	RAPY	PAPC	RPCPA	Cropland per capita	Irrigation percentage	Fertilization consumption
R_p	−0.45**	−0.08	0.29**	−0.11	−0.2*	−0.16
R_s	−0.54**	−0.26**	0.19*	−0.07	−0.18	−0.53**
R_k	−0.37**	−0.19**	0.14**	−0.05	−0.13	−0.38**

Note: * denotes those two variables were significantly correlated (p<0.05), and ** indicates that two variables were significantly correlated (p<0.01)

3.4. Arable land use and poverty

According to skewness and kurtosis values in Fig. 4, none of the six indicators follow a normal distribution. So, we selected non-parametric CART and random forest models to reveal specific linkage rules and variables for arable land use indicators. For the decision tree, the terminal number (tree size) was the key parameter. With an increase in tree size accuracy could become very high; however, the model would become complicated leading to overfitting. Therefore, we kept the tree as simple as possible to avoid overfitting (James et al., 2013). The relationship between tree size (terminal number) and training accuracy is shown in Fig. 6a. To avoid overfitting, we chose a tree size of 7 because training accuracy converges with increasing tree size. In the end, the accuracy of the pruned decision tree was 73.8%.

The pruned classification tree (Fig. 6c), in which the root node was RAPY, indicated that it was most important for poverty prediction in the CART model. Overall, there were seven rules in the decision tree. A poor country typically had RAPY values less than 0.33 and fertilizer consumption less than 7.31 kg/ha. For a country with RAPY values less than 0.33 and fertilizer consumption greater than 7.31 kg/ha and if the RAPC was less than 0.18, the country was usually a moderately poor country; otherwise, the country was likely to be in the developed country class. Overall, the poor and moderately poor countries always appeared on the left side of the root node, which means the RAPY in these countries was less than 0.33.

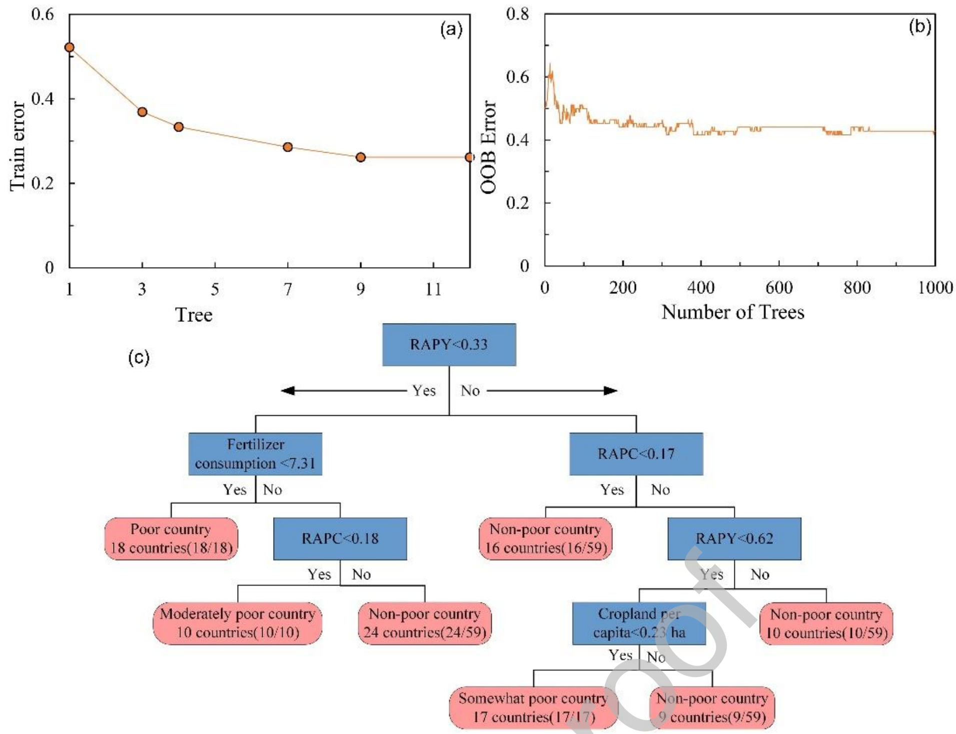
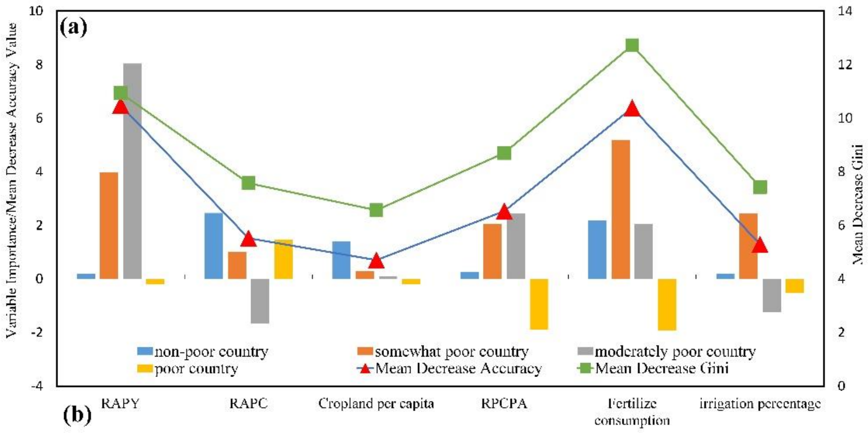


Fig. 6. (a) the relationship between tree size (terminal number) and training accuracy for the decision trees, (b) the relationship between number of trees and OOB accuracy in random forest and (c) the pruned decision tree from CART.

Note: RAPC: ratio of actual to potential cropland; RAPY: ratio of actual to potential yield

The non-parametric random forest model had an overall OOB estimation error of 44.0% (an effective accuracy of 55.9%). According to the decrease in accuracy and the Gini index when an indicator variable was removed, fertilizer use was the most important variable in predicting poverty classes, followed by RAPY, RPCPA, RAPC and irrigation percentage (Fig. 7a), while cropland per capita was least important. Nevertheless, RAPY contributed slightly more than fertilizer use.

The rank of variable importance for each poverty level is shown in Fig. 7b. RAPC was the most important variable in predicting developed and poor categories. RAPY was most important in predicting moderately poor countries, while fertilizer use was most important for the prediction of developing countries. The most important variable in developing countries was the RAPC.



(b)

non-poor country	somewhat poor country	moderately poor country	poor country
RAPC	Fertilizer consumption	RAPY	RAPC
Fertilizer consumption	RAPY	RPCPA	Cropland per capita
Cropland per capita	Irrigation percentage	Fertilizer consumption	RAPY
RPCPA	RPCPA	Cropland per capita	Irrigation percentage
Irrigation percentage	RAPC	Irrigation percentage	RPCPA
RAPY	Cropland per capita	RAPC	Fertilizer consumption

Fig. 7. (a) is the variable importance of six indicators derived from random forest. Coloured bars show that variable importance differed depending on the poverty class of the country. The lines show two different methods for calculating variable importance. (b) is the variable importance for each poverty level according to the mean decrease in the OOB accuracy in random forest, ranking from most to least important.

Note: RAPC is the ratio of actual to potential cropland; RAPY is the ratio of actual to potential yield and RPCPA is the ratio of potential cropland in protected areas.

4. Discussion

Our paper explores the linkage between arable land use and poverty at global perspective using non-parametric machine learning methods of CART and random forest. We find there are strong linkages between poverty and crop yield and cropland use. The crop yield gap has the highest relevance to poverty, while RAPC had a non-linear relationship with poverty. In addition, the RPCPA was positively related to poverty. Fertilizer use was significant and negatively related to the poverty rate (r_s=−0.53, r_k=−0.38). Fertilizer use was the most important variable in poverty class prediction, followed by RAPY, RPCPA, RAPC and irrigation percentage according to the variable importance in random forest. In the pruned classification tree, the root node is RAPY, indicating that this variable was the most important in poverty class prediction in the CART model. Overall, arable land use is highly linked with poverty at a global level.

4.1 Big geodata for supporting arable land assessment

Data from questionnaires, common in previous research, only reflects the relationship between poverty and agricultural development at the individual level due to small sample sizes (Minten and Barrett, 2008; Namara et al., 2010; Kassie et al., 2011; Watmough et al., 2016; Leonardo et al., 2018). This limitation introduces bias, particularly when respondents know how the data were going to be used. Census data was used in some studies, but these data are often unavailable in developing countries due to costs, leading bias as well. So the big-geo data was used in our research to avoid these problems.

Big geodata were used to assess the arable land use and explore its links with poverty. Specifically,, the following big geodata were in this research:

1). The geo-statistic data from World Bank and FAO including cropland per capita and fertilization use.

2). Crop simulation data from GAZE and FAO. A simulated model of crop productivity and cropland suitability by global agro-ecological zones (GAEZ) (IIASA and FAO, 2012) was used in grid format to reflect the spatial variation across regions (Watmough et al., 2016; Watmough et al., 2019). Using the potential yield from the GAEZ model and actual yield from FAO, it is feasible to estimate the yield gap in gridded format. The utilization of arable land can be estimated based on the GAZE model's distribution of cropland and potential cropland.

3) Global protected area (PA) data were estimated using the World Database on Protected Areas (United Nations Environment World Conservation Monitoring Centre), which supported the analysis related to protected area and protection of biodiversity.

4) Potential irrigation area simulation data from the FAO was used to estimate a novel "irrigation percentage", which is the ratio of actual and potential irrigation area. This methods was used because irrigation percentage does not reflect the irrigation development level for one region, because not all cropland is water scare. These data

support the decision making in achieving SDGs, especially for SDG1 and SDG2.

We have included human alterations and excluded differences in natural conditions by comparing actual utilization and its potential. In other words, we have only considered anthropological factors. For example, climatic conditions can affect potential and actual yields. Instead of using absolute values of potential and actual yields, we used the ratio of actual to potential yields to estimate the level of arable land use, which excludes climatic differences, but includes the effects of technology for agricultural development. Another example is irrigation percentage. In our research. the irrigation percentage was estimated by the ratio of actual irrigated area to potential irrigated area to negate the effect of differences in natural endowment. However, we recognize that the actual irrigation area is influenced by human inputs and technology development.

To further explore the relationship between poverty and arable land use, we identified 12 countries where the poverty rate decreased by more than 10% since 1990 and plotted poverty rates and cereal yields as shown in Fig. 8. The historical poverty rates and cereal yields were obtained from the World Bank. We concluded that the decrease in the poverty rate was accompanied by an increase in crop yields.

Fig. 8. Trends in cereal yields and poverty rates for twelve countries where the poverty rate decreased by more than 10% since 1990.

4.2 Non-parametric machine learning methods

The relationship between poverty and arable land use is complex. It is unclear whether poverty causes less arable land use or inefficient arable land use results in more poverty. A two-way relationship between poverty and arable land use is possible. Our original hypothesis does not assume that poverty is dependent on arable land, but that a links exists between arable land use and poverty. A data-driven approach was used to analyze these links using data mining methods. So, the link differs from previous causal relationships and is more of an association that already exists. Watmough et al. (2016; 2019) used this method to explore the poverty-environment relationship in Assam India as well as predict rural household poverty with remote sensed data. Tian et al. (2020) used data mining to understand the links between poverty and water resource development and found that the ratio of water utilization to reservoir density is highly related to poverty from a global perspective. While this research did not begin with a hypothesis some interesting patterns were discovered.

As opposed to previous research that used linear models (Huang et al., 2006; Minten and Barrett, 2008; Kassie et al., 2011; Burney and Naylor, 2012), we found that non-linear relationships were dominant among poverty and arable land use indicators. This finding reflects the complex relationship between poverty and agricultural development, such as the non-linear relationship between cropland area and crop income found by Harris and Orr (2014), which may not be depicted clearly by a linear model. The poverty-arable land use relationship was detected by parametric (Pearson) and non-parametric methods (Spearman and Kendall tau) as a comparison. Parametric methods did not show a significant correlation between poverty-RAPC and the poverty-fertilizer-consumption relationship, while some significant non-linear relationships were detected using non-parametric methods (Table 3).

Previous research supports the hypothesis that agricultural development can alleviate poverty (Huang et al., 2006; Minten and Barrett, 2008; Hanjra et al., 2009; Leonardo et al., 2018). RAPY is a direct measurement of cropland utilization efficiency and reflects crop intensification. We argue that a high RAPY and low poverty rate are significantly linked, which supports previous research findings (Davis et al., 2012; Leonardo et al., 2018). As shown by the Green Revolution, increased crop productivity reduced poverty rates, increased GDP per capita (Thirtle et al., 2003; Pingali, 2012) and decreased labour share in agriculture (Ravallion and Datt, 1996; McArthur and McCord, 2017), thereby transformation agriculture (Johnston, 1962;

Mellor, 2017; Nin-Pratt et al., 2018). Our research has found a link between poverty and arable land use and delivers some implications of this relationship.

4.3 Implications from links between poverty and arable land use

According to the random forest, feature importance varies with poverty levels and country. Globally, RAPY is the most important variable in predicting poverty, which may cloud the significance of increased crop yields and income. Some studies have found that current cropland production will substantially exceed food demand in 2050 through improved farm management and enhanced crop efficiency without any increases in croplands (Mauser et al., 2015). The importance of RAPY changed depending on the country's poverty classification. While RAPY was not important in developed countries, it was important for somewhat poor, moderately poor and developing countries.

For each poverty level, the links with arable land use were different. RAPC was the most important variable for developing countries, indicating that crop extensification was highly linked with poverty and should be given more attention, which coincides with the findings of Benfica et al. (2019) in Mozambique. In Mozambique, where RAPC was only 8%, an agricultural investment project (Programa Nacional de Investimento do Sector Agrário) was proposed to increase the share of irrigation and fertilizer subsidies, however it failed to reduce poverty (Benfica et al., 2019).

IAPY was the most important factor in predicting poverty in moderately poor communities Thus enhancing crop productivity should be given more attention considering the RAPY value of 0.31 in moderately poor countries. The RPCPA (considered a secondary factor) value of 0.60 was also higher than that in somewhat poor and developed countries. In sub-Saharan Africa, potential cropland was under-utilized, especially in protected areas, as shown in Fig. 3. This ensures high biodiversity, which benefits both local and global populations. However, potential croplands in protected areas results in reduced agricultural development in these areas. We agree with Brandon et al. (2005) that biodiversity conservation and agricultural development should be reconciled by creating core protected areas in low population areas where biodiversity conservation is a priority, while cropland expansion should be allowed in high population density areas to meet food demand.

4.4 Limitations and outlook

In this research, we only used cross-sectional data to analyze the relationship between poverty rates and arable land use due to the lack of historical poverty rate and arable land data for all countries. To ensure data conformity, most data were obtained from reputable sources like the World Bank or FAO. We tried to fill the gaps in poverty rate data for 2016, but some data were not available. Although estimating poverty rates using data from recent years will introduce some bias when determining the relationship between poverty and arable land use, the influence will be acceptable overall after the poverty rate is divided into four classes. For RAPY, RPCPA and RAPC, the value was estimated by the GAEZ model and will not change significantly over a decade.

The linkage found by machine learning methods (random forest) was not strong, and we were unable to fully explain its mechanisms. Limited accuracy means that the linkage is sub-optimal for individual countries, but can show global patterns. Many other factors are related to poverty from the macroscopic perspective including: industrial development (Kimura and Chang, 2017); services (Joshi, 2004); natural resources (Barbier, 2007); access to land (Besley and Burgess, 2000); and political and socioeconomic conditions (Chaux et al., 2009). These impacts are not directly considered in this research, however prolonged war and political instability may lead to a decrease in arable land and production, which would be reflected in RAPY and RAPC.

According to the CART method, arable land use reflects the poverty level in 73.8% of the countries in the world. This means that in 26.2% of countries arable land use cannot explain poverty levels with other factors having greater importance. The poverty rates for some countries have been affected by war such as in the Syrian Arab Republic (82.5%) and Somalia (73.0%). With abundant natural resources, some countries benefit and have low poverty rates, such as Oman and Qatar (0%). On the other hand, the development of arable land use reflects the capacity for industrialization as it can ensure the provision of fertilizer and agricultural mechanization (Osakwe, 2019).

We did not find significant links between poverty and irrigation on a national scale. This is partly due to the fact that irrigated agriculture is not globally predominant, even in developed countries (Harris and Orr, 2014). Therefore, irrigation did not show a significant relationship between poverty rate and irrigation percentage based on our methods. However this does not suggest that irrigation is unhelpful for enhancing crop productivity and mitigating the influence of drought, as some studies have found that irrigation agriculture is helpful in specific countries or regions such as China (Huang et al., 2006) and sub-Saharan Africa (Burney and Naylor, 2012).

Data mining methods aim to find patterns hidden in data. For example, the linkage between RAPY and poverty in which a high RAPY leads to a low poverty rate or a low RAPY results in a high poverty rate should be carefully interpreted because the mechanisms between poverty and arable land use are not clear using non-parametric data mining methods. Hence, more research needs to be done to explore these mechanisms. This will be possible when big geodata becomes more comprehensive and includes indicators such as cropland abandonment, crop diversification and crop rotation.

5. Conclusions

There are strong linkages between poverty and crop yield and cropland utilization. The crop yield gap is most relevant to poverty, while RAPC has a non-linear relationship with poverty. In addition, the RPCPA was positively related to poverty. Fertilizer consumption was negatively related to the poverty rate (r_s=−0.53, r_k=−0.38). Arable land use was able to predict poverty levels with an accuracy of 73.8%. Fertilizer use was the most important variable when predicting poverty levels, followed by RAPY, RPCPA and RAPC. RAPY was most important in predicting poverty levels in moderately poor countries, while RAPC was the most variable in developing countries. So, crop productivity enhancement should be given more attention in relatively poor countries. In developing countries, cropland expansion is also one of the potential ways to alleviate poverty besides increasing RAPY. Globally, arable land use, represented by agricultural inputs, crop intensification and crop extensification, is highly linked with poverty because zero hunger is a prerequisite of poverty eradication.

Author Contributions

Fuyou Tian contributed to the research experiments and manuscript preparation. Bingfang Wu contributed to conceptual designing this research and was responsible for the research. Hongwei Zeng, Miao Zhang, Gary R Watmough and Yurui Li gave useful comments which improved the paper. All of the co-authors helped revise and polish the manuscript.

Declaration of Competing Interest

None

Acknowledgements

We thank the World Bank, FAO, IIASA for providing data for this research. This study was supported financially by the National Key Research and Development Program (2016YFA0600304).

References

Adela, F.A., Aurbacher, J., Abebe, G.K., 2019. Small-scale irrigation scheme governance - poverty nexus: Evidence from Ethiopia. Food Secur. 11(4), 897-913.

Alene, A.D., Coulibaly, O., 2009. The impact of agricultural research on productivity and poverty in sub-Saharan Africa. Food Policy 34(2), 198-209.

United Nations, 2015. Transforming our world: The 2030 Agenda for sustainable development. United Nations General Assembly, New York.

Barbier, E.B., 2007. Natural Resources and Economic Development. Cambridge University Press, Cambridge.

Barbier, E.B., Hochard, J.P., 2018. Land degradation and poverty. Nat. Sustain. 1(11), 623-631.

Barrow, C.J., 2012. Biochar: Potential for countering land degradation and for improving agriculture. Appl. Geogr. 34, 21-28.

Benfica, R., Cunguara, B., Thurlow, J., 2019. Linking agricultural investments to growth and poverty: An economywide approach applied to Mozambique. Agric. Syst. 172, 91-100.

Besley, T., Burgess, R., 2000. Land reform, poverty reduction, and growth: Evidence from India. Q. J. Econ. 115(2), 389-430.

Brandon, K., Gorenflo, L.J., Rodrigues, A.S., Waller, R.W., 2005. Reconciling biodiversity conservation, people, protected areas, and agricultural suitability in Mexico. World Dev. 33(9), 1403-1418.

Breiman, L., 2001. Random forests. Mach. Learn. 45(1), 5-32.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and Regression Trees. Routledge, Boca Raton.

Burney, J.A., Davis, S.J., Lobell, D.B., 2010. Greenhouse gas mitigation by agricultural intensification. Proc. Natl. Acad. Sci. U.S.A. 107(26), 12052-12057.

Burney, J.A., Naylor, R.L., 2012. Smallholder irrigation as a poverty alleviation tool in Sub-Saharan Africa. World Dev. 40(1), 110-123.

Chaux, E., Molano, A., Podlesky, P., 2009. Socio-economic, socio-political and socio-emotional variables explaining school bullying: A country-wide multilevel analysis. Aggress. Behav. 35(6), 520-529.

Chen, J., Yang, S., Li, H., Zhang, B., Lv, J., 2013. Research on geographical environment unit division based on the method of natural breaks (Jenks). Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 40(4W3),47-50.

CIA, 2016. The world Factbook. https://www.cia.gov/library/publications/the-world-factbook/fields/2046.html. (accessed 20 June 2018).

Dasgupta, S., Deichmann, U., Meisner, C., Wheeler, D., 2005. Where is the poverty–environment nexus? Evidence from Cambodia, Lao PDR, and Vietnam. World Dev. 33(4), 617-638.

Davis, K., Nkonya, E., Kato, E., Mekonnen, D.A., Odendo, M., Miiro, R., Nkuba, J., 2012. Impact of farmer field schools on agricultural productivity and poverty in East Africa. World Dev. 40(2), 402-413.

Dawson, N., Martin, A., Sikor, T., 2016. Green Revolution in Sub-Saharan Africa:

Implications of imposed innovation for the wellbeing of rural smallholders. World Dev. 78, 204-218.

de Sherbinin, A., 2008. Is poverty more acute near parks? An assessment of infant mortality rates around protected areas in developing countries. Oryx 42(1), 26-35.

Diao, X., Hazell, P., Thurlow, J., 2010. The role of agriculture in African development. World Dev. 38(10), 1375-1383.

Duan, W., Wen, Y., 2017. Impacts of protected areas on local livelihoods: Evidence of giant panda biosphere reserves in Sichuan Province, China. Land Use Policy 68, 168-178.

Dzanku, F.M., Jirström, M., Marstorp, H., 2015. Yield gap-based poverty gaps in rural Sub-Saharan Africa. World Dev. 67, 336-362.

Ellis, F., 2005. Small farms, livelihood diversification, and rural-urban transitions: Strategic issues in Sub-Saharan Africa. In: International Food Policy Research Institute (IFPRI). The Future of Small Farms: Proceedings of a Research Workshop. Wye, London, UK, 26–29 June 2005.

Evenson, R.E., Gollin, D., 2003. Assessing the impact of the Green Revolution, 1960 to 2000. Science 300(5620), 758-762.

Genuer, R., Poggi, J.M., Tuleau, C., 2008. Random Forests: Some methodological insights. arXiv:0811.3619v1[stat.ML].

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques. Elsevier Inc.

Hanjra, M.A., Ferede, T., Gutta, D.G., 2009. Reducing poverty in sub-Saharan Africa through investments in water and other priorities. Agric. Water Manag. 96(7), 1062-1070.

Harris, D., Orr, A., 2014. Is rainfed agriculture really a pathway from poverty? Agric. Syst. 123, 84-96.

Huang, Q., Rozelle, S., Lohmar, B., Huang, J., Wang, J., 2006. Irrigation, agricultural performance and poverty reduction in China. Food Policy 31(1), 30-52.

IIASA, FAO, 2012. Global Agro-ecological Zones (GAEZ v3.0). IIASA, Laxenburg, Austria and FAO, Rome, Italy.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: With Applications in R. Springer, New York.

Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science 353(6301), 790-794.

Jiang, L., Chen, X., Meng, L., Zhang, G., Pan, Z., An, P., 2021. Increased grain production of cultivated land by closing the existing cropping intensity gap in Southern China. Food Secur. 13(2), 385-398.

Johnston, B.F., 1962. Agricultural development and economic transformation: A comparative study of the Japanese experience. Food Res. Inst. Stud. 3, 223-276.

Joshi, S., 2004. Tertiary sector-driven growth in India: Impact on employment and poverty. Econ. Polit. Wkly. 39, 11-17.

Kassie, M., Shiferaw, B., Muricho, G., 2011. Agricultural technology, crop income, and poverty alleviation in Uganda. World Dev. 39(10), 1784-1795.

Kendall, M.G., 1938. A new measure of rank correlation. Biome 30(1/2), 81-93.

Kimura, F., Chang, M.S., 2017. Industrialization and poverty reduction in East Asia: Internal labor movements matter. J. Asian Econ. 48, 23-37.

Leonardo, W., van de Ven, G.W.J., Kanellopoulos, A., Giller, K.E., 2018. Can farming provide a way out of poverty for smallholder farmers in central Mozambique? Agric. Syst. 165, 240-251.

Liverpool, L.S.O., Winter-Nelson, A., 2010. Poverty Status and the Impact of Formal Credit on Technology Use and Wellbeing among Ethiopian Smallholders. World Dev. 38(4), 541-554.

Mauser, W., Klepper, G., Zabel, F., Delzeit, R., Hank, T., Putzenlechner, B., Calzadilla, A., 2015. Global biomass production potentials exceed expected future demand without the need for cropland expansion. Nat. Commun. 6, 8946.

Maxwell, S., Slater, R., 2004. Food Policy Old and New. Blackwell, Oxford.

McArthur, J.W., McCord, G.C., 2017. Fertilizing growth: Agricultural inputs and their effects in economic development. J. Dev. Econ. 127, 133-152.

Mellor, J.W., 2017. Agricultural Development and Economic Transformation: Promoting Growth with Poverty Reduction. Palgrave Macmillan, New York.

Minten, B., Barrett, C.B., 2008. Agricultural technology, productivity, and poverty in Madagascar. World Dev. 36(5), 797-822.

Müller, D., Leitão, P.J., Sikor, T., 2013. Comparing the determinants of cropland abandonment in Albania and Romania using boosted regression trees. Agric. Syst. 117, 66-77.

Myers, J.L., Well, A.D., Lorch Jr, R.F., 2013. Research Design and Statistical Analysis. Routledge, New York.

Namara, R.E., Hanjra, M.A., Castillo, G.E., Ravnborg, H.M., Smith, L., Van Koppen, B., 2010. Agricultural water management and poverty linkages. Agric. Water Manag. 97(4), 520-527.

Nin-Pratt, A., El-Enbaby, H., Figueroa, J.L., ElDidi, H., Breisinger, C., 2018. Agriculture and economic transformation in the Middle East and North Africa: A review of the past with lessons for the future. International Food Policy Research Institute (IFPRI), Washington D.C.

Osakwe, P.N., 2019. Unlocking the potential of the power sector for industrialization and poverty alleviation in Nigeria. In: Wamboye, E.F., Nyaronga, P.J. (Eds.), The Service Sector and Economic Development in Africa, Routledge, pp.159-180.

Özokcu, S., Özdemir, Ö., 2017. Economic growth, energy, and environmental Kuznets curve. Renew Sust. Energ. Rev. 72, 639-647.

Pingali, P.L., 2012. Green revolution: Impacts, limits, and the path ahead. Proc. Natl. Acad. Sci. U.S.A. 109(31), 12302-12308.

Ravallion, M., Datt, G., 1996. How important to India's poor is the sectoral composition of economic growth? World Bank Econ. Rev. 10(1), 1-25.

Ricker-Gilbert, J., Jayne, T.S., 2012. Do Fertilizer Subsidies Boost Staple Crop Production and Reduce Poverty Across the Distribution of Smallholders in Africa?

Quantile Regression Results from Malawi. International Association of Agricultural Economists, Brazil.

- Steele, J.E., Sundsoy, P.R., Pezzulo, C., Alegana, V.A., Bird, T.J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.-A., Iqbal, A.M., Hadiuzzaman, K.N., Lu, X., Wetter, E., Tatem, A.J., Bengtsson, L., 2017. Mapping poverty using mobile phone and satellite data. J. R. Soc. Interface 14(127), 20160690.

Stern, D.I., 2004. The rise and fall of the environmental Kuznets curve. World Dev. 32(8), 1419-1439.

Sumberg, J., 2012. Mind the (yield) gap(s). Food Secur. 4(4), 509-518.

Thirtle, C., Lin, L., Piesse, J., 2003. The impact of research-led agricultural productivity growth on poverty reduction in Africa, Asia and Latin America. World Dev. 31(12), 1959-1975.

Tian, F., Wu, B., Zeng, H., Ahmed, S., Yan, N., White, I., Zhang, M., Stein, A., 2020. Identifying the links among poverty, hydroenergy and water use using data mining methods. Water Resour. Manag. 34(5), 1725-1741.

Tian, X., Yu, X., 2019. Crop yield gap and yield convergence in African countries. Food Secur. 11(6), 1305-1319.

Tomich, T.P., Lidder, P., Coley, M., Gollin, D., Meinzen-Dick, R., Webb, P., Carberry, P., 2019. Food and agricultural innovation pathways for prosperity. Agric. Syst. 172, 1-15.

United Nations Environment World Conservation Monitoring Centre, 2019. Protected Planet. https://www.protectedplanet.net/ (accessed 7 August 2019).

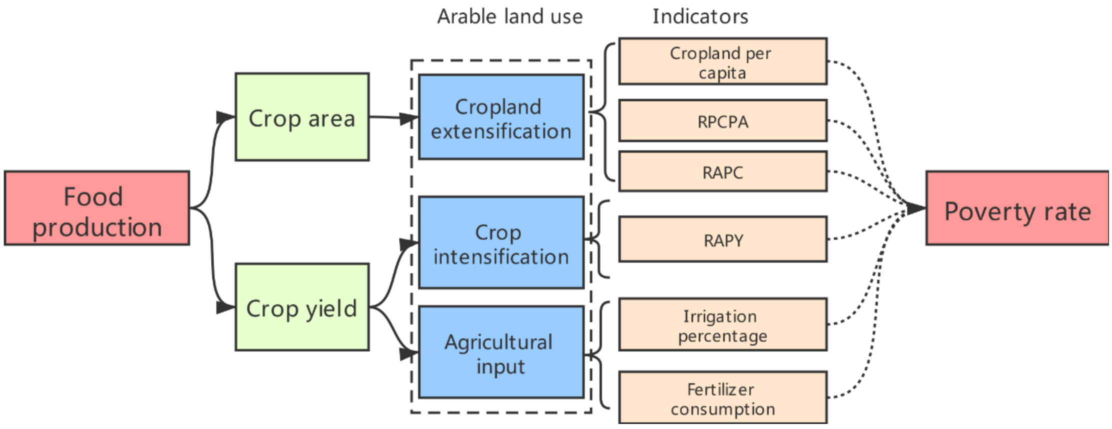
Watmough, G.R., Atkinson, P.M., Saikia, A., Hutton, C.W., 2016. Understanding the evidence base for poverty–environment relationships using remotely sensed satellite data: an example from Assam, India. World Dev. 78, 188-203.

Watmough, G.R., Marcinko, C.L.J., Sullivan, C., Tschirhart, K., Mutuo, P.K., Palm, C.A., Svenning, J.C., 2019. Socioecologically informed use of remote sensing data to predict rural household poverty. Proc. Natl. Acad. Sci. U.S.A. 116(4), 1213-1218.

World Bank, 2019. World Bank Open Data. https://data.worldbank.org/. (accessed 23 January 2019).

Wu, W., Yu, Q., You, L., Chen, K., Tang, H., Liu, J., 2018. Global cropping intensity gaps: Increasing food production without cropland expansion. Land Use Policy 76, 515-525.

Graphical Abstract: Graphical Abstract


Arable land use
Indicators
Food production
Crop area
Crop yield
Cropland extensification
Crop intensification
Agricultural input
Cropland per capita
RPCPA
RAPC
RAPY
Irrigation percentage
Fertilizer consumption
Poverty rate