Institute for Food and Resource Economics

University of Bonn

Discussion Paper 2024:1 (Formatvorlage: Cover Paper number)

Agricultural Mechanization Around the World

Shahrear Roman, Hadi, David Wuepper*

Editor: Thomas Heckelei

Institute for Food and Resource Economics

University of Bonn

Nußallee 21

53115 Bonn, Germany

Phone: +49-228-732332

Fax: +49-228-734693

E-mail: thomas.heckelei@ilr.uni-bonn.de

Agricultural and Resource Economics, Discussion Paper 2024:1

Agricultural Mechanization Around the World

Shahrear Roman, Hadi, David Wuepper*

Abstract

Mechanization is one of the key ingredients for achieving high agricultural productivity. Despite its importance, there is currently no globally comprehensive information about countries' agricultural mechanization. Here, we propose and demonstrate a machine learning approach, relying on a large, novel training dataset, to not only produce an up-to-date and comprehensive dataset of countries' average agricultural mechanization, but also a global gridded map at ~ 5km resolution. Comparing our results to previously available data we find major improvements in accuracy, completeness, timeliness etc., and we notice that several countries are by now much more mechanized than reported so far. When investigating the association between mechanization and crop yield gaps we find a strong and robust link: For each 10-percentage point increase in mechanization, the associated crop yield gap decreases by 4 – 5 percentage points.

Keywords: Agricultural Mechanization, Labor, Tractors, Crop Yield Gap

JEL classification: C51, C53, C55, Q16, Q52, Q55

1 Introduction

Agricultural mechanization is globally relevant for addressing several United Nations Sustainable Development Goals, including No Poverty (SDG 1), Zero Hunger (SDG 2), Decent Work and Economic growth (SDG 8), Climate Action (SDG 13), and Life on Land[1] (SDG 15). It reduces the physical demands of farming, facilitates the timely execution of agricultural tasks, and improves productivity (Hamilton et al. 2022; Wuepper et al. 2023) and it improves working conditions, including allowing farmers to have more leisure time (Caunedo and Kala 2021). It also promotes more efficient use of inputs and minimizes spoilage losses (Yan et al. 2024). Finally, it also plays a crucial role in contributing to the mitigation of climate-related hazards (Emami et al. 2018; FAO 2017; Xinshen, Jed, and Hiroyuki 2016).

So far, however, at larger scales, there is comparably little information on countries' and regions' degree of agricultural mechanization, posing a severe constraint on our understanding of agricultural mechanization (FAO 2017).

Figure **1** illustrates the current data situation, showing the agricultural mechanization data published by World Bank (World Bank, 2024) which has just been taken from the web in autumn 2024 because it was so outdated and unreliable by now. This data is considered so far the main source of information about countries' agricultural mechanization (Léautier and Hanson 2013; World Bank Group 2017).

Figure 1, however, clearly shows the lack of harmonization across regions, the lack of within-country variation, and how outdated the data is by now. Even in regions like Europe and North America, the data is still outdated by about 15 to 50 years. This highlights a major gap in our understanding of agricultural mechanization, as there are no other comprehensive sources on this topic available.

In this study, we propose and demonstrate a machine learning (ML) framework to produce a fine-grained global map of agricultural mechanization around the world. This allows us to examine countries' current degree of agricultural mechanization, as well as the degree of agricultural mechanization within countries, and to quantify the link between agricultural mechanization and crop yield gaps at the global level.

Our ML approach starts with a large, newly assembled training dataset of known mechanization around the world. Then, the algorithm is trained to use environmental, infrastructural, agricultural, economic, and policy data to predict a detailed global map of agricultural mechanization rates. The final

[1] https://sdgs.un.org/goals

step is then an extensive testing and validation phase, to quantify how well the produced map predicts known mechanization rates.

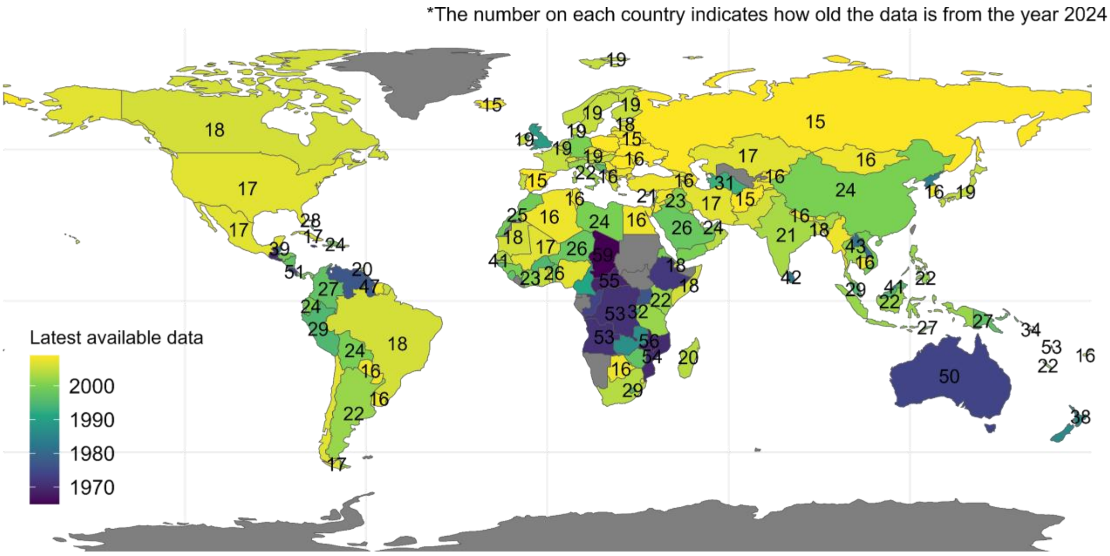

Figure 1: The Available Mechanization Data, Tractor Use Per 100 sq km of Arable Land. This is the most up-to-date information on agricultural mechanization in various countries. The latest available data, sourced from the World Bank, is from the year 2024. The mechanization data reflects the number of tractors used per 100 square kilometers of arable land. On the map, darker shades indicate older data, while lighter shades represent more recent data from 1961 to 2009. The most recent data is from 2009 and covers only 10 countries.

One of our several robustness checks is based on a Multivariate Environmental Similarity Surface (MESS), to quantify the similarity in conditions between grid-cells we have training data for and grid-cells for which we do not have training data. We also generate and compare a null model to our actual model and compare how much our algorithm improves upon a purely random map.

While a comparable fine-gridded global map for agricultural mechanization does not exist to date, previous research has produced comparable maps for other agricultural and environmental variables. This includes the research by Siebert et al., (2005) for the global distribution of irrigation infrastructure, Kamau, Roman, and Biber-Freudenberger (2023) for the profitability of diversified farming systems, and Wuepper et al. (2021) for different kinds of land degradation.

Prior research on agricultural mechanization has contributed valuable evidence on adoption patterns and socioeconomic impacts of mechanization already. For example, Pingali (2007) highlights that while many regions in Asia and Latin America have adopted labor-saving technologies, many Sub-Saharan African regions lag behind, indicating barriers rooted in economic and institutional constraints. Studies by Caunedo and Kala (2021) and Meng et al., (2024) emphasized the labor reallocation effects of mechanization, showing how it can enhance productivity and increase off-farm income while shifting agricultural labor dynamics. Lu, Du, and Qiu (2022) found mechanization significantly boosts crop yields, although this also increases risk under some conditions. Wuepper et al. (2023) show that countries' degree of economic freedom leads to more agricultural mechanization and this in turn increases crop yields.

Here, we make three contributions to this literature. First and foremost, we provide a harmonized, up-to-date, global map of agricultural mechanization, at a resolution of ~5 km (i.e., ~5 km × ~5 km grid cells). This will be useful in a wide range of agricultural economics applications, not only at the global level, but also within individual countries. Secondly, we demonstrate how to synthesize individually available empirical data points on agricultural and environmental variables in a machine learning algorithm to create a fine-grained global map. Finally, we analyze the link between agricultural mechanization and crop yield gaps globally. This analysis cannot establish causality, but we show that the link is both strong and robust, and survives the inclusion of various control variables.

The rest of the paper is organized as follows. In section 2, we explain our data and methods. In section 3, we explain our main results. Section 4 presents a discussion on agricultural mechanization and especially its relationship with crop yields as well as the potential and main limitations of our modelling framework. The last section presents our conclusions.

2 Data and Methods

Two distinct types of data are required for our modelling approach. First, precise field or farm-level information, including the geographic coordinates of areas where the specific motorized machine (farming and processing technologies, for example, tractors, seed drills, cultivators, harvesters, etc.) are utilized. We call this 'training data'. Secondly, based on the relevant theoretical background (Foster and Rosenzweig 2010; Shang et al. 2021) we collated different spatially explicit variables (environmental, economic, infrastructure, demographic, and policy variables) that are predictive of agricultural mechanization adoption and diffusion. Then, we applied our machine-learning algorithm to model the global distribution of agricultural mechanization.

2.1 *Data*

Training Data

We use 961 unique training data points with geo-coordinates based on two sources. First, approximately 59% (n=560) of the geo-locations were obtained from the Global Database on Sustainable Land

Management (SLM) of WOCAT[2] (the World Overview of Conservation Approaches and Technologies) (WOCAT 2024). This database contains detailed data on 2,433 SLM practices in 136 countries. We only selected those SLM practices where farmers use any kind of motorized machine. To account for missing data from certain countries or regions, we filled in data gaps with an additional 41% (n=401) training data that were collected with a large global expert survey (Appendix A1)

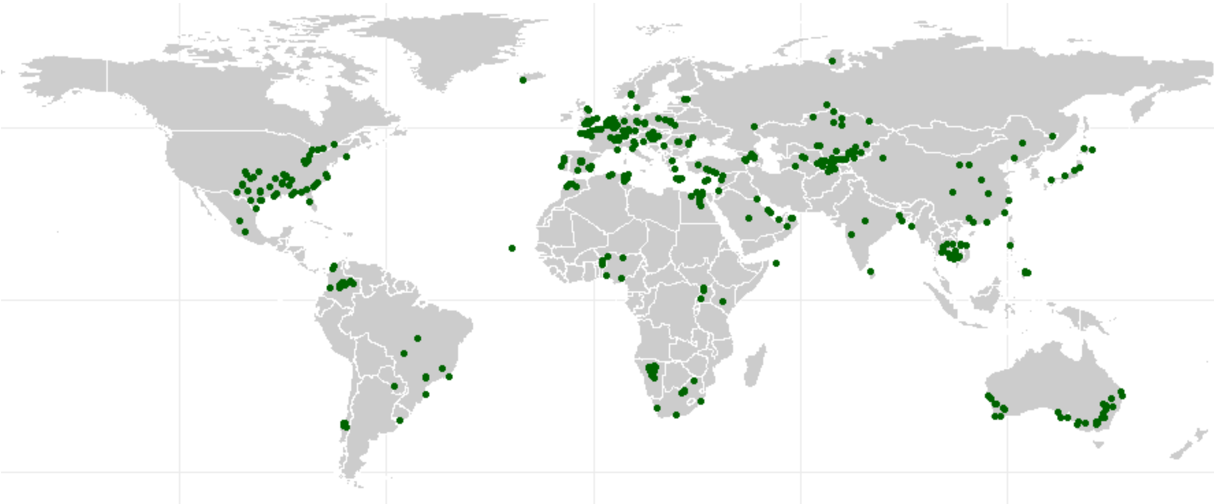

Figure 2: Global Locations Where Agricultural Mechanization is Practiced. Mechanization refers to farming and processing technologies, which include motorized equipment, for example, tractors, seed drills, cultivators, harvesters, etc. (N= 961).

Variables

Different variables explain agricultural mechanization at different scales. We used a theory-informed set of fourteen different variables as our predictors for the machine learning algorithm (see **Table 1**). Broadly, the predictor variables include economic variables, environmental variables, infrastructural variables, sociodemographic variables, policy variables and agricultural variables.

We used temperature, precipitation, a terrain ruggedness index, and soil organic carbon at the grid cell level as environmental variables. Environmental variables (Anwar et al. 2015; Lal 2018) are critical as they directly affect crop growth, and with mechanization, farmers can work effectively and timely on the field even in suboptimal situations. Moreover, soil health, indicated by organic carbon levels, affects crop productivity and determines land aptness for mechanization (Lal 2018).

We also included the country's Economic Freedom Index and subnational Human Development Index as economic variables (Rochecouste et al., 2015) as well as socio-demographic variables (Ricker-

[2] https://wocat.net/en/

Gilbert, Jumbe, and Chamberlin 2014) comprising population density and demographic characteristics of people aged 15 to 59 at the sub-regional level. Higher economic freedom and human development levels generally correlate with increased mechanization due to improved access to capital, technology, skilled labor, favorable business conditions, and government support (Xinshen, Jed, and Hiroyuki 2016). Socio-demographic variables provide insights into labor availability (Hiroyuki 2016) and workforce demographics, which are crucial considerations for agricultural mechanization. Areas with labor shortages drive the demand for mechanization as an alternative to manual labor. Infrastructural variables, e.g., travel time to the nearest cities (accessibility) is important for mechanization (Mvodo and Liang 2012), are used at the grid cell level. These variables significantly influence the feasibility and efficiency of mechanization. Agricultural variables, such as cropland area, and field size (Rasouli, Sadighi, and Minaee 2009) are more conducive to mechanization, enabling larger machinery and economies of scale. Additionally, proximity to urban centers facilitates access to markets, services, and infrastructure necessary for mechanized farming operations. We also incorporated property rights protection at the country level as a policy variable because higher property rights protection increases institutional support for mechanization (Hagedorn 2004). Finally, the number of crop types per pixel was added to capture agricultural diversity. All of these indices reflect the economic environment, governance quality, and investment climate, collectively influencing farmers' ability and willingness to invest in mechanized technologies. We believe understanding these variables' roles and interplay is essential to model the agricultural mechanization in agriculture.

Table 1. Model Input Variables

Variables	**Description**	**Units**	**Year**	**Resolution**	**Source**
Economic freedom index	Measures countries’ economic freedom by the security of property rights, inflation control, absence of excessive regulation, and various other indicators	Index (1:100)	2023	Country level	(the heritage foundation 2024)
Human development index	A composite index assessing average achievements in education, income, and health	Index (0:1)	2015	~10-km grid	(kummu, taka, and guillaume 2018)
Soil organic carbon	the carbon stored in the soil's organic matter, contributing to soil health, fertility, and overall ecosystem functioning	Cg/kg	2021	~5-km grid	(poggio et al. 2021)

Soil water content	Indicates the volume of water retained in the soil, a critical factor for agricultural potential	Kilopascals (-10 kpa)	2021	~5-km grid	(poggio et al. 2021)
Precipitation	The cumulative amount of rainfall or other forms of precipitation that an area receives over a year	mm	Average (1970 - 2000)	~5-km grid	(fick and hijmans 2017)
Temperature	The average temperature recorded in a given area over the calendar year	Degree celsius	Average (1970 - 2000)	~5-km grid	(fick and hijmans 2017)
Terrain ruggedness index (tri)	Calculated as the mean of the absolute differences in elevation between a focal cell and its eight neighboring cells	Index	2020	~90-m grid	(amatulli et al. 2020)
Travel time to cities	Assessing the time required to reach the nearest densely populated area with at least 1,500 inhabitants as proxy for market access	Minutes	2015	1-km grid	(weiss et al. 2018)
Population	Represents the number of individuals residing per square kilometer, based on 2020 population estimates that align with national census data and population registers	Number of persons per km^2	2020	~1-km grid	(ciesin 2018)
Demographics	Provides estimates of the working-age population (ages 15-64) and their densities (individuals per square kilometer), consistent with national census data	Number of persons (15-64y) per km^2	2010	~1-km grid~1-km grid	(ciesin 2018)
Property rights protection	Measures perceptions of the security of property rights, distinct from other aspects of the rule of law. This index synthesizes 18 individual indicators to	Index (1:100)	Mean across years	Country level	(wuepper et al., 2024)

	offer a comprehensive view of property rights protection				
Number of crop types	Encompasses 42 individual crops and broader crop categories, offering a detailed representation of global agricultural production and its distribution across various regions	Number of crop types per pixel (count)	2010	~10-km grid	(yu et al. 2020)
Cropland area	Cropland refers to land used for the cultivation of annual and perennial herbaceous crops intended for human consumption, forage (including hay), and biofuel production	Percent of cropland per pixel	2019	3-km grid	(potapov et al. 2022)
Field size	Categorizes fields in very small (class 3506: <0.64 ha), small (3505: 0.64–2.56 ha), medium (3504: 2.56–16 ha), large (3503: 16–100 ha), and very large (3502: >100 ha)	Categorical	2019	30-m grid	(lesiv et al. 2019)

We conducted a Pearson correlation test (Appendix A2.1), which reveals that most variables do not show a strong correlation with one another. However, there is a notable correlation between population density and the demographic characteristics of the working-age population, specifically those aged 15 to 64. Higher population densities lead to increased food demand, which in turn necessitates greater mechanization in agriculture to effectively enhance productivity. Additionally, the proportion of the working-age population (defined as individuals aged 15 to 59) is crucial for sustaining agricultural productivity, as a larger workforce is essential for efficient agricultural operations. Furthermore, we find a significant correlation between the economic freedom index and the protection of property rights, which is mechanistic, as protection of property rights is one out of many indicators for economic freedom.

2.2 *Modeling Approach*

We use the machine learning approach as it can easily accommodate complex relationships among multiple outputs, and its flexibility allows it to represent a range of linear and nonlinear relationships, as well as complex interactions, which results in high predictive performance (Baylis et al., 2021). Here, we implemented the Maxent modelling which apply machine-learning technique called maximum entropy modeling in R version 4.3.2 (R Core Team, 2023) using the package 'ENMeval 2.0' (Kass et al., 2021). Following best practices, we tested an extensive set of candidate models, with different functional forms and model complexity. We use linear (L) to complex nonlinear (quadratic (Q), hinge (H), product (P), threshold (T)) feature classes (L, LQ, H, LQH, LQHP, LQHPT). We also use low to high regularization multipliers (0.5, 1, 1.5, 2, 3, 5), which controls the generalization ability of the model. In total, this extensive model parameters combinations result in seventy-two candidate models.

Subsequently, out of seventy-two models, we selected the best model that balances predictive performance and model simplicity (i.e., minimize overfitting). This was done based on several validation metrics, including area under the receiver operating characteristic curve (AUC) (Hirzel et al., 2006) omission rate (or.10p.avg), and Continuous Boyce Index (CBI), following best practices (Hirzel et al., 2006; Kass et al., 2021; Radosavljevic & Anderson, 2014). AUC values range from 0 to 1, with a value of 1 indicating perfect discrimination ability, with 0.5 or less indicating no discrimination ability of the model (random prediction). Omission rate ranges from 0 to 1, with 0 indicating no omission errors (perfect prediction). CBI (Continuous Boyce Index) is used as a metric to assess the accuracy of a model, specifically through the Spearman correlation between the predicted-to-expected (P/E) ratio of habitat suitability values and the mean habitat suitability index (HSI). Its values range from -1 to +1, where a value of 1 indicates a perfect prediction, 0 reflects a random prediction, and -1 signifies a highly inaccurate prediction (Farrell et al., 2019).

We use the spatial cross-validation method 'checkerboard' partition (Radosavljevic & Anderson, 2014)). Spatial cross-validation accounts for the geographic locations of the training data in the process of partitioning the data into the model calibration set and the model evaluation set. This helps optimize the model transferability (i.e., extrapolation ability) from locations and regions with observed (sampled) training data, to other (unsampled) locations and regions, based on the learned statistical relationship between the outcome variable and the predictor variables.

Our different model specifications (e.g., functional forms) tested also allow us to generate a spatially explicit understanding of model uncertainty. We estimated the model-based uncertainty as the variability (standard deviations) of the predictions from our all-candidate models created through different hyperparameter tuning (linear to a complex feature class and low to high regularization multiplier. We also conduct MESS (Multivariate Environmental Similarity Surface) analysis to evaluate the reliability of model predictions in various conditions (Appendix A3.1).

We also run a null model test to see if the model's accuracy is significantly higher than the null model. This would indicate whether the model captures meaningful patterns in distributions of the agricultural mechanization rather than just fitting to noise in the data or random predictions. Finally, we calculated variable importance in the optimal model, based on the permutation feature importance. This measures the contribution of each variable based on the resulting decrease in model accuracy if we randomly permute the values of that variable among the training points.

2.3 *Further analysis*

In our analysis of the relationship between mechanization levels and crop yield gaps, we utilized our latest available mechanization data (Figure 3a) and crop yield gap data (Gerber et al., 2024). The yield gap is defined as the difference between observed crop yields at any given location and attainable (i.e., exploitable (van Ittersum et al. 2013) not agronomic potential) yields of these crops in this location, expressed as percentage relative to attainable yields. The yield gap data represents the ten most important global crops (wheat, maize, rice, barley, sorghum, cassava, soybean, rapeseed, oil palm, and sugar cane). Our approach employs ordinary least squares (OLS) regression models, with and without incorporating various sets of control variables to enhance the robustness of the model. In the first model, we only used agricultural mechanization as an explanatory variable, while the yield gap is the dependent variable. From the second to fifth model, we controlled for various variables to assess mechanization's impact on yield gaps. First, we included environmental variables namely precipitation, temperature, and terrain ruggedness to understand their associations together with mechanization. Next, we added economic indicators such as GDP, the Human Development Index, and the Economic Freedom Index. We also incorporated policy-related variables, specifically property rights protection, to evaluate how governance affects mechanization's influence. Finally, by applying country-fixed effects with the complete set of controls, we acknowledged country-specific factors, emphasizing their role in shaping mechanization's effectiveness in reducing yield gaps both within and between countries.

3 Results

Our study presents a comprehensive analysis of agricultural mechanization across the globe, offering critical insights into its spatial distribution and potential impact on crop yield gaps. In the following sections, we present and analyse key findings from our study. At first, we explore global patterns of agricultural mechanization, offering insights of agricultural mechanization across and within countries. in the next section, we provide a detailed comparison of the datasets with other agricultural variables. Finally, Section 3.3 examines the association between mechanization and crop yield gaps, shedding light on how varying mechanization levels associated with yield gaps.

3.1 *Agricultural Mechanization Around the World*

Our map of the global distribution of agricultural mechanization is displayed as **Figure 3a**. What is modeled here is the probability that a farm is mechanized (i.e., tractors and similar machinery are used), ranging from 0 to 100%, corresponding to the share of mechanized farms in a given region on the same scale.

Western Europe is the world's most mechanized region, with an average mechanization rate of 68%, followed by North America at 62% and Oceania at 56% (Appendix A4.1). In Western Europe, over 92% of farms/cropland areas have a mechanization level above 50%, with minimum and maximum mechanization rates of 44% and 86%, respectively. Similarly, North America has about 85% of its farms with mechanization levels over 50%, ranging from 44% to 82%. Oceania follows with 69% of farms surpassing 50% mechanization, ranging from 40% to 76%.

On the other hand, the least mechanized regions are Sub-Saharan Africa and Middle Eastern countries, with an average mechanization rate of 37%. Only ~16% of cropland areas in this region have a mechanization level above 50%, with maximum and minimum mechanization rates of 69% and 24%, respectively.

In South East Europe and Western Asia, average mechanization rates are 53%, with 54.37% and 55.9% of farms exceeding the 50% mechanization mark, respectively. Central America and the Caribbean Islands have an average mechanization rate of 51%, with nearly half of the farms /farmers having more than 50% mechanization. Northern Western Africa and Central Asia show similar patterns, with average rates of 50% and around 48% of farms having high mechanization levels.

South Asia and Southern Africa both have an average mechanization rate of 48%, with nearly half of the farms in these regions having more than 50% mechanization. Eastern Asia and South America have lower average mechanization rates of 45%, with 35.59% and 32.59% of farms, respectively, exceeding the 50% mechanization level.

Figure 3b shows the degree of model uncertainty, measured as the standard deviation of all models developed by tuning the hyperparameter (feature class and randomization multiplier). Overall, most of the map has a very low uncertainty, with the global median of 0.04. The density plot's standard deviation illustrates the uncertainty values' distribution, mean, and median. However, there are regional uncertainty hotspots with uncertainty values up to ~0.20 in some parts of East Europe and Asia.

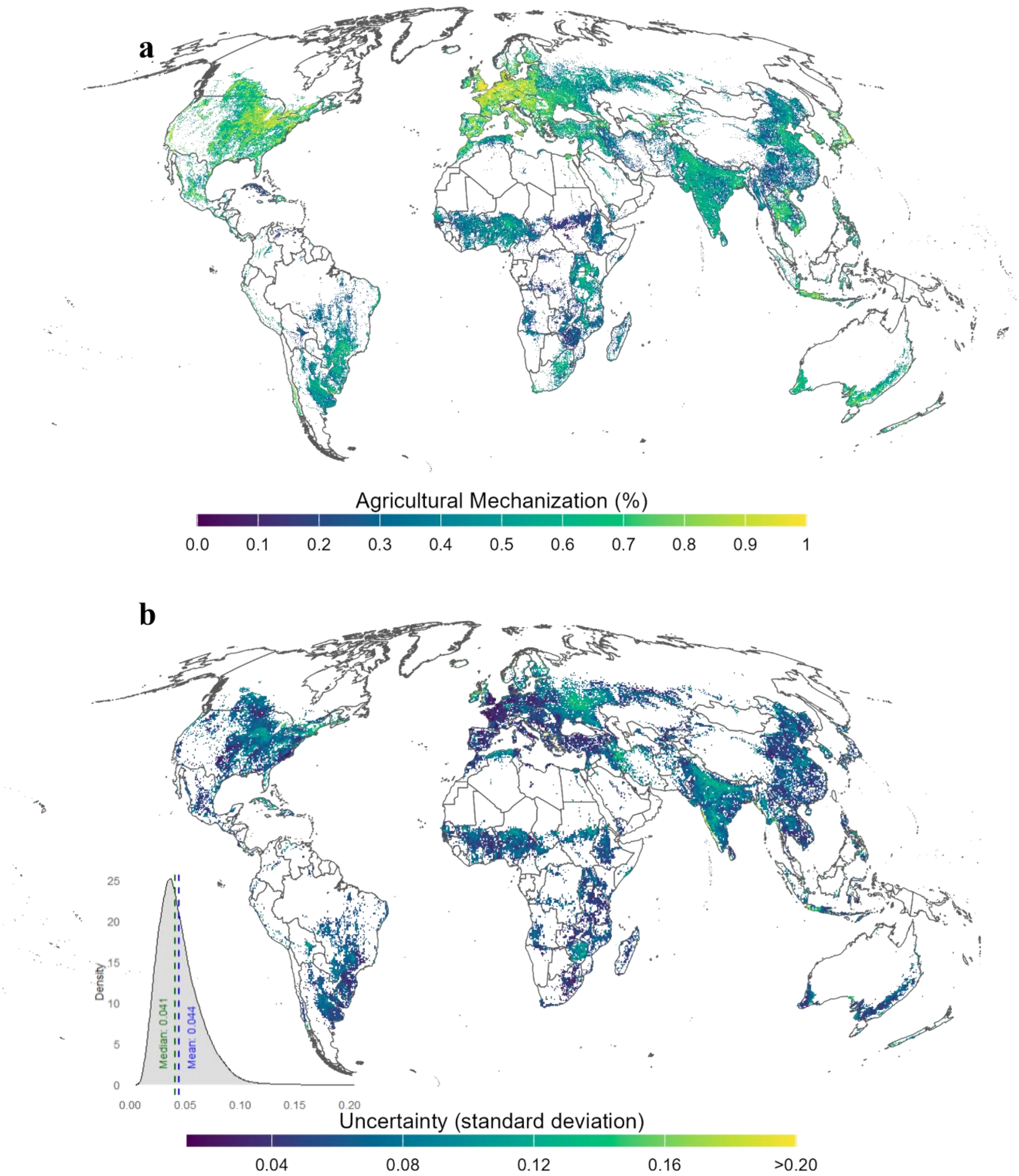

Figure 3: Global distribution of agricultural mechanization. (a) The global map of mechanization levels at 5-km resolution. Dark blue to yellow, ranges from low to high levels of mechanization. (b) Uncertainty map of the prediction of agricultural mechanization levels, measured as the variability in the predictions from the thirty-six different specifications model (standard deviation of the predictions).

To better understand the pattern of mechanization in agriculture within and across countries, we closely examined countries/regions on various continents, as depicted in **Figure 4**. This figure explains the considerable variation in mechanization levels among agricultural nations, highlighting the disparities between countries in the northern and southern hemispheres. The countries in the Northern Hemisphere, particularly Western Europe and North America, demonstrates a high and relatively uniform agricultural mechanization both across and within countries. In contrast, most countries in the Southern Hemisphere, with the exception of Oceania, exhibit lower levels of mechanization.

Western Europe is the region with the highest agricultural mechanization and the most evenly distributed mechanization levels between and within a country (Figure 4). For example, France, Belgium, Netherlands, and Germany have a similar agricultural mechanization with the mean of 72%, 77%, 78% and 76% respectively (Appendix A5.1. Among these Western European countries, France has visible heterogeneity in mechanization within the northern and southern parts. For example, Northern France (highest 88%), well known for its grain production, has high mechanization levels, whereas the southern regions have slightly less mechanization (lowest 14%). Nonetheless, 98% of farm's mechanization level is above 50% (Appendix A5.1).

Figure 4 displays US agriculture is also highly mechanized, with an average of ~63% (~85% farm's mechanization level more than 50%) showing a high concentration of mechanization in the areas in the Midwest, the nation's agricultural heartland, often referred to as the "Corn Belt". This region (including, marked in chartreuse (yellow-green), signifies highly mechanized farming practices in that region (**Figure 4**). The western and southern regions exhibit less mechanized areas than the Midwest. In Canada, about 85% of firms have a mechanization level above 50%, with an average of 62%. But in Mexico, about 56% of firms have a mechanization level above 50%, with an average of 54%. On the other hand, Brazil's mean mechanization level is relatively lower (44%), which is more concentrated in the south and southeast (maximum mechanization level 76%), reflecting the dominance of mechanized soybean, sugarcane, and cotton farming in these regions. The northern areas, particularly the Amazon basin, show sparse agricultural activity, which aligns with environmental protection policies and less arable land.

Compared to Europe or North America, the mechanization level in Asia is visibly lower (apart from Japan and South Korea). For example, the average mechanization level in India and China is 51% and 42%, with the highest at 75% and 83%, respectively, which is noticeable in the map.

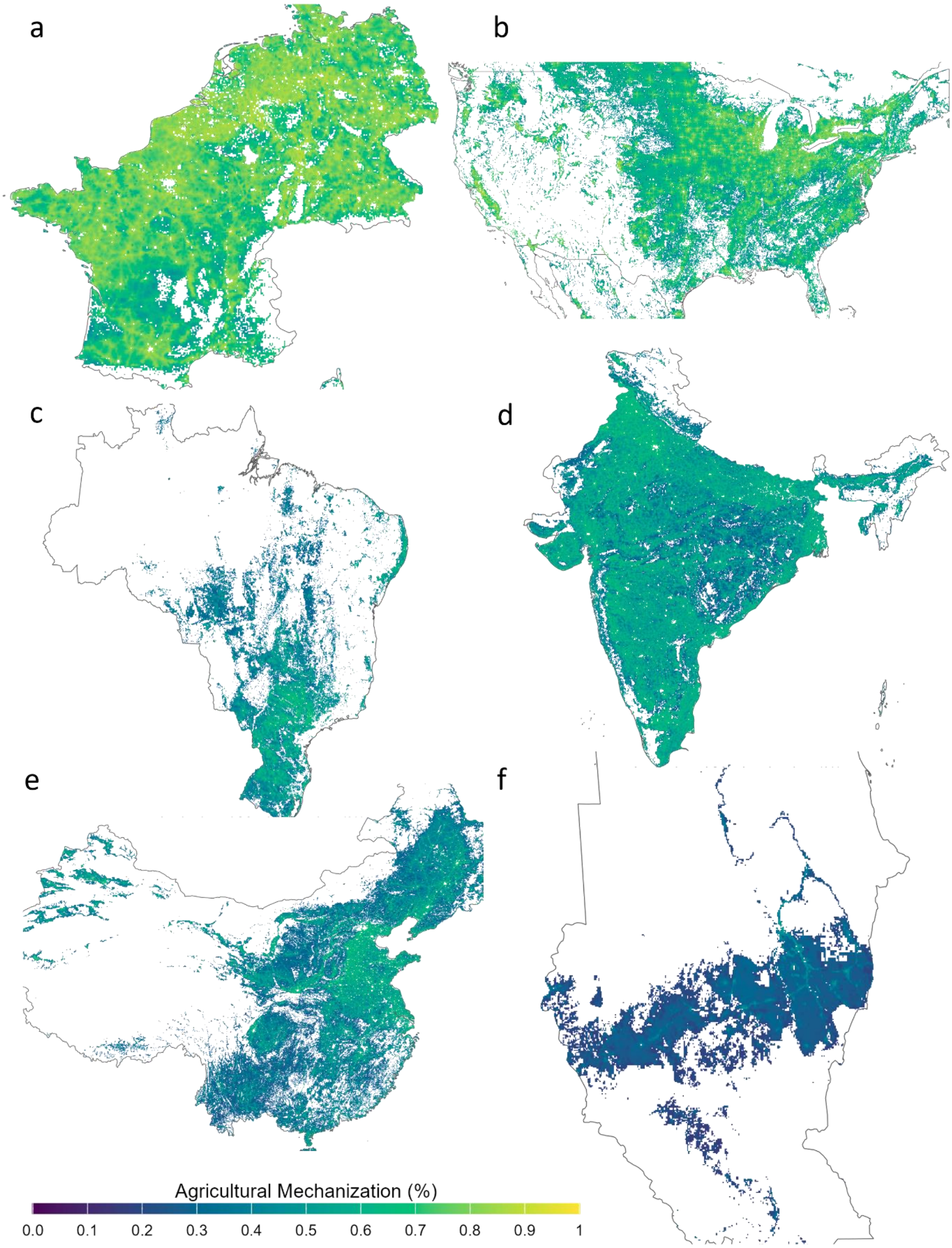

Figure 5: Regional distribution of agricultural mechanization. High-resolution (5-km) map reveals substantial variability in agricultural mechanization levels both within regions and countries. The spatial distribution of agricultural mechanization in a) France, Belgium, Netherlands, and Germany; b) the United States; c) Brazil; d) India; e) China and f) Sudan

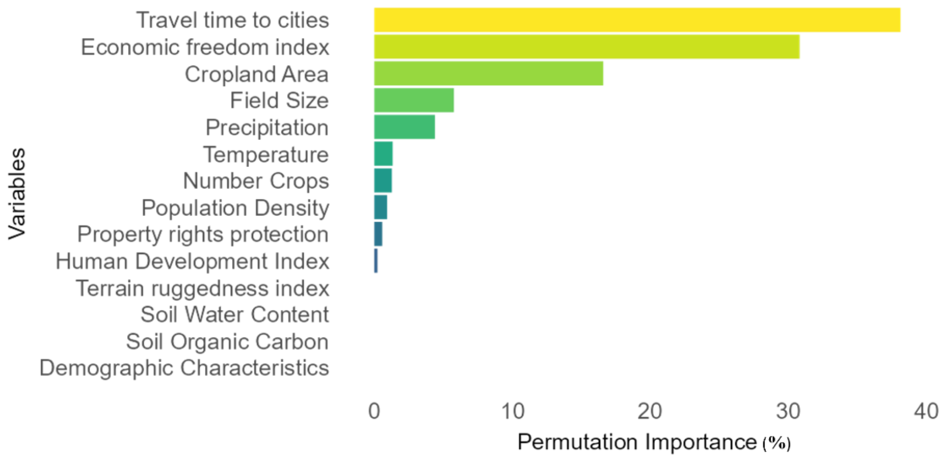

Figure 5: Permutation Variable importance. The values represent each variable's relative importance (normalized to percentages) in the model. The higher values of the variables indicate a greater importance for mechanization in agriculture.

In India, there is high agricultural mechanization in the northern plains, particularly around Punjab and Haryana, known as the "breadbasket", and also in the southern part, namely Andhra Pradesh, Tamil Nadu, and Kerala. In contrast, the central and eastern regions show less mechanization in agriculture. On the other hand, Eastern and northern China, especially the Yangtze River basin and the North China Plain regions, are more mechanized, where farmers mainly produce Corn, Wheat, and soybeans. In contrast, in south China, where rice is produced, it is less mechanized, for example, in the Pearl River basin region.

On the other hand, Sudan is one of the least mechanized countries in Figure 4(f), with average mechanization of 26%, with high and low 15% and 64% showing sole mechanization along the central part of the country, likely around the Nile River basin. In Sudan, only 0.64% of firms have a more than 50% mechanization level. Apart from that, most of the cropland is not mechanized.

Model Performance

Our machine learning model has a high predictive power; the spatial cross-validation AUC is 0.74. The model's AUC is significantly higher (p-val < 0.01 & z-score 3.818) than the null model's AUC (0.59). The model has an average test omission rate (at a 10% threshold (average)) of 0.10, an AUC difference between training and validation of 0.01, and the Continuous Boyce Index (CBI) of 0.97, suggesting the empirical model effectively and accurately captures the patterns in the data, with high predictive capability while minimizing overfitting.

Variable Importance

We find that infrastructural, economic, agricultural and environmental variables are the most important predictors of agricultural mechanization. Notably, "Travel time to cities (proxy for accessibility)," an infrastructural variable, is the most influential factor, having 38% of the total importance across all features/predictors. This highlights the critical role of infrastructure in enabling access to mechanized farming. Economic variables, notably the "Economic Freedom Index," make the second-largest contribution, accounting for approximately 31% of the total importance across all features. This underscores the importance of a conducive economic environment in fostering mechanization. "Cropland area" and "Field size" rank third and fourth in their importance, with nearly 16% and ~6%, respectively. These variables emphasize the significance of land availability and management practices in adopting mechanized agriculture. Environmental variables, such as precipitation (4.3%) and temperature (1.3%), contribute only slightly to the prediction. While their influence is smaller, it remains crucial, particularly in regions where extreme weather conditions can hinder mechanization.

3.2 *Datasets Comparison*

We also explore the relationship between the agricultural mechanization at the country level with other agricultural variables. We considered the number of tractors per 100 square kilometers of arable land (World Bank 2024), and fertilizer consumption per hectare (World Bank 2024) to compare the agricultural mechanization.

Agricultural machinery includes wheel and crawler tractors, reflecting the mechanization intensity across regions. Fertilizer consumption, which measures the use of nitrogenous, potash, and phosphate fertilizers, serves as a key indicator of agricultural input efficiency (excluding traditional nutrients like manure). Due to the lack of comprehensive data for tractors per 100 square kilometers of arable land after 2000 in the World Bank Databank for many countries, we took the average from 1997-2002. However, we used the latest available data for fertilizer consumption kilograms per hectare of arable land (2021) to ensure an accurate and relevant comparison.

Figure 6 shows a scatter plot on the relationship between our modelled agricultural mechanization and reports number of tractors per 100 square kilometers of arable land (logged), as provided by the World Bank so far. Overall, the correspondence is strong (Pearson’s correlation of 0.42), but unsurprisingly, there are several outliers. Several countries are by now much more mechanized than indicated in the data of the World Bank. Examples are Rwanda, Bangladesh, Indonesia, Togo, and Senegal, to name a few.

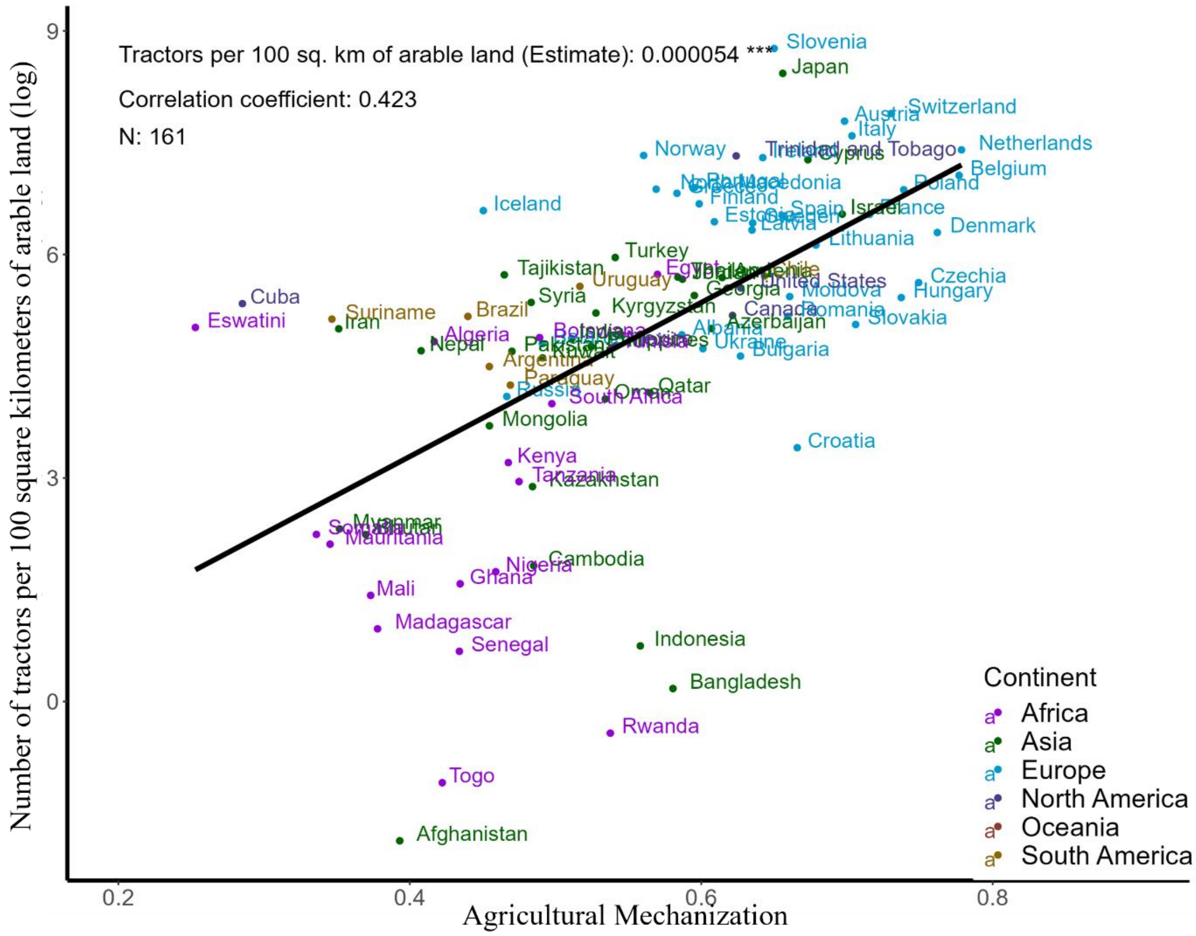

Figure 6: Mechanization and Number of Tractors Use. The relationship between modelled agricultural mechanization based on machine learning and the reported number of tractors per 100 square kilometers of arable land in the official data, as provided by the World Bank so far. Here, mechanization and fertilizer use, emphasizing the stark contrasts in agricultural inputs and practices across regions.

Figure 7 shows the relationship between modelled agricultural mechanization and fertilizer consumption per hectare of arable land (logged) at the country level (World Bank, 2024). The upward trend of the regression line, supported by a correlation coefficient of 0.27, underscores a good correspondence between mechanization and fertilizer use. Countries in Europe and North America, represented by blue and cyan dots, stand out with both high mechanization and high fertilizer use, indicative of their intensive agricultural practices. Asian nations tend towards higher fertilizer consumption, though their mechanization levels remain moderate or low. On the other hand, African countries, marked by purple dots, exhibit lower levels of mechanization.

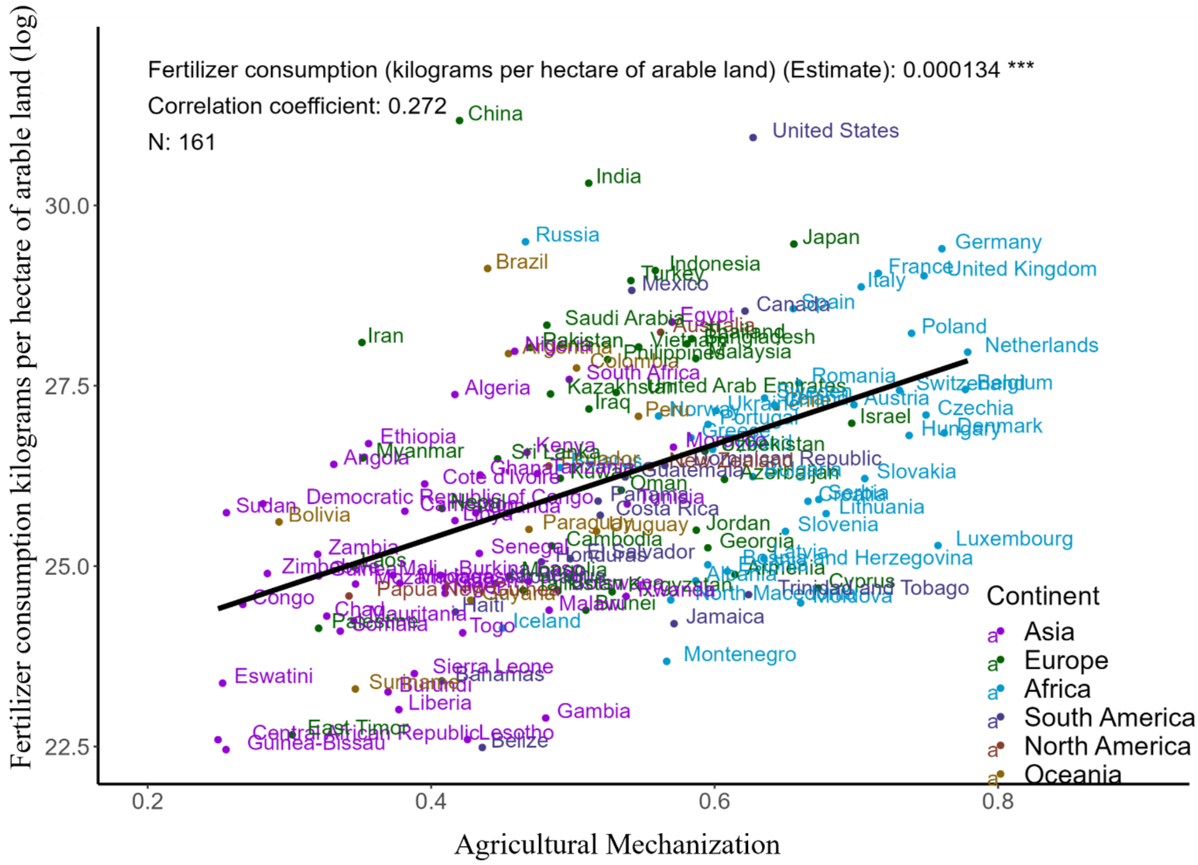

Figure 7: Mechanization and Fertilizer Use. The relationship between modelled agricultural mechanization based on machine learning and the reported fertilizer consumption (kilograms per hectare of arable land).

3.3 *Mechanization and Crop Yield Gaps*

Figure 8 shows the association between the agricultural mechanization and the average yield gap, globally for the ten most important global crops.

Overall, this analysis indicates that globally, for each one percentage point increase in mechanization, the yield gap decreases by 0.36 to 0.48 percentage points. **Figure 8a** presents the regression coefficients with various confounding factors controlled for. Specification 1 is the baseline model without any control variables, where the coefficient of mechanization is −0.48. This implies that overall, one percentage point increase in mechanization is associated with a 0.48 percentage point decrease in the yield gap. This is statistically highly significant (p-val < 0.00 and a Z-score of −233.24). Specification 2 adds environmental variables, such as precipitation, temperature, and terrain ruggedness index. The mechanization coefficient is −0.36 (p-value < 0.00). In Specification 3, we additionally control for economic variables, including GDP, the Human Development Index, and the Economic Freedom Index, along with environmental variables. Here, the coefficient slightly increased to −0.37 (p-< 0.00). In

Specification 4, we further add the policy 'property rights protection' and obtained the coefficient of −0.38 (p-< 0.00), indicating that policy variables further contribute to the observed effect of mechanization on yield gaps. Finally, in Specification 5, we incorporate country-fixed effects and the full set of controls. Here, the mechanization coefficient drops to −0.16. This suggests that country-specific factors explain a substantial portion of the variation, and mechanization in agriculture can play a meaningful role in reducing the yield gap across the country and in between the countries.

Figure 8b shows a binscatter plot of the association between the agricultural mechanization (x-axis) and the average yield gap (y-axis). A regression line with a 95% confidence interval (Conley standard error) visualizes the very clear pattern in the data that there is a strongly negative and mostly linear association between the agricultural mechanization and the average yield gap.

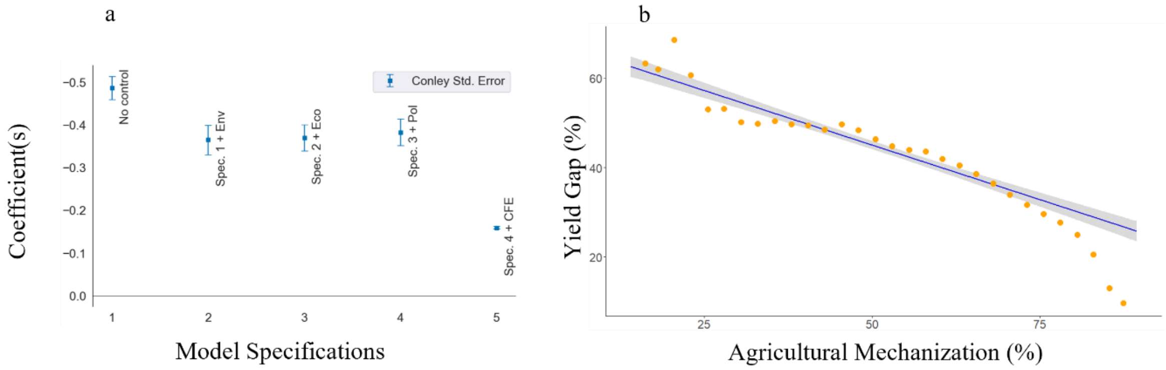

Figure 8: Agricultural Mechanization and Crop Yield Gaps. (a) regression coefficients of the agricultural mechanization (%) for different regression model specifications with yield gap (%) as the dependent variable. Specification 1 is without any controls. Specification 2 includes environmental controls. Specification 3 includes additional economic controls. Specification 4 further adds policy controls. Finally, specification 5 additionally includes country-fixed effects. (b) Binscatter plot of the relationship between the agricultural mechanization (%) and the yield gap (%). (For both a &b panel, the p-value < 0.00).

Figure 9 is a bivariate map depicting the relationship between the agricultural mechanization and the yield gap. In the map, yellow represents regions with both high mechanization levels and high yield gaps, while gray indicates areas where both are low. Blue signifies regions with a high yield gap but low mechanization, whereas red highlights areas with high and low yield gaps.

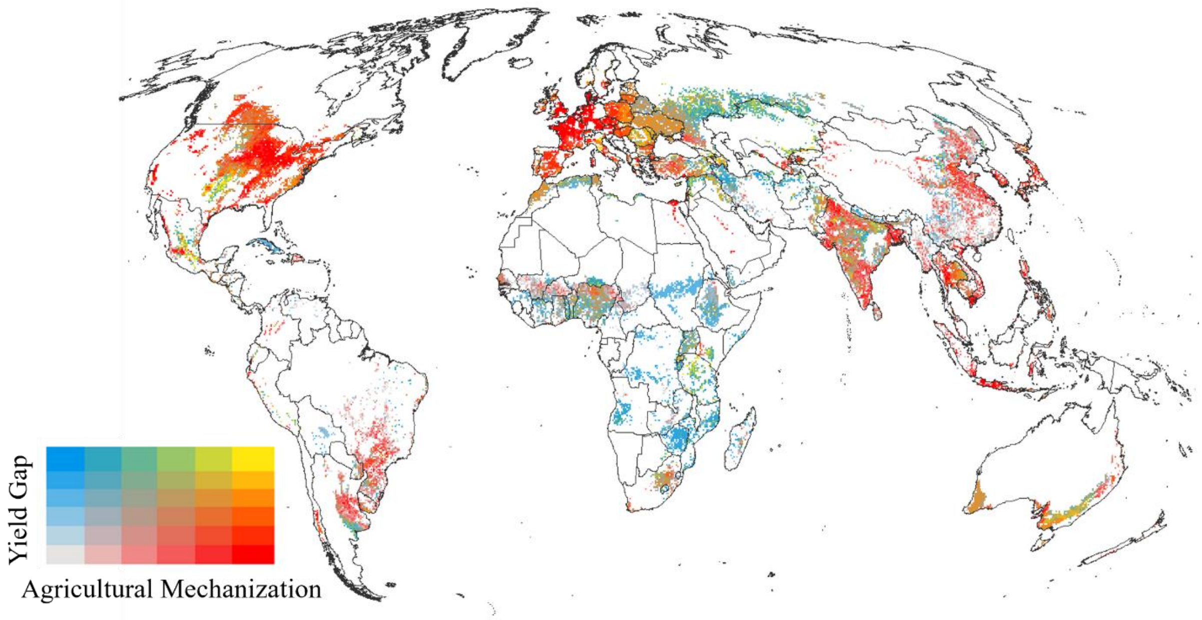

Figure 9: Bivariate Map for Agricultural Mechanization and Yield Gaps. Yellow indicates high agricultural mechanization and a high yield gap, while gray indicates both are low. Blue indicates a high yield gap but a low agricultural mechanization, while red indicates a high agricultural mechanization with a low yield gap.

This map reveals a clear pattern in the agricultural mechanization and yield gap across different regions. In North America and Western Europe, high levels of mechanization have effectively minimized the yield gap, reflecting advanced agricultural practices. Conversely, large parts of Sub-Saharan and middle Africa and some scattered regions of Asia are characterized by a significant yield gap coupled with low mechanization levels, indicating substantial room for potential improvement in crop yields through increased mechanization. Interestingly, Eastern Europe, Western Asia, and Central Asia exhibit moderate mechanization but still struggle with a considerable yield gap. However, in South and Eastern Asia, mechanization is high for most of the area, and the yield gap is low for most of the area, though there is some scattered cropland area that is not highly mechanized, and the yield gap is high. This opens doors to finding specific areas or regions to enhance agricultural productivity and reduce the yield gap by improving mechanizations.

4 Discussion

We propose and demonstrate a novel way to model the rate of agricultural mechanization at the global scale in high resolution. The resulting map provides highly policy and research relevant information and our approach can be used to model many other variables of interest.

Our approach is innovative in its method, scale and resolution and in the use of advanced machine-learning modeling techniques to capture the nuanced patterns of mechanization across various geographies. To our knowledge, no prior study has examined the spatial distribution of mechanization at national, continental, or global levels with such a comprehensive and integrated perspective, providing critical insights for policymakers aiming to optimize agricultural strategies worldwide. We argue that agricultural economics has a lot of use for this approach, as agricultural mechanization is by far not the only agricultural phenomenon for which harmonized, high-resolution data at the global scale is currently unavailable.

Our analysis highlights significant global disparities in agricultural mechanization, underlining how access to mechanization varies drastically across and within regions. Mechanization is largely concentrated in high-income, agriculturally intensive countries, where variables such as economic freedom, infrastructural development, and large-scale farming (field size) contribute to higher mechanization levels. Regions like North America, Europe, and Australia stand out with high levels of mechanization, driven by favorable policy environments, economic freedom, larger field sizes, and better connectivity to urban centers. Within these regions, our findings reveal that proximity to cities is a critical determinant—longer travel times reduce mechanization adoption significantly. This relationship is particularly pronounced in North America, where the western and southern fringes show lower mechanization levels compared to the highly accessible midwestern and eastern regions. In Latin America, mechanization is primarily observed in the southern parts of Brazil and Argentina, where large commercial farms dominate. However, within-country disparities are evident; in Brazil, while the southern and southeastern regions are highly mechanized, the Amazon Basin and northeastern states show significantly lower mechanization levels, driven by infrastructural challenges and distinct land-use patterns.

Conversely, in many low- and middle-income countries, especially in Africa and parts of South Asia, mechanization remains limited due to smaller farm sizes, low levels of economic freedom, inadequate policy support, and infrastructural bottlenecks. Africa has the most considerable agricultural mechanization gaps globally. Despite some progress in northern African countries like Egypt, much of sub-Saharan Africa continues to rely heavily on manual labor and animal traction, with mechanization limited to a few commercial hubs (Baudron et al. 2019). The low agricultural mechanization levels here are not only a function of economic and infrastructural constraints but also a result of inadequate policy frameworks that fail to address the unique needs of smallholder farmers (Dileepkumar 2013).

Infrastructure-related variables, travel time to the nearest city, Agricultural variable, field size, and cropland area, alongside the Economic Freedom Index (economic variable), emerge as the top four most significant variables influencing mechanization levels. Environmental variables also play a critical but complex role. Our findings reveal that extreme precipitation and rugged terrain are linked to negatively associated agricultural mechanization, likely due to operational difficulties for machinery in these conditions. Whereas demographic variables, e.g., population density and demographic characteristics, did not appear as very important variables for mechanization.

We also incorporated yield gap estimates to identify areas where mechanization could bridge productivity deficits. By mapping mechanization levels alongside yield gaps, our study also discovered the regions where improving mechanization could effectively close these productivity shortfalls. For example, regions like Sub-Saharan Africa and parts of Southeast Asia show high yield gaps and low mechanization, indicating a significant opportunity for mechanized solutions to elevate low productivity and reduce reliance on manual labor. In contrast, areas with relatively higher mechanization but persistent yield gaps (Gerber et al. 2024), such as parts of Eastern Europe, suggest that variable like fertilizer, crop management, or access to quality inputs may also need to be addressed along with mechanization. Thus, our findings underscore that the relationship between mechanization and yield gaps is highly context-specific, influenced not only by economic variable but also by a blend of infrastructural, agricultural and environmental variables that must be considered in any strategy to enhance agricultural productivity globally.

Limitations. Similar to other machine learning modelling approaches, our model prediction of agricultural mechanization is subject to uncertainties stemming from diverse input sources, and the inherent complexities of the model (Jain 2020). Another key source of uncertainty is the integration of variables from different years and types with varying spatial resolutions, a compromise that is necessary due to the lack of a globally harmonized dataset on mechanization. To mitigate these issues, we employed rigorous validation techniques to validate our models, as well as evaluated an extensive set of candidate models, obtaining an optimal model with high accuracy and robust generalization capability. This ensures our prediction at the global level remains robust despite these data constraints. To test how much our model is better than a randomly guessing model, we also incorporated null tests. Additionally, we applied uncertainty analyses among all the models to distinguish different levels of mechanization instead of relying on a fixed classification to capture the inherent uncertainty as comprehensively as possible.

5 Conclusion

This study provides a major update to our understanding of global agricultural mechanization. Because the available data so far has been considerably outdated and it was not harmonized in any way, we find that actually, many countries around the world are likely far more mechanized than reported and widely assumed so far. This information offers policymakers valuable insights to target specific regions and devise tailored intervention strategies—particularly in developing countries where mechanization could substantially improve productivity and close yield gaps. While recognizing data limitations and uncertainties, our findings lay the foundation for a more precise understanding of agricultural mechanization's potential role in sustainable agricultural development. Our global map on agricultural mechanization developed in this study is an invaluable tool for policymakers, researchers, and development agencies aiming to pinpoint areas where mechanization could be a game-changer for food security and sustainable development. As mechanization continues to shape the future of agriculture, our work lays a foundation for more precise strategies that can drive transformative change in agricultural systems worldwide. Beyond this particular use case, we suggest that our here presented approach will be useful to model many other variables relevant for agricultural economics research. We provide a full replication package with this study, to enable and simplify follow-up studies.

References

Amatulli, Giuseppe, Daniel McInerney, Tushar Sethi, Peter Strobl, and Sami Domisch. 2020. "Geomorpho90m, Empirical Evaluation and Accuracy Assessment of Global High-Resolution Geomorphometric Layers." *Scientific Data* 7 (1). Nature Publishing Group: 162. doi:10.1038/s41597-020-0479-6.

Anwar, Muhuddin Rajin, De Li Liu, Robert Farquharson, Ian Macadam, Amir Abadi, John Finlayson, Bin Wang, and Thiagarajah Ramilan. 2015. "Climate Change Impacts on Phenology and Yields of Five Broadacre Crops at Four Climatologically Distinct Locations in Australia." *Agricultural Systems* 132 (January): 133–44. doi:10.1016/j.agsy.2014.09.010.

Baudron, Frédéric, Michael Misiko, Bisrat Getnet, Raymond Nazare, John Sariah, and Pascal Kaumbutho. 2019. "A Farm-Level Assessment of Labor and Mechanization in Eastern and Southern Africa." *Agronomy for Sustainable Development* 39 (2): 17. doi:10.1007/s13593-019-0563-5.

Caunedo, Julieta, and Namrata Kala. 2021. "Mechanizing Agriculture." Working Paper. Working Paper Series. National Bureau of Economic Research. doi:10.3386/w29061.

Center for International Earth Science Information Network - CIESIN - Columbia University. 2018. *Gridded Population of the World, Version 4 (GPWv4): Basic Demographic Characteristics, Revision 11*. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H46M34XX.

Center For International Earth Science Information Network-CIESIN-Columbia University. 2018. "Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets." Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). doi:10.7927/H45Q4T5F.

Dileepkumar, G. 2013. "Appropriate Knowledge and Tools for Smallholder Farmers." *New Agriculturist*. WRENmedia, 1–3.

Emami, Mohammad, Morteza Almassi, Hossein Bakhoda, and Issa Kalantari. 2018. "Agricultural Mechanization, a Key to Food Security in Developing Countries: Strategy Formulating for Iran." *Agriculture & Food Security* 7 (1): 24. doi:10.1186/s40066-018-0176-2.

FAO. 2017. *The Future of Food and Agriculture: Trends and Challenges*. Rome: Food and Agriculture Organization of the United Nations.

Fick, Steve, and Robert Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37 (May). doi:10.1002/joc.5086.

Foster, Andrew D., and Mark R. Rosenzweig. 2010. "Microeconomics of Technology Adoption." *Annual Review of Economics* 2 (Volume 2, 2010). Annual Reviews: 395–424. doi:10.1146/annurev.economics.102308.124433.

Gerber, James S., Deepak K. Ray, David Makowski, Ethan E. Butler, Nathaniel D. Mueller, Paul C. West, Justin A. Johnson, et al. 2024. "Global Spatially Explicit Yield Gap Time Trends Reveal Regions at Risk of Future Crop Yield Stagnation." *Nature Food* 5 (2). Nature Publishing Group: 125–35. doi:10.1038/s43016-023-00913-8.

Hagedorn, Konrad. 2004. "Property Rights Reform on Agricultural Land in Central and Eastern Europe." *Quarterly Journal of International Agriculture* 43 (4): 409–38.

Hamilton, Stephen F., Timothy J. Richards, Aric P. Shafran, and Kathryn N. Vasilaky. 2022. "Farm Labor Productivity and the Impact of Mechanization." *American Journal of Agricultural Economics* 104 (4): 1435–59. doi:10.1111/ajae.12273.

Hiroyuki, Takeshima. 2016. "Effects of Agricultural Mechanization on Smallholders and Their Self-Selection into Farming: An Insight from the Nepal Terai." *Research Papers in Economics*, January. International Food Policy Research Institute (IFPRI). https://typeset.io/papers/effects-of-agricultural-mechanization-on-smallholders-and-5ewg3gg1sb.

Ittersum, Martin K. van, Kenneth G. Cassman, Patricio Grassini, Joost Wolf, Pablo Tittonell, and Zvi Hochman. 2013. "Yield Gap Analysis with Local to Global Relevance—A Review." *Field Crops Research*, Crop Yield Gap Analysis – Rationale, Methods and Applications, 143 (March): 4–17. doi:10.1016/j.fcr.2012.09.009.

Jain, Meha. 2020. "The Benefits and Pitfalls of Using Satellite Data for Causal Inference." *Review of Environmental Economics and Policy* 14 (January): 157–69. doi:10.1093/reep/rez023.

Kamau, Hannah, Shahrear Roman, and Lisa Biber-Freudenberger. 2023. "Nearly Half of the World Is Suitable for Diversified Farming for Sustainable Intensification." *Communications Earth & Environment* 4 (1). Nature Publishing Group: 1–9. doi:10.1038/s43247-023-01062-3.

Kummu, Matti, Maija Taka, and Joseph H. A. Guillaume. 2018. “Gridded Global Datasets for Gross Domestic Product and Human Development Index over 1990-2015.” *Scientific Data* 5 (February): 180004. doi:10.1038/sdata.2018.4.

Lal, Rattan. 2018. “Digging Deeper: A Holistic Perspective of Factors Affecting Soil Organic Carbon Sequestration in Agroecosystems.” *Global Change Biology* 24 (8): 3285–3301. doi:10.1111/gcb.14054.

Léautier, Frannie, and Kobena Hanson. 2013. *Jobless Economic Growth: Lessons from Africa.* doi:10.13140/2.1.4702.4325.

Lesiv, Myroslava, Juan Carlos Laso Bayas, Linda See, Martina Duerauer, Domian Dahlia, Neal Durando, Rubul Hazarika, et al. 2019. “Estimating the Global Distribution of Field Size Using Crowdsourcing.” *Global Change Biology* 25 (1): 174–86. doi:10.1111/gcb.14492.

Lu, Qinan, Xiaodong Du, and Huanguang Qiu. 2022. “Adoption Patterns and Productivity Impacts of Agricultural Mechanization Services.” *Agricultural Economics* 53 (5): 826–45. doi:10.1111/agec.12737.

Meng, Meng, Wuke Zhang, Xi Zhu, and Qinghua Shi. 2024. “Agricultural Mechanization and Rural Worker Mobility: Evidence from the Agricultural Machinery Purchase Subsidies Programme in China.” *Economic Modelling* 139 (C). Elsevier. https://ideas.repec.org//a/eee/ecmode/v139y2024ics0264999324001408.html.

Mvodo, Elise Stephanie Meyo, and Dapeng Liang. 2012. “Cassava Sector Development in Cameroon: Production and Marketing Factors Affecting Price.” *Agricultural Sciences* 3 (5). Scientific Research Publishing: 651–57. doi:10.4236/as.2012.35078.

Pingali, Prabhu. 2007. “Chapter 54 Agricultural Mechanization: Adoption Patterns and Economic Impact.” In *Handbook of Agricultural Economics*, edited by R. Evenson and P. Pingali, 3:2779–2805. Agricultural Development: Farmers, Farm Production and Farm Markets. Elsevier. doi:10.1016/S1574-0072(06)03054-4.

Poggio, Laura, Luís De Sousa, Niels Batjes, Gerard Heuvelink, Bas Kempen, Eloi Ribeiro, and David Rossiter. 2021. “SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty.” *SOIL* 7 (June): 217–40. doi:10.5194/soil-7-217-2021.

Potapov, Peter, Svetlana Turubanova, Matthew C. Hansen, Alexandra Tyukavina, Viviana Zalles, Ahmad Khan, Xiao-Peng Song, Amy Pickens, Quan Shen, and Jocelyn Cortez. 2022. “Global Maps of Cropland Extent and Change Show Accelerated Cropland Expansion in the Twenty-First Century.” *Nature Food* 3 (1). Nature Publishing Group: 19–28. doi:10.1038/s43016-021-00429-z.

Rasouli, F, Hassan Sadighi, and Saeed Minaee. 2009. "Factors Affecting Agricultural Mechanization: A Case Study on Sunflower Seed Farms in Iran." *Journal of Agricultural Science and Technology* 11 (January).

Ricker-Gilbert, Jacob, Charles Jumbe, and Jordan Chamberlin. 2014. "How Does Population Density Influence Agricultural Intensification and Productivity? Evidence from Malawi." *Food Policy*, Boserup and Beyond: Mounting Land Pressures and Development Strategies in Africa, 48 (October): 114–28. doi:10.1016/j.foodpol.2014.02.006.

Shang, Linmei, Thomas Heckelei, Maria K. Gerullis, Jan Börner, and Sebastian Rasch. 2021. "Adoption and Diffusion of Digital Farming Technologies - Integrating Farm-Level Evidence and System Interaction." *Agricultural Systems* 190 (May): 103074. doi:10.1016/j.agsy.2021.103074.

Siebert, S, P Doll, and J Hoogeveen. 2005. "Development and Validation of the Global Map of Irrigation Areas." *Hydrology and Earth System Sciences*.

The Heritage Foundation. 2024. "Index of Economic Freedom: All Country Scores | The Heritage Foundation." *Index of Economic Freedom | The Heritage Foundation*. https://www.heritage.org/index.

Weiss, D. J., A. Nelson, H. S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, et al. 2018. "A Global Map of Travel Time to Cities to Assess Inequalities in Accessibility in 2015." *Nature* 553 (7688). Nature Publishing Group: 333–36. doi:10.1038/nature25181.

WOCAT. 2024. "The Global Database on Sustainable Land Management (WOCAT)." *Wocat.Net*. January 7. https://qcat.wocat.net/en/wocat/.

World Bank. 2024. "Agricultural Machinery, Tractors per 100 Sq. Km of Arable Land." *World Bank Open Data*. https://data.worldbank.org.

World Bank. 2024. "Fertilizer Consumption (Kilograms per Hectare of Arable Land)." *World Bank Open Data*. https://data.worldbank.org/indicator/AG.CON.FERT.ZS.

World Bank Group. 2017. *Enabling the Business of Agriculture 2017*. World Bank, Washington, DC. doi:10.1596/978-1-4648-1021-3.

Wuepper, David, Pasquale Borrelli, Panos Panagos, Thomas Lauber, Thomas Crowther, Amy Thomas, and David A. Robinson. 2021. "A 'Debt' Based Approach to Land Degradation as an Indicator of Global Change." *Global Change Biology* 27 (21): 5407–10. doi:10.1111/gcb.15830.

Wuepper, David, Haoyu Wang, Wolfram Schlenker, Meha Jain, and Robert Finger. 2023. "Institutions and Global Crop Yields." *NBER Working Papers*, NBER Working Papers, , July. National Bureau of Economic Research, Inc. https://ideas.repec.org//p/nbr/nberwo/31426.html.

Wuepper, David, Ilsabe Wiebecke, Lara Meier, Sarah Vogelsanger, Selina Bramato, Andrea Fürholz, and Robert Finger. 2024. "Agri-Environmental Policies from 1960 to 2022." *Nature Food* 5 (4). Nature Publishing Group: 323–31. doi:10.1038/s43016-024-00945-8.

Xinshen, Diao, Silver Jed, and Takeshima Hiroyuki. 2016. *Agricultural Mechanization and Agricultural Transformation*. Intl Food Policy Res Inst.

Yan, Fuzhen, Xuetao Sun, Suisui Chen, and Guilin Dai. 2024. "Does Agricultural Mechanization Improve Agricultural Environmental Efficiency?" *Frontiers in Environmental Science* 11 (January). Frontiers. doi:10.3389/fenvs.2023.1344903.

Yu, Qiangyi, Liangzhi You, Ulrike Wood-Sichra, Yating Ru, Alison K. B. Joglekar, Steffen Fritz, Wei Xiong, Miao Lu, Wenbin Wu, and Peng Yang. 2020. "A Cultivated Planet in 2010 – Part 2: The Global Gridded Agricultural-Production Maps." *Earth System Science Data* 12 (4). Copernicus GmbH: 3545–72. doi:10.5194/essd-12-3545-2020.

SUPPLEMENTARY MATERIALS

Agricultural Mechanization Around the World

Shahrear Roman[1*], Hadi[1], David Wuepper[1]

[1] Land Economics Group, University of Bonn, Germany

*Corresponding author: s.roman@ilr.uni-bonn.de

25.11.2024

A1. Appendix A1. Training data

The mechanization data for this analysis comes from two main sources. The first is the Global Database on Sustainable Land Management (SLM) by WOCAT (World Overview of Conservation Approaches and Technologies). This database serves as a vital resource for understanding sustainable practices across various countries. The SLM data highlights sustainable land management techniques aimed at enhancing long-term productivity while preserving ecosystem functions. It includes information on soil and water conservation practices, as well as strategies to prevent land degradation, drawing from over 2,446 documented SLM practices in 136 countries. The WOCAT database is essential for providing baseline data on sustainable agricultural practices.

To complement the WOCAT data, we also integrated information from the Economics of Agricultural Technology Database, a newly completed survey done by University of Bonn, Germany. This dataset focuses on the economic aspects of agricultural innovation, including the adoption and profitability of mechanization and other technologies. By surveying experts from academia, industry, extension services, and the farming community, this database provides insights into the profitability of digital mechanization practices, including specific locations. The combination of these data sources ensures our analysis reflects a diverse range of agricultural contexts and innovations, addressing gaps left by the WOCAT dataset and facilitating a more comprehensive evaluation.

A2. Pearson correlation test

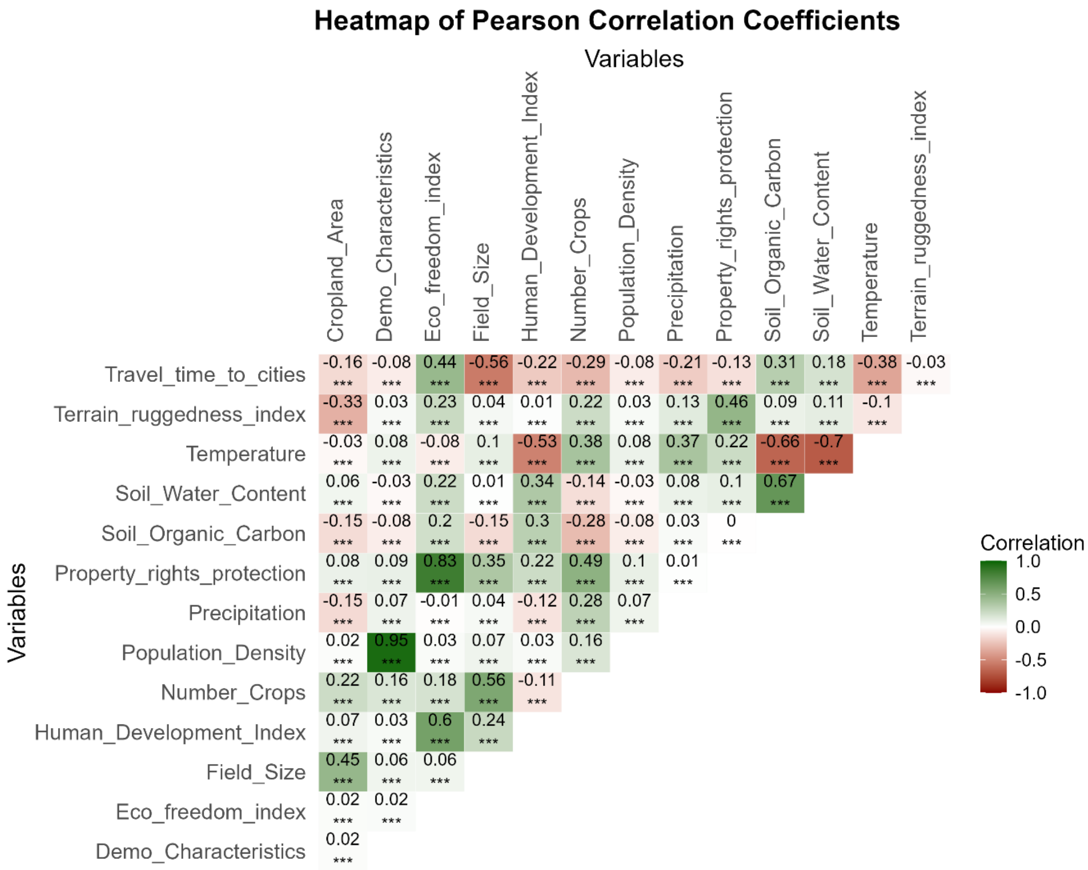

Figure A2.1: Heatmap of Pearson correlation coefficients between predictor variables. Note: Significance levels: p≤ 0.001 ***, p ≤0.01 **, p≤ 0.05*, p≤ 0.10.

A3. Appendix A2. Multivariate Environmental Similarity Surface (MESS) analysis

The Multivariate Environmental Similarity Surface (MESS) analysis evaluates the reliability of model predictions in various environmental conditions by comparing these conditions to those present in the training data. Positive MESS values indicate that the environmental conditions at a location fall within the range observed during model training, suggesting that predictions in these areas are more dependable. Higher positive values indicate a closer match to the average training conditions, which boosts confidence in the predictions.

Conversely, negative MESS values indicate that at least one environmental variable at a location falls outside the training data range, implying that the model must extrapolate. As a result, predictions in these cases should be approached with caution. A MESS value of 0 represents the boundary of environmental similarity, marking conditions that are on the edge of the training data. This analysis helps identify regions where model predictions are robust versus those where they are less certain.

The MESS plot provides a visual representation of this similarity across global environmental conditions. The color gradient ranges from light shades (yellow to light pink), indicating high similarity to the training data, to dark shades (purple to black), which show significant divergence from the training conditions. The locations of training points are marked with black cross markers, making it easy to identify regions that closely align with the model's foundational data, as well as areas where environmental conditions are different. This map serves as a valuable tool for understanding where model predictions are based on interpolation versus extrapolation, enhancing confidence in areas with higher environmental similarity.

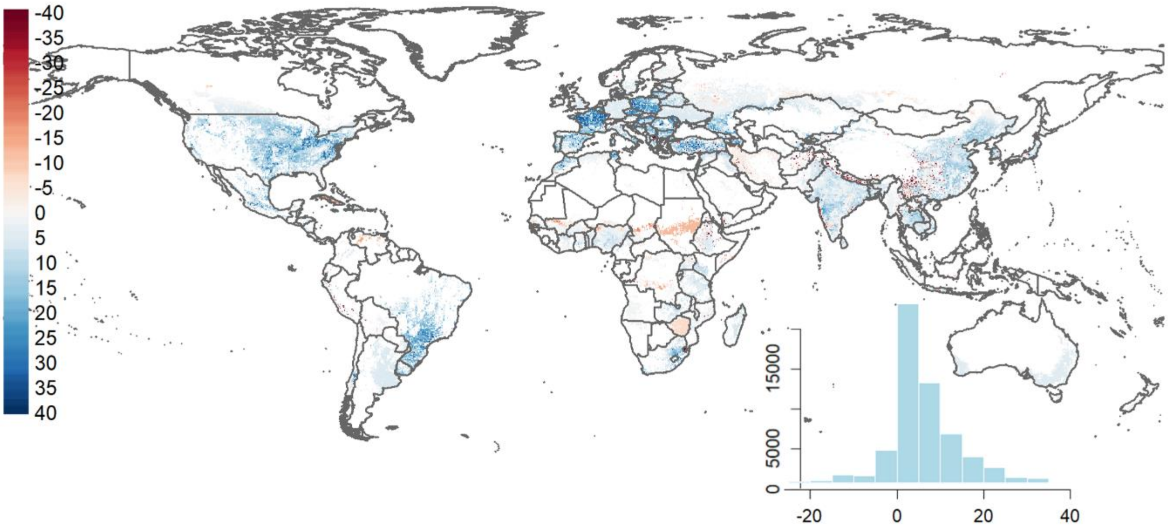

Figure A3.1: The map illustrates similarity across various regions, generated using the MESS (Multivariate Environmental Similarity Surface).

A4. Appendix A3. Agricultural Mechanization at Regional Level

Table A4.1: Average, minimum and maximum rate of agricultural mechanization at regional level.

REGION	AVERAGE	MIN	MAX	PERCENTAGE > 0.50
WESTERN EUROPE	0.68	0.44	0.86	92.44%
NORTH AMERICA	0.63	0.44	0.82	84.88%
OCEANIA	0.56	0.40	0.76	68
EAST EUROPE	0.53	0.33	0.81	54.37%
WESTERN ASIA	0.53	0.31	0.77	55.9%
CENTRAL AMERICA AND CARIBBEAN ISLAND	0.51	0.24	0.79	49.93%
NORTHERN WESTERN AFRICA	0.50	0.25	0.75	48.17%
CENTRAL ASIA	0.50	0.31	0.78	41.23%
SOUTH ASIA	0.48	0.25	0.67	48.3%
SOUTHERN AFRICA	0.48	0.31	0.69	41.1%
EASTERN ASIA	0.45	0.25	0.74	35.59%
SOUTH AMERICA	0.45	0.24	0.69	32.59%
SUB-SAHARAN, MIDDLE AFRICAN GROUP	0.37	0.20	0.64	15.7%

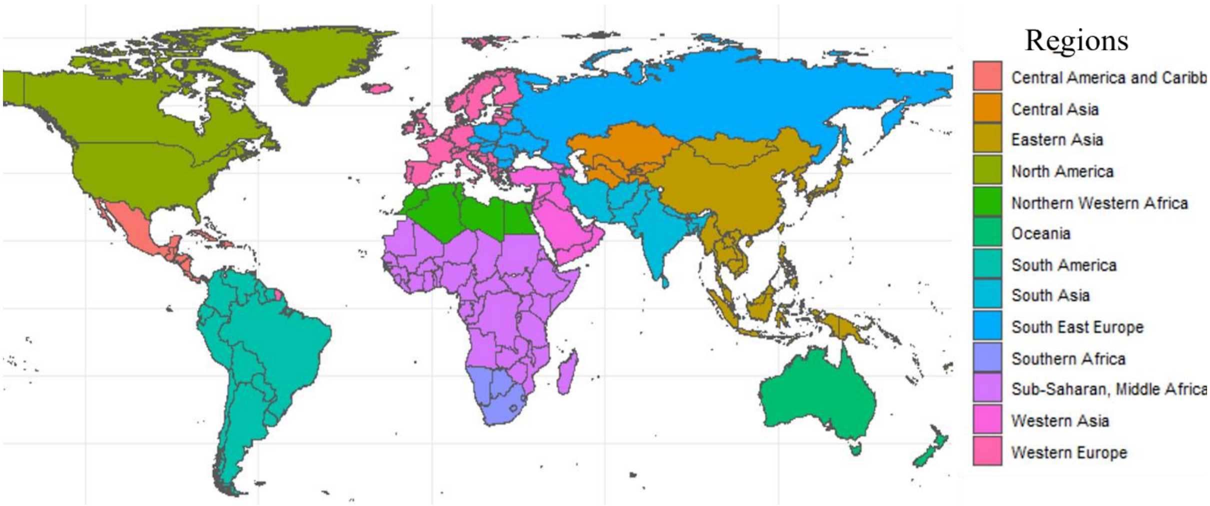

Figure A4.2: Map of the regions which is explained in the Table A3.1

A5. Appendix A4. Agricultural Mechanization at Country Level

Table A4.1 presents a detailed summary of mechanization data across 172 countries, including mean, minimum, maximum, standard deviation, and range metrics. The NAME_EN column lists countries according to the World Bank nomenclature, while the Mean value reflects each country's average mechanization level across pixels. The Median value signifies the midpoint of mechanization levels arranged in order, and the Minimum and Maximum values indicate the lowest and highest recorded mechanization levels for each country, respectively. SD (Standard Deviation) quantifies the variability of mechanization values around the mean, while the Range illustrates the difference between maximum and minimum values. The N column represents each country's total number of observations (i.e., the 5-km grid cells), with "N above 0.50" column, counting observations that surpass a mechanization level of 0.50. Finally, Mechanization (N) and Mechanization (%) indicate the proportion of observations and the overall percentage of mechanization exceeding 0.50 within each country.

The very first column shows the name of the country (according to World Bank data). The Second and third columns show the mean and median of agricultural mechanization for each country with minimum (column 4) and maximum (column 5). We also calculate the standard deviation (column 6) and range (column 7). In column 8, we showed the number of observations per country (pixel 5km*5km) per country alongside the number of pixel's agricultural mechanization over 0.50 (column 9) with percentage (column 10).

Table A5.1: Statistics of mechanization at the country level.

Country	Mean	Median	Min	Maxi	Sd	Range
Afghanistan	0,39	0,36	0,23	0,75	0,1	0,51
Albania	0,58	0,57	0,35	0,83	0,12	0,47
Algeria	0,42	0,41	0,23	0,8	0,12	0,57
Andorra	0,4	0,3	0,29	0,63	0,19	0,34
Angola	0,33	0,3	0,21	0,68	0,07	0,48
Antigua And Barbuda	0,23	0,26	0,14	0,31	0,07	0,17
Argentina	0,45	0,45	0,2	0,75	0,1	0,55
Armenia	0,61	0,61	0,24	0,86	0,12	0,62
Australia	0,56	0,58	0,28	0,89	0,1	0,61
Austria	0,7	0,71	0,43	0,88	0,1	0,45
Azerbaijan	0,61	0,62	0,24	0,86	0,13	0,62
Bahrain	0,62	0,64	0,52	0,69	0,09	0,17
Bangladesh	0,58	0,6	0,22	0,69	0,09	0,46
Barbados	0,68	0,69	0,54	0,79	0,07	0,25
Belarus	0,49	0,49	0,27	0,85	0,09	0,58
Belgium	0,77	0,81	0,51	0,88	0,09	0,37

Belize	0,44	0,43	0,31	0,69	0,07	0,37
Benin	0,47	0,45	0,27	0,74	0,1	0,47
Bhutan	0,37	0,34	0,27	0,66	0,07	0,4
Bolivia	0,29	0,29	0,2	0,71	0,07	0,51
Bosnia And Herzegovina	0,59	0,58	0,37	0,86	0,1	0,49
Botswana	0,49	0,48	0,2	0,74	0,08	0,54
Brazil	0,44	0,43	0,2	0,76	0,11	0,56
Brunei	0,51	0,51	0,39	0,63	0,17	0,24
Bulgaria	0,63	0,62	0,35	0,87	0,1	0,52
Burkina Faso	0,4	0,4	0,26	0,72	0,08	0,46
Burundi	0,37	0,36	0,19	0,66	0,09	0,47
Cambodia	0,48	0,5	0,24	0,75	0,12	0,51
Cameroon	0,38	0,37	0,19	0,69	0,1	0,5
Canada	0,62	0,61	0,36	0,88	0,1	0,52
Central African Republic	0,24	0,22	0,17	0,64	0,06	0,48
Chad	0,33	0,31	0,17	0,64	0,07	0,47
Chile	0,64	0,63	0,31	0,9	0,11	0,59
Colombia	0,5	0,49	0,17	0,82	0,11	0,65
Costa Rica	0,52	0,51	0,28	0,78	0,1	0,5
Croatia	0,67	0,67	0,37	0,86	0,1	0,49
Cuba	0,28	0,28	0,15	0,48	0,07	0,33
Cyprus	0,68	0,69	0,41	0,86	0,11	0,44
Czech Republic	0,75	0,76	0,45	0,89	0,08	0,44
Democratic Republic Of The Congo	0,28	0,25	0,17	0,66	0,08	0,49
Denmark	0,76	0,76	0,41	0,89	0,07	0,48
Dominican Republic	0,56	0,56	0,26	0,81	0,11	0,55
East Timor	0,32	0,29	0,21	0,64	0,09	0,43
Ecuador	0,48	0,48	0,27	0,76	0,1	0,49
Egypt	0,57	0,61	0,29	0,74	0,14	0,45
El Salvador	0,5	0,5	0,29	0,73	0,09	0,44
Eritrea	0,28	0,24	0,17	0,63	0,09	0,46
Estonia	0,6	0,6	0,36	0,88	0,11	0,52
Eswatini	0,24	0,23	0,14	0,64	0,07	0,5
Ethiopia	0,36	0,34	0,15	0,65	0,1	0,5
Finland	0,6	0,58	0,16	0,89	0,11	0,73
France	0,72	0,73	0,14	0,88	0,09	0,74
Gabon	0,41	0,38	0,29	0,57	0,12	0,28
Georgia	0,59	0,59	0,35	0,86	0,1	0,51
Germany	0,76	0,77	0,41	0,89	0,08	0,48

Ghana	0,44	0,43	0,27	0,73	0,1	0,46
Greece	0,57	0,57	0,32	0,83	0,11	0,5
Guantanamo Bay Naval Base	0,33	0,33	0,24	0,42	0,13	0,18
Guatemala	0,54	0,54	0,3	0,8	0,11	0,5
Guinea	0,32	0,3	0,2	0,62	0,07	0,42
Guinea-Bissau	0,25	0,23	0,19	0,55	0,06	0,35
Guyana	0,45	0,45	0,27	0,75	0,12	0,48
Haiti	0,41	0,41	0,23	0,63	0,08	0,4
Honduras	0,48	0,47	0,3	0,76	0,1	0,45
Hong Kong	0,27	0,27	0,27	0,27	Na	0
Hungary	0,73	0,74	0,43	0,88	0,08	0,45
Iceland	0,45	0,43	0,36	0,71	0,05	0,34
India	0,51	0,52	0,23	0,75	0,11	0,52
Indonesia	0,56	0,55	0,25	0,8	0,14	0,55
Iran	0,35	0,33	0,18	0,85	0,1	0,67
Iraq	0,51	0,5	0,22	0,8	0,12	0,58
Ireland	0,64	0,62	0,37	0,88	0,09	0,5
Israel	0,71	0,76	0,19	0,89	0,14	0,7
Italy	0,7	0,72	0,37	0,87	0,09	0,5
Ivory Coast	0,4	0,38	0,24	0,74	0,08	0,5
Jamaica	0,57	0,58	0,31	0,77	0,11	0,46
Japan	0,65	0,65	0,35	0,86	0,1	0,52
Jordan	0,59	0,6	0,21	0,84	0,13	0,63
Kazakhstan	0,49	0,45	0,25	0,87	0,1	0,62
Kenya	0,47	0,46	0,24	0,75	0,13	0,51
Kosovo	0,68	0,7	0,37	0,86	0,11	0,48
Kuwait	0,51	0,5	0,37	0,76	0,1	0,4
Kyrgyzstan	0,53	0,51	0,3	0,85	0,13	0,55
Laos	0,34	0,31	0,22	0,76	0,1	0,54
Latvia	0,63	0,63	0,31	0,87	0,1	0,57
Lebanon	0,44	0,43	0,25	0,82	0,09	0,57
Lesotho	0,43	0,41	0,27	0,72	0,11	0,45
Liberia	0,38	0,41	0,27	0,44	0,06	0,18
Libya	0,42	0,39	0,27	0,73	0,12	0,46
Liechtenstein	0,73	0,72	0,66	0,83	0,05	0,17
Lithuania	0,68	0,69	0,31	0,88	0,09	0,57
Luxembourg	0,76	0,76	0,61	0,87	0,06	0,26
Madagascar	0,38	0,34	0,28	0,76	0,09	0,47
Malawi	0,48	0,48	0,24	0,73	0,12	0,5

Malaysia	0,59	0,6	0,34	0,83	0,12	0,49
Mali	0,37	0,36	0,24	0,69	0,08	0,45
Mauritania	0,33	0,3	0,25	0,71	0,08	0,47
Mexico	0,54	0,52	0,27	0,87	0,13	0,61
Moldova	0,66	0,66	0,4	0,86	0,08	0,46
Mongolia	0,45	0,43	0,29	0,77	0,07	0,47
Montenegro	0,56	0,56	0,35	0,75	0,09	0,4
Morocco	0,57	0,57	0,25	0,83	0,12	0,58
Mozambique	0,35	0,32	0,14	0,7	0,09	0,56
Myanmar	0,35	0,33	0,21	0,75	0,11	0,55
Namibia	0,4	0,37	0,26	0,77	0,07	0,51
Nepal	0,41	0,38	0,24	0,66	0,13	0,42
Netherlands	0,78	0,79	0,42	0,88	0,07	0,46
New Zealand	0,55	0,55	0,36	0,87	0,1	0,51
Nicaragua	0,44	0,43	0,26	0,72	0,1	0,45
Niger	0,41	0,41	0,26	0,73	0,09	0,46
Nigeria	0,46	0,45	0,25	0,71	0,1	0,46
North Korea	0,52	0,51	0,25	0,83	0,16	0,58
Norway	0,57	0,54	0,33	0,88	0,13	0,55
Oman	0,53	0,51	0,37	0,78	0,1	0,41
Pakistan	0,47	0,49	0,23	0,7	0,13	0,47
Palestine	0,39	0,33	0,17	0,85	0,16	0,68
Panama	0,52	0,52	0,34	0,81	0,11	0,47
Papua New Guinea	0,34	0,33	0,25	0,63	0,08	0,38
Paraguay	0,47	0,47	0,22	0,78	0,09	0,55
People's Republic Of China	0,42	0,41	0,22	0,83	0,12	0,61
Peru	0,55	0,52	0,23	0,89	0,13	0,66
Philippines	0,52	0,53	0,27	0,75	0,13	0,48
Poland	0,74	0,75	0,34	0,88	0,08	0,54
Portugal	0,59	0,59	0,4	0,88	0,1	0,48
Qatar	0,56	0,56	0,43	0,72	0,07	0,29
Republic Of Macedonia	0,57	0,56	0,32	0,78	0,11	0,46
Republic Of The Congo	0,27	0,24	0,21	0,57	0,06	0,36
Romania	0,66	0,66	0,4	0,87	0,1	0,48
Russia	0,47	0,47	0,24	0,84	0,1	0,61
Rwanda	0,54	0,55	0,2	0,67	0,1	0,47
San Marino	0,6	0,63	0,35	0,79	0,16	0,44
Saudi Arabia	0,48	0,46	0,3	0,82	0,09	0,52
Senegal	0,43	0,43	0,2	0,76	0,11	0,55

Serbia	0,67	0,69	0,4	0,87	0,1	0,47
Sierra Leone	0,37	0,37	0,23	0,62	0,08	0,39
Singapore	0,64	0,64	0,63	0,64	0,01	0,02
Slovakia	0,71	0,72	0,46	0,88	0,09	0,42
Slovenia	0,65	0,64	0,47	0,83	0,09	0,36
Somalia	0,34	0,33	0,21	0,62	0,09	0,42
South Africa	0,5	0,49	0,14	0,81	0,1	0,67
South Korea	0,62	0,62	0,37	0,83	0,11	0,46
South Sudan	0,21	0,19	0,16	0,63	0,05	0,47
Spain	0,66	0,67	0,38	0,89	0,11	0,51
Sri Lanka	0,45	0,46	0,25	0,66	0,1	0,41
Sudan	0,26	0,26	0,15	0,64	0,06	0,49
Suriname	0,34	0,32	0,22	0,58	0,09	0,36
Sweden	0,64	0,62	0,4	0,89	0,11	0,49
Switzerland	0,73	0,74	0,41	0,88	0,08	0,47
Syria	0,49	0,49	0,21	0,84	0,11	0,63
Tajikistan	0,46	0,44	0,28	0,8	0,14	0,53
Tanzania	0,47	0,47	0,19	0,8	0,11	0,61
Thailand	0,58	0,6	0,22	0,79	0,11	0,57
The Bahamas	0,42	0,37	0,35	0,61	0,09	0,26
The Gambia	0,49	0,49	0,29	0,73	0,1	0,44
Togo	0,42	0,41	0,26	0,72	0,09	0,46
Trinidad And Tobago	0,62	0,59	0,53	0,72	0,07	0,18
Tunisia	0,54	0,55	0,24	0,77	0,1	0,53
Turkey	0,54	0,54	0,23	0,83	0,12	0,6
Turkmenistan	0,43	0,42	0,22	0,82	0,11	0,6
Uganda	0,43	0,44	0,17	0,69	0,1	0,52
Ukraine	0,6	0,6	0,29	0,85	0,09	0,55
United Arab Emirates	0,54	0,54	0,38	0,73	0,08	0,34
United Kingdom	0,74	0,78	0,37	0,89	0,11	0,52
United States Of America	0,63	0,63	0,34	0,9	0,11	0,57
Uruguay	0,52	0,51	0,22	0,85	0,11	0,62
Uzbekistan	0,59	0,6	0,27	0,84	0,14	0,57
Vanuatu	0,47	0,47	0,42	0,52	0,07	0,1
Venezuela	0,24	0,22	0,16	0,71	0,07	0,56
Vietnam	0,55	0,54	0,23	0,79	0,13	0,56
Yemen	0,37	0,35	0,24	0,69	0,1	0,45
Zambia	0,32	0,3	0,19	0,75	0,08	0,56
Zimbabwe	0,28	0,28	0,18	0,66	0,07	0,48