

*Full Length Research Paper*

## **Rural public policies and the state of smallholders: Recent evidence from Brazil**

**Gabriel Paes Herrera<sup>1\*</sup>, Fernando Silva Bernardes<sup>2</sup>, Reginaldo Brito da Costa<sup>1</sup>, Benedito Albuquerque da Silva<sup>1</sup>, Dany Rafael Fonseca Mendes<sup>3</sup> and Michel Constantino<sup>1</sup>**

<sup>1</sup>Department of Environmental Sciences and Sustainable Agro-Livestock Technologies, Catholic University Dom Bosco, Campo Grande, Mato Grosso do Sul, Brazil.

<sup>2</sup>Department of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul, Brazil.

<sup>3</sup>Department of Law, University Center of Brasília, Brasília, Distrito Federal, Brazil.

Received 23 May, 2018; Accepted 24 July, 2018

**The aim of this article is to characterize the current situation of family farmers or smallholders in Brazil and establish a connection with the rural public policies that exist in the country. This study analyzed the most current available data regarding family farming in Brazil, which included almost 4.7 million smallholders and their characteristics. Two analytical tools for unsupervised learning were combined, Principal Component Analysis (PCA) and K-means clustering, which enabled the analysis of such a large database and the extraction of information concerning this sector. It was found that cooperative smallholders are considerably more likely to achieve higher incomes. A family farmer's income and productivity are related to their region and are higher in the South and Southeast and lower in the Northeast region. Crop diversification presented a negative impact on family farming activity, although this practice is considered highly important for agricultural sustainability. These results confirm, based on the data, empirical findings regarding the sector and also reveal new information such as the negative impact that rural assistance services are demonstrated to have on smallholders' income. Therefore, this study provides essential information to support policy makers in the process of formulating better and more efficient policies in order to strengthen smallholders in Brazil and guarantee food security in the future.**

**Key words:** Family farm, rural programs, unsupervised learning, government support.

### **INTRODUCTION**

According to the Food and Agriculture Organization of the United Nations - FAO (2014), by 2050, there will be approximately 9.6 billion people in the world, and food production will have to increase by 60% to meet this new demand, thereby placing more pressure on natural

resources that are already scarce and showing signs of more food, but production must be undertaken with sustainability.

In this context, family farmers, also known as smallholders, are considered part of the solution for

\*Corresponding author. E-mail: [gabrielherrera27@hotmail.com](mailto:gabrielherrera27@hotmail.com).

achieving food security and sustainable rural development (FAO, 2014). Smallholding is the prevalent agricultural arrangement, as almost 90% of farms, or approximately 500 million farms in the world, are owned and operated by families. These farmers occupy more than half of the total agricultural land and produce at least 53% of the world's food (Graeub et al., 2016; Lowder et al., 2014). The efficiency of smallholder farming relative to larger farms has been widely documented, and these farmers are capable of achieving high production levels per unit of land through the use of family labor in diversified production systems (Bosc et al., 2013).

Brazil plays a decisive role in the agricultural international market which is among the ten largest economies in the world. As it has the fifth-largest surface area and favorable location and climate, the country became the largest supplier of sugar, orange juice and coffee (OECD/FAO, 2015). In Brazil, family farmers represent more than 80% of the production units and play an essential role in the domestic market food supply. In 2006, smallholders were responsible for 38% of the gross value of Brazilian agricultural production according to the Brazilian Institute of Geography and Statistics – IBGE (IBGE, 2006). Additionally, according to the IBGE, in Brazil, approximately 4.3 million rural units are owned by families, and more than 12 million people depend on this activity for their subsistence.

The size of this sector and the enormous amount of information that it contains creates a massive quantity of high dimensional data that needs to be managed and carefully explored. According to Chakraborty and Joseph (2017), the development of new analytical tools, such as machine learning techniques, enables us to untangle important information and patterns that would pass unnoted if conventional approaches were used. Therefore, it is essential that new studies rely on updated databases and analytical tools that can generate new insights and relevant information to support policy makers and important decisions. The two-step cluster methodology employed in this research has been used in similar studies worldwide and generated excellent results, with a few examples of such coming from Herrero et al. (2014), Gaspar et al. (2011) and Sepúlveda et al. (2010).

The former traditional agricultural system based on the massive use of agrochemicals and fertilizers is no longer accepted as the best one. Globalization, climate change and a general societal perception of the importance of natural resources has led farmers to rediscover efficient sustainable practices and also consumers to demand more environmental friendly products (de Roest et al., 2018).

The Brazilian government has a range of public policies targeting family farmers that aim to increase their incomes, welfare and reduce social inequality. In order to have access to these policies, smallholders need to maintain a register in the Ministry of Agrarian Development (MDA) by completing a declaration form,

known as the “DAP” (Declaration of Aptitude to Pronaf), and keep it updated. The present study is based on the information provided in these forms relative to millions of family farmers from every state of Brazil. Public policies can make a difference in the success or failure of an entire agricultural sector; therefore, studies to drive and point where investments should focus on are essential. Without public support and the correct investments, the only way to production growth would be through the expansion of agricultural land (Anang and Yanwen, 2014).

The remainder of this paper is structured as follows. First is a description of the database, techniques and methodology that was followed in this analysis. This is followed by a presentation of the results and discussion, and thereafter, the study's conclusions and outlined recommendations for further research and policy-making.

## MATERIALS AND METHODS

### Data source

The analyses conducted in this paper are based on the data from Brazilian family farmers. The database was obtained through the MDA in October 2014 and contains the most updated information about these farmers in Brazil. When filling the DAP form, smallholders provide detailed information regarding themselves and their farms, such as their age, gender, schooling, farm area, number of crops produced and total income. Therefore, the database creates a plentiful source of information about family farming in the country. Most studies about this sector in Brazil are based on the Agricultural Census data, which was last conducted in 2006 and can be easily accessed by everyone. Studies using data from the MDA are still scarce due to the restrictive bureaucracy involved in obtaining it. The database was refined by removing cases with missing values or highly distorted values (outliers) to minimize errors in the results. Approximately 3% (133,000 DAPs) were excluded, and the final database used for the analysis contained approximately 4.7 million declaration forms of family farmers from all states in Brazil. The most important variables were selected and are presented in Table 1. The variables include the age and schooling of the household head, the area in hectares and the state where the farm was located. Also included were the coefficient of production diversification, which was measured by Simpson's Diversity Index – SDI (Simpson, 1949), regardless of whether the farmer was part of a cooperative, whether the farmer received rural assistance and the farmer's income and productivity. The analyses were conducted using the software R Studio (R Core Team, 2017).

### Principal component analysis (PCA)

Currently, with the ever-growing massive quantity of high dimensional data, researchers have found some obstacles to performing certain analyses. Principal component analysis (PCA) is a statistical technique for unsupervised dimension reduction, which is closely related to unsupervised learning and is used in very broad areas, such as meteorology, image processing, genomic analysis and information retrieval (Ding and He, 2004; Bishop, 2006). As defined by Hotelling (1933), PCA is the orthogonal projection of the data onto a lower dimensional linear space such that the variance of the projected data is maximized. According to Howley et al.

**Table 1.** Variable descriptions and summary statistics.

Variable	Description	Mean	Std. Dev.
Age	Household head age in years	44.8358	15.2110
Area	Total area of the farm in hectares	19.0604	33.3236
State	Brazilian state code	27.8471	8.8080
Production diversification	Simpson's diversity index coefficient	0.3529	0.2821
Cooperative	Dummy (0 = no; 1 = yes)	0.0497	0.2175
Rural assistance	Dummy (0 = no; 1 = yes)	0.0768	0.2663
Income	Total annual on-farm income in BRL	18,404.13	37,667.88
Productivity	Annual productivity in BRL/ha	8,345.119	334,852.2
Schooling	Schooling ranked from 1 to 10*	3.3010	1.4858

\*1 – Illiterate, 2 – Literate, 3 – Elementary school incomplete, 4 – Elementary school complete, 5 – High school incomplete, 6 – High school complete, 7 – Certificate program incomplete, 8 – Certificate program complete, 9 – College incomplete, and 10 – College complete.

(2006), PCA enables us to transform the attributes of a dataset into a new set of uncorrelated attributes called principal components (PCs), thereby reducing the dimensionality of the original dataset while still retaining as much of the variability as possible. Each PC is a linear combination of the original inputs and each PC is orthogonal, which therefore eliminates the problem of collinearity. Thus, the PCA technique is commonly used to reduce high dimensional data, such as the one exploited in this paper, to enable a certain analysis. In addition, using this technique as a preprocessing step can improve the performance of machine learning techniques, especially in the classification of high dimensional data (He et al., 2011; Howley et al., 2006). As reported by Ding and He (2004), principal component analysis dimensional reduction is particularly beneficial for K-means clustering, thereby improving the cluster accuracy. This methodology proved to be very efficient in similar studies such as those conducted by Herrero et al. (2014), Gaspar et al. (2011) and Sepúlveda et al. (2010).

### K-means clustering

Machine learning techniques can be divided into supervised and unsupervised learning, with this last one being characterized by having no category labels that tag objects with prior identifiers, and as such, the algorithm merely aims to find structure in the data, which has to be interpreted by the researcher (Chakraborty and Joseph, 2017; Jain, 2010). Most unsupervised algorithms aim to group observations according to common patterns. According to Ding and He (2004) and Jain (2010), the K-means algorithm is one of the most commonly used clustering techniques for large scale data due to its easy implementation, simplicity, efficiency and empirical success. Likewise, MacQueen (1967) states that the K-means procedure is easily programmed and is computationally economical, thereby making it is feasible to obtain qualitative and quantitative understandings of large amounts of N-dimensional data. As defined by MacQueen (1967), the K-means procedure consists of simply starting with K groups, each of which consists of a single random point and thereafter assign each new point to its closest centroid. After a point is added to a group, the mean of that group is adjusted in order to take account of the new point. Thus, at each stage, the K-means are, in fact, the means of the groups that they represent. According to Bishop (2006), the goal, therefore, is to find an assignment of data points to clusters, as well as a set of vectors  $\{\mu_k\}$ , such that the sum of the squares of the distances of each data point to its closest vector  $\mu_k$  which is its minimum. Following Jain (2010), the corresponding function is defined as follows:

$$J(C) = \sum_{k=1}^K \sum_{\chi_i \in C_k} \|\chi_i - \mu_k\|^2$$

Clustering is used for knowledge discovery rather than prediction. It provides insight into the natural groupings found within data, resulting in meaningful and actionable data structures that reduce complexity (Lantz, 2013). As stated by Chakraborty and Joseph (2017), the K-means technique tries to minimize the differences within each cluster and maximize the differences between the clusters, thereby providing insights regarding the commonalities between observations.

## RESULTS AND DISCUSSION

Initially, the variables “income” and “productivity” were transformed into logarithms following the current literature, as outlined by Wooldridge (2015) and Venables and Ripley (2013). In order to conduct PCA analysis, the data was standardized, transforming all features into comparable numerical ranges, and the conventions for such an analysis can be found in Bishop (2006), Chakraborty and Joseph (2017) and Jolliffe (2002). The PCA results can be found in Table 2. According to the results, more than fifty percent of the total variance can be explained by the first three PCs. The first PC is shown to be positively related to the farmer’s income, productivity, state, cooperativism and schooling.

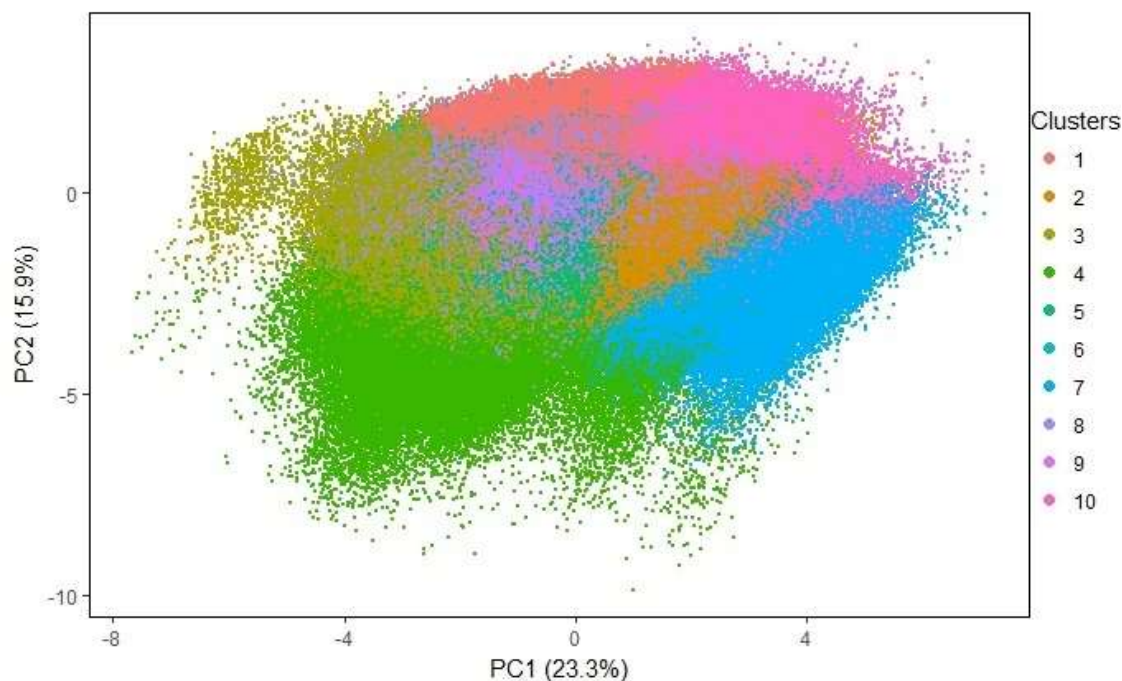
Subsequently, the k-means clustering technique was applied using the eigenvalues of each observation. A few tests were conducted with different combinations of numbers of PCs and numbers of clusters, seeking to identify the best sum of squares ratio. the first five principal components were decidedly kept, since these PCs together explain more than 75% of the data variance, and jointly with a set of ten clusters, attain a 67.4% sum of squares ratio. Figure 1 shows the scatter plot of smallholders divided into groups on the first two dimensions of PCA, making it possible to visualize the structures of some clusters.

**Table 2.** Importance of each principal component and the loadings of the variables.

Statistical variable	Importance of each principal component								
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Standard Deviation	1.4510	1.1996	1.1204	1.0101	0.9721	0.9192	0.8342	0.7270	0.4072
Proportion of Variance	0.2339	0.1599	0.1395	0.1134	0.1050	0.0938	0.0773	0.0587	0.0184
Cumulative Proportion	0.2339	0.3938	0.5333	0.6466	0.7516	0.8455	0.9228	0.9815	1.0000

Demographic variable	Loadings of the variable								
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
State	0.451	-0.218	0.050	-0.102	0.028	-0.204	-0.730	-0.388	0.089
Age	0.013	-0.547	0.486	0.160	-0.109	0.125	-0.174	0.613	0.060
Schooling	0.295	0.225	-0.610	-0.063	0.114	0.009	-0.250	0.638	0.047
Area	-0.147	-0.548	-0.532	0.125	-0.189	0.237	0.109	-0.196	0.484
Diversification	-0.118	-0.262	0.020	-0.356	0.878	0.103	0.070	-0.004	0.057
Cooperative	0.331	-0.284	-0.050	-0.135	-0.046	-0.745	0.470	0.088	0.035
Rural Assistance	-0.015	-0.032	0.050	-0.888	-0.399	0.210	0.024	0.056	0.001
Income	0.537	-0.248	-0.115	0.094	0.004	0.442	0.255	-0.124	-0.592
Productivity	0.523	0.296	0.294	0.037	0.072	0.281	0.262	-0.022	0.629

**Figure 1.** Scatter plot (Projection of smallholders divided by clusters on the first two dimensions of PCA).

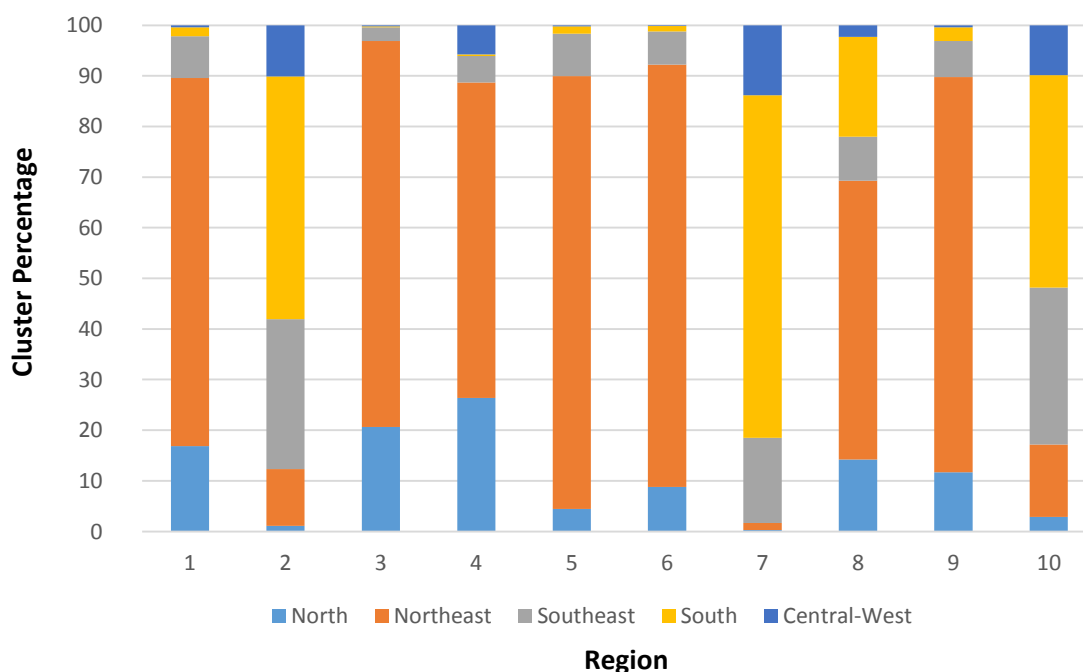
After applying this two-step analysis jointly, it was possible to find the patterns and structures in these almost 5 million family farmers. For instance, the cluster number four, as shown in Table 3, is the smallest and has the highest mean area. However, smallholders in this group have low average income and the second lowest productivity, thereby demonstrating that though they own

a larger area (in ha), these do not guarantee greater cash revenues. A fraction of farms can be in naturally unproductive areas or suffer from mismanagement, and more than 95000 (2%) family farmers in Brazil are in this situation.

One of the main points presented in the results can be seen in cluster number seven. This group has the largest

**Table 3.** Clusters' summary and means.

Cluster	Total	%	Age	Schooling	Area	Income	Diversity Index	Cooperative (%)	Rural Assistance (%)	Productivity
1	613,773	13.1	31.66	3.78	6.2	7,602.41	0.0178	0.055	0.00	11,192.83
2	617,266	13.1	54.03	3.31	15.83	42,889.77	0.3050	2.265	0.00	28,021.01
3	423,491	9.0	44.49	2.94	54.4	7,757.08	0.4172	0.144	0.01	115.99
4	95,632	2.0	47.88	3.28	184.96	23,068.28	0.3791	1.844	6.09	125.77
5	821,990	17.5	56.46	2.19	10.14	6,598.12	0.5896	0.107	0.00	1,555.76
6	810,682	17.3	32.48	3.81	7.8	6,481.60	0.5756	0.051	0.00	2,091.07
7	213,942	4.6	53.28	3.75	28.95	78,423.15	0.3622	88.804	8.91	6,671.22
8	335,090	7.1	43.88	3.1	14.79	12,554.32	0.3788	1.685	100.00	3,615.77
9	490,767	10.4	54.28	2.28	8.71	6,241.64	0.0266	0.071	0.00	3,602.55
10	276,789	5.9	32.49	6.24	18.64	54,666.04	0.2799	7.229	0.36	27,489.92

**Figure 2.** Smallholders by clusters and regions.

number of cooperated smallholders and the highest mean income, aside from its great productivity. Although other clusters achieved better productivity averages with much lower cooperativism, the mean income of this highly cooperated group is significantly greater than the mean income found in other clusters. Also, more than 80% of the farmers in this group are from the South and Southeast regions (Figure 2), thereby presenting evidence of the relation between the regions and income and productivity.

This study's results corroborate other studies, such as FAO (2014) and Ito et al. (2012), which state that cooperativism is a key factor to strengthen family farmers, since it plays an important role in their

production and access to markets, and is an avenue for farmers to improve their incomes. However, this characteristic raises serious concerns, since only 5% of Brazilian family farmers are members of agricultural cooperatives (Herrera et al., 2017). The lack of cooperativism is also cited by Theodossiou et al. (2018) as one of the weaknesses of Greek agricultural sector. In order to guarantee the future of this sector, increasing cooperativism needs to be one of the targets of policy makers.

For decades, public policies have supported a production system based on specialization, intensification and scale enlargement developing a commercial food system driven by supermarkets where family farmers

struggle to compete against industrial farmers. Additionally, the government has not encouraged smallholder's organization and empowerment, and thus they cannot act strongly as an oppositional force (de Roest et al., 2018; Córdoba et al., 2018; Blanc and Kledal, 2012).

One of the main public policies targeting the smallholder sector in Brazil is the PRONAF (National Program for the Strengthening of Family Farming), and despite the fact that the program has helped to improve small farmers' livelihoods, it is also criticized for making them highly dependent on the government. Furthermore, credit lines offered by PRONAF target the production of specific crops, such as commodities for exportation, leading family farmers to monoculture and specialization (Guanziroli et al., 2012; Córdoba et al., 2018).

Cluster numbers five and six are the largest ones. The first has the highest mean age and the second, conversely, has one of the lowest mean ages. Both groups have the highest average diversity indexes and have very low productivity and annual income. This presents some evidence that crop diversification may have a negative impact on smallholders' income and productivity, and also presents more evidence of the relation between the region and these two variables, since both clusters with more than 80% of the smallholders are located in the Northeast region of Brazil. One may also infer that the age of the household head does not interfere with income and productivity, considering that the mean ages found in between these two clusters are very distinct.

According to Li et al. (2009) and Meraner et al. (2015), intercropping promotes sustainable productivity growth that reduces agribusiness's negative environmental impacts. The vast majority of machinery and technologies that are developed are targeted to non-family farmers that practice monoculture, and thus diversified systems are very dependent on labor and manual harvesting, which increase production costs. As stated by Silva et al. (2018) and Coser et al. (2018), crop diversification and integrated agricultural systems are promising strategies to revert widespread land degradation and increase ecological production intensification. A study conducted by Steward (2013) in a village in the Amazon Estuary in Brazil showed that, with the emergence of new markets for agricultural crops, farmers are abandoning annual fields and replacing it with cash crops agroforests. Similarly, a research with approximately 3000 farmers in Kenya revealed that diversification with cash crops is a key intensification strategy in the country (Herrero et al., 2014).

One public policy with relative success in Brazil is the *Agricultura de Baixo Carbono* (ABC) – Low Carbon Agriculture program, which is strongly related with Brazil's Nationally Determined Contribution, offered at COP 21, for the reduction of greenhouse gas emissions. This program encourages farmers to adopt mitigation

technologies such as pasture restoration that aim to reduce deforestation and increase the implementation of integrated agricultural systems (Silva et al., 2018). According to Coser et al. (2018), the strategy of the ABC program is to convert 15 million hectares of low-productivity pastures to agri-silviculture systems, which would account for a reduction of carbon emissions of 79.50 Tg ha<sup>-1</sup> year<sup>-1</sup> during the first four years after its implementation.

Therefore, although this research results have shown that crop diversification might have negative impacts on farmer's income and productivity, this practice should be encouraged in order to increase agricultural activity sustainability. This can be achieved with more public policies that invest in the research and development of technologies for diversified systems that are accessible for smallholders. Theodossiou et al. (2018) reported similar findings, and the authors stated that agricultural policy and rural development should be designed concerning the protection of the environment.

By examining clusters two and ten, once more, we find evidence of the neutral effect of age on income and productivity, as observed in clusters five and six. These two groups with the two highest productivity averages also have great mean incomes. Nevertheless, one has a high mean age while the other has one of the lowest mean ages. In addition, the region seems to be very correlated to productivity and income when considering that more than 70% of family farmers from clusters two and ten are from the South and Southeast regions.

As stated by Guilhoto et al. (2011) and Fernandes and Woodhouse (2008), this great relation between regions and income and productivity may be due to the contrasting structures found in different regions of the country. Family farmers from the South and Southeast regions of Brazil are more likely to succeed, while farmers from the Northeast region are more similar to peasants. The South is a very developed region with great infrastructure; while the Northeast region, as stated by Simões et al. (2010) and Berdegue and Fuentealba (2011) concentrates the country's poorest population and suffers from the lack of investment. Such enormous differences are a consequence of the high inequality found in this country that affects both smallholders and other sectors as well. Sietz (2014) adds that smallholders in Northeast are more vulnerable due to dryland condition and therefore need special attention from policymakers.

In an attempt to reduce these huge contrasts and improve family farmers' livelihoods, in 2003, Brazil's Federal Government implemented the Family Farming Food Acquisition Program (PAA) to provide incentives to smallholders to increase food production both for self-consumption and for sale at guaranteed prices to public sector procurement agencies. Later, in 2009, the National School Meal Program (PNAE) required public schools to allocate at least 30% (that is, BRL 1.1 billion) of food expenditures to direct purchases from smallholders.

Under the PNAE, an estimated 47 million free-of-charge meals are served in schools every day; and between 2003 and 2014, about BRL 3.3 billion was spent under the PAA program (OECD/FAO, 2015). However, according to Graeb et al. (2016), although both policies have shown good results, these are still short-term solutions. Other measures such as the research and development of technologies for family farmers, incentives to cooperatives and the availability of quality rural assistance service are seen to be of greater importance to ensure the future of the smallholder sector (Salazar et al., 2016; Anang and Yanwen, 2014).

Another relevant fact concerns the schooling level of smallholders, which does not show a significant impact on improving family farming. According to the results, cluster number ten has the highest schooling average and the second highest mean income and productivity. However, other groups achieved great income and productivity with much lower schooling levels. As stated by Yue et al. (2010) and Greiner and Sakdapolrak (2013), this can be explained by the fact that farmers who seek higher levels of education tend to gradually move to urban areas and secure a non-farm job, thus reducing their time and attention spent with the agricultural activity. Conversely, farmers that have lower levels of education but dedicate their full time and attention to their agricultural activity are capable of achieving greater incomes and productivity.

It was also found that in cluster number eight, in which all family farmers had received rural assistance, the income and productivity averages were very low. This result is the opposite of what we expected. Several studies highlight that the rural assistance provided to smallholders are essential for their development and production improvement (Muatha et al., 2017; Fernandes and Woodhouse, 2008; Marennya and Barrett, 2007). Regardless of their income and productivity, all family farmers in the country have the right to receive rural assistance, not only the poorest. For example, in cluster number seven, which has the higher mean income, almost ten percent of smallholders use this service. A study by Vasconcelos et al. (2013) shows how rural assistance service in Santa Catarina State, Brazil, is using landraces to help smallholders adapt to climate change. Therefore, the findings here raise an important question about the quality of the rural assistance service provided to Brazilian family farmers.

## Conclusions

This study examined a database with the most current information regarding smallholders in Brazil. The innovative approach of using machine learning techniques revealed important characteristics of this farmers and the diversity of groups inside this sector. It is believed that this study provides interesting elements of discussion on the process of formulating public policies that are

capable of delivering real solutions to smallholders in Brazil.

As a major contribution of this research, the importance of cooperativism to increase family farmers' incomes and productivity was highlighted. Only a small percentage of smallholders are part of agricultural cooperatives, a characteristic that needs to change in order for these farmers to gain access to markets, obtain better selling prices and improve their returns to scale. Also, a fraction of family farmers benefit from living in one region while others are impaired by living in a different location, a sign of the great inequality found in the country that is the reason for major differences in this sector.

In addition, crop diversification was demonstrated to negatively affect family farming. Intercropping is an important practice to increase agricultural sustainability and reduce environmental impacts. Further research should focus on how to improve diversification while simultaneously increasing farmers' incomes and productivity and seeking sustainable development.

It is necessary to take a closer look at the quality of the rural assistance service provided to smallholders in Brazil in order to understand why this variable presented a negative effect on family farming, which is contrary to what is found in other countries. Finally, this study has examined several public policies targeting this sector and, despite that some have shown relative success, especially those that are focused only on short-term solutions such as low interest credit lines and price supports. In order to guarantee the future of smallholders, public policies should focus on the research and development of technologies for family farmers, incentives to cooperatives and providing of quality rural assistance services. Alongside several other studies, it is believed that family farmers can ensure food security in the future. Nevertheless, they are still not considered a priority and are ignored in policy makers' agendas.

## CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

## ACKNOWLEDGEMENTS

The authors thank FUNDECT – Foundation for the development of teaching, science and technology in the State of Mato Grosso do Sul, for a Master's scholarship for Gabriel Paes Herrera.

## REFERENCES

- Anang SA, Yanwen T (2014). Influence of agricultural trade liberalization policies on poverty reduction among Ghanaian rural smallholder rice farmers. *Bulgarian Journal of Agricultural Science* 20(4):746-753. <https://www.agrojournal.org/20/04-04.pdf>
- Bishop CM (2006). Graphical models. *Pattern recognition and machine*



- Learning 4:359-422.
- Blanc J, Kledal PR (2012). The Brazilian organic food sector: Prospects and constraints of facilitating the inclusion of smallholders. *Journal of Rural Studies* 28(1):142-154.
- Bosc PM, Berdegué J, Goita M, van der Ploeg JD, Sekine K, Zhang L (2013). Investing in smallholder agriculture for food security (No. 6) HLPE.
- Chakraborty C, Joseph A (2017). Machine learning at central banks.
- Córdoba D, Selifa T, Abrams JB, Sombra D (2018). Family farming, agribusiness and the state: Building consent around oil palm expansion in post-neoliberal Brazil. *Journal of Rural Studies* 57:147-156.
- Coser TR, de Figueiredo CC, Jovanovic B, Moreira TN, Leite GG, Cabral Filho SLS, Marchão R L (2018). Short-term buildup of carbon from a low-productivity pastureland to an agrisilviculture system in the Brazilian savannah. *Agricultural Systems*.
- Food and Agriculture Organization (FAO) (2014). The State of Food and Agriculture (SOFA Report). Rome: FAO. <http://www.fao.org/3/a-i4040e.pdf>
- Fernandes LADO, Woodhouse PJ (2008). Family farm sustainability in southern Brazil: An application of agri-environmental indicators. *Ecological Economics* 66(2-3):243-257.
- Gaspar P, Escribano AJ, Mesías FJ, Escribano M, Pulido AF (2011). Goat systems of Villuercas-Ibores area in SW Spain: Problems and perspectives of traditional farming systems. *Small Ruminant Research* 97(1-3):1-11.
- Graeb BE, Chappell MJ, Wittman H, Ledermann S, Kerr RB, Gemmill-Herren B (2016). The state of family farms in the world. *World development* 87:1-15.
- Greiner C, Sakdapolrak P (2013). Rural-urban migration, agrarian change, and the environment in Kenya: a critical review of the literature. *Population and Environment* 34(4):524-553.
- Guanziroli CE, Buainain AM, Di Sabbato A (2012). Dez anos de evolução da agricultura familiar no Brasil:(1996 e 2006). *Revista de Economia e Sociologia Rural* 50(2):351-370.
- Guilhoto J, Azzoni CR, Silveira FG, Ichihara SM, Diniz BPC, Moreira GRC (2011). PIB da agricultura familiar: Brasil-Estados.
- He R, Hu BG, Zheng WS, Kong XW (2011). Robust principal component analysis based on maximum correntropy criterion. *IEEE Transactions on Image Processing* 20(6):1485-1494.
- Herrera GP, da Costa RB, de Moraes PM, Mendes DRF, Constantino M (2017). Smallholder farming in Brazil: An overview for 2014. *African Journal of Agricultural Research* 12(17):1424-1429.
- Herrero M, Thornton PK, Bernués A, Baltenweck I, Vervoort J, van de Steeg J, Staal SJ (2014). Exploring future changes in smallholder farming systems by linking socio-economic scenarios with regional and household models. *Global Environmental Change* 24:165-182.
- Hotelling H (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology* 24(6):417.
- Howley T, Madden MG, O'Connell ML, Ryder AG (2006). The effect of principal component analysis on machine learning accuracy with high dimensional spectral data. In *Applications and Innovations in Intelligent Systems XIII* (pp. 209-222). Springer London.
- IBGE, Sidra (2006). Sistema IBGE de recuperação automática. Censo Agropecuário 2006: Segunda apuração.
- Ito J, Bao Z, Su Q (2012). Distributional effects of agricultural cooperatives in China: Exclusion of smallholders and potential gains on participation. *Food policy* 37(6):700-709.
- Jain AK (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters* 31(8):651-666.
- Jolliffe IT (2002). Graphical representation of data using principal components. *Principal component analysis* pp. 78-110.
- Lantz B (2013). *Machine learning with R*. Packt Publishing Ltd.
- Li C, He X, Zhu S, Zhou H, Wang Y, Li Y, Long Y (2009). Crop diversity for yield increase. *PLoS One* 4(11):e8049.
- Lowder SK, Scoot J, Singh S (2014). What do we really know about the number and distribution of farms and family farms in the world. *Background paper for the State of Food and Agriculture* 8:1-45.
- Marennya PP, Barrett CB (2007). Household-level determinants of adoption of improved natural resources management practices among smallholder farmers in western Kenya. *Food policy* 32(4):515-536.
- Meraner M, Heijman W, Kuhlman T, Finger R (2015). Determinants of farm diversification in the Netherlands. *Land Use Policy* 42:767-780.
- Muatha IT, Otieno DJ, Nyikal RA (2017). Determinants of smallholder farmers awareness of agricultural extension devolution in Kenya. *African Journal of Agricultural Research* 12(51):3549-3555.
- OECD/FAO (2015). *OECD-FAO Agricultural Outlook 2015*, OECD Publishing, Paris. [https://www.oecd-ilibrary.org/docserver/agr\\_outlook-2015-en.pdf?expires=1535652366&id=id&accname=guest&checksum=8CE8F823715769CCA93EFA69B25F807E](https://www.oecd-ilibrary.org/docserver/agr_outlook-2015-en.pdf?expires=1535652366&id=id&accname=guest&checksum=8CE8F823715769CCA93EFA69B25F807E)
- R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.agrojournal.org/20/04-04.pdf>
- de Roest K, Ferrari P, Knickel K (2018). Specialisation and economies of scale or diversification and economies of scope? Assessing different agricultural development pathways. *Journal of Rural Studies* 59:222-231.
- Salazar L, Aramburu J, González-Flores M, Winters P (2016). Sowing for food security: A case study of smallholder farmers in Bolivia. *Food Policy* 65:32-52.
- Sepúlveda WS, Maza MT, Pardos L, Fantova E, Mantecón ÁR (2010). Farmers' attitudes towards lamb meat production under a Protected Geographical Indication. *Small ruminant research* 94(1-3):90-97.
- Sietz D (2014). Regionalisation of global insights into dryland vulnerability: Better reflecting smallholders' vulnerability in Northeast Brazil. *Global Environmental Change* 25:173-185.
- Silva RDO, Barioni LG, Pellegrino GQ, Moran D (2018). The role of agricultural intensification in Brazil's Nationally Determined Contribution on emissions mitigation. *Agricultural Systems* 161:102-112.
- Simões AF, Kligerman DC, La Rovere EL, Maroun MR, Barata M, Obermaier M (2010). Enhancing adaptive capacity to climate change: The case of smallholder farmers in the Brazilian semi-arid region. *Environmental science and policy* 13(8):801-808.
- Steward A (2013). Reconfiguring agrobiodiversity in the Amazon estuary: market integration, the Açaí trade and smallholders' management practices in Amapá, Brazil. *Human ecology* 41(6):827-840.
- Theodossiou G, Karelakis C, Goulas A (2018). Perceptions and clustering of Greek farmers on the new CAP: Opportunity or threat?. *African Journal of Agricultural Research* 13(10):440-451.
- Vasconcelos ACF, Bonatti M, Schindwein SL, D'Agostini LR, Homem LR, Nelson R (2013). Landraces as an adaptation strategy to climate change for smallholders in Santa Catarina, Southern Brazil. *Land Use Policy* 34:250-254.
- Venables WN, Ripley BD (2013). *Modern applied statistics with S-PLUS*: Springer Science and Business Media.
- Wooldridge JM (2015). *Introductory econometrics: A modern approach*. Nelson Education.
- Yue Z, Li S, Feldman MW, Du H (2010). Floating choices: A generational perspective on intentions of rural-urban migrants in China. *Environment and Planning A* 42(3):545-562.
- MacQueen J (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1(14):281-297.
- Ding C, He X (2004). K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning* (P 29). ACM.
- Berdegué JA, Fuentealba R (2011). Latin America: The state of smallholders in agriculture. In *IFAD conference on new directions for smallholder agriculture* 24:25.
- Simpson EH (1949). Measurement of diversity. *Nature* <https://www.nature.com/articles/163688a0>