Global Food Security 30 (2021) 100555

Contents lists available at ScienceDirect

Global Food Security

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

Assessing food availability: A novel approach for the quantitative estimation of the contribution of small farms in regional food systems in Europe

Nuno Guiomar[a,*], Sérgio Godinho[a], Maria Rivera[a], Teresa Pinto-Correia[a], Rui Machado[a], Marta Czekaj[b], Ewa Tyran[b], Jacek Puchała[b]

[a] MED, Mediterranean Institute for Agriculture, Environment and Development, University of Évora, Ap. 94, 7002-554, Évora, Portugal
[b] University of Agriculture in Krakow, The Faculty of Agriculture and Economics Al. Mickiewicza 21, 31-120, Krakow, Poland

A R T I C L E I N F O

Keywords:
Small-scale farming
Remote sensing data
Self-reported yields
Farm size-productivity debate

A B S T R A C T

Recent findings on the contribution of smallholders to global food production and security challenge the values used in several reports of international organizations. The skewed distribution of the number of farms and the agricultural area by farm size may explain overestimations in small farms' food production. In fact, the highest values found in literature seem to be more strongly correlated to the total number of small farms than to the actual area they cover, suggesting errors in the estimation procedures. Additionally, a significant part of the small farms is not considered in official statistics, thus limiting the use of the data and also leading to underestimations. New efforts are thus needed to develop and apply methodologies to reduce the error and uncertainty of these estimates. In this paper we demonstrate the progress obtained by using a novel approach to provide new and more accurate estimates on the availability of food produced in small farms in 17 European regions (NUTS-3 level) distributed in 8 countries. Our assessment was carried out using two data sets: [1] data on crop area and production for *a priori* selected key products in each reference region, collected through questionnaires to small producers; [2] remote sensing-based products derived from Sentinel-1 and Sentinel-2 images, including crop type maps with ground-truth validation and small-scale farming systems probability maps. To reduce error propagation resulting from self-reported yield estimates, we used robust measures of central tendency based on Tukey's bi-weight function to compute the overall production in small farms in each region, minimizing the effect of outliers. The self-reported yields by small farmers were also compared with national and regional values of productivity per unit area and discussed in light of previous findings. Our results highlight not only the importance of small farms in the European context, but also their diversity in productivity levels. In addition to the novel methodological steps that underlie our study, which involve the combination of remote sensing data with data resulting from field surveys, the approach undertaken allows to better understand the contribution of small farmers to food security in each regional context, and the potential they have to support short food supply chains. Our findings can be key in supporting policy options that aim to enhance food security by reducing the EU footprint through strengthening and diversifying regional food systems.

1. Introduction

FAO (2014) defined "food security" through four quantitative components: availability, access, utilization, and stability; which today are widely accepted in scientific literature and practice assessments (e.g., Fan and Brzeska, 2016; Alonso et al., 2018). Burchi and de Muro (2016) discussed the term in light of five approaches (food availability, income-based, basic needs, entitlement, and sustainable livelihoods), where also food availability is at the basis. The above mentioned components are described by Barrett (2010) as being "inherently hierarchical". Although food availability is insufficient to achieve food security, it is critical to safeguard the other dimensions (Barrett, 2010; Stringer, 2016). Economic crises such as the one triggered by COVID-19, can have severe implications in all dimensions of food security. Particularly, food availability and food access have been directly affected by lockdowns, but also by the limitations imposed on the transport sector or

* Corresponding author.
E-mail address: nunogui@uevora.pt (N. Guiomar).

https://doi.org/10.1016/j.gfs.2021.100555
Received 31 December 2019; Received in revised form 25 May 2021; Accepted 25 May 2021

changes in market prices (e.g., Deaton and Deaton, 2020; Devereaux et al., 2020). In such a context, enhancing or revitalizing short supply chains linked to small farms can be a strategy to minimize the constraints on food supply and, particularly, to local food consumption (Altieri and Nicholls, 2020). However, data on food production in small scale farming systems is inappropriately or insufficiently collected, and this information is critical to estimate their overall production capacity and, therefore, how they can contribute to food availability in the regional food systems.

Nevertheless, the assessment of the spatial distribution of small farms, their productivity and contribution to the food systems is a challenging process (Ricciardi et al., 2018) since the diversity of small farms' types is high and difficult to address (Guarín et al., 2020). Small farms, in terms of their structural size, can generate high turnovers and incomes representing significant business operations when they are highly-specialized or produce and process high value products, whereas small farms oriented to agricultural commodities are less economically sustainable and less competitive (Davidova et al., 2012; Tocco et al., 2013; Guarín et al., 2020).

Samberg et al. (2016) estimated the contribution of smallholders in 83 countries distributed by Latin America, sub-Saharan Africa, and South and East Asia. They obtained an estimated average area covered by small farms of 35% of cropland. They analysed 41 crops and determined that small farms produced 52.5% of food calories in their study areas and 70% of the ones produced in the smallholder-dominated subnational units. However, their representativeness of small farms does not correspond to the area effectively covered by them, but rather to an estimate based on the mean agricultural area determined for each administrative unit analysed, considering as "smallholders" all of these with mean agricultural area less than 5 ha.

As highlighted by Ricciardi et al. (2018), this is a major limitation of the Samberg's approach, due to the highly skewed distributions of farm sizes. According to Lowder et al. (2016) the total number of farms worldwide exceeds 570 million. These authors showed that most of these are small-scale (less than 2 ha) and family farms, operating about 10% and 75% of the global agricultural area, respectively. This skewed distribution is also evident in Europe, where *ca.* 45% of total farms have less than 5 ha, covering about 4% of total utilized agricultural area (Guiomar et al., 2018). However, marked differences can be found between regions across Europe. For example, in Romania, Slovenia, Poland, and Estonia, between 30% and 40% of the agricultural area is operated by small-scale farmers, whereas in Bulgaria, Czech Republic and Hungary ~80% of the agricultural area is covered by the farms in the 90th percentile of farm sizes (Blacksell, 2010; Guiomar et al., 2018; Tudor, 2014). The mean is an unsuitable measure for data asymmetrically distributed (e.g., Killeen, 1985) and does not guarantee, under these conditions, that the largest proportion of the area in each of the Samberg's subnational units is effectively covered by farms below that threshold. Asymmetrically distributed data of farm sizes require more accurate approaches to assess the relative importance of smallholder agriculture and their contribution to food production and supply (Ricciardi et al., 2018). Alternatively, the authors should have used the median or any other robust statistical measure of location (e.g., Hampel et al., 1986).

Herrero et al. (2017) combined data of 41 crops, 7 livestock, and 14 aquaculture and fish products from different databases (Herrero et al., 2013; Ray et al., 2013; Watson, 2017) with spatially-explicit data on field sizes (Fritz et al., 2015) and national-based farm size distributions (e.g., Lowder et al., 2016). They estimated agricultural and nutrient production by farm size and concluded that 51–77% of all commodities and nutrients produced globally come from farms below 50 ha and ~18% of food calories come from small farms bellow 2 ha. The authors highlighted relevant regional differences both in the distribution of dominant farm sizes and in their relative importance of food production, and showed that the diversity of agricultural and nutrient production decreases with increasing farm size. Nevertheless, Ricciardi et al. (2018) stressed that both Samberg et al. (2016) and Herrero et al. (2017) did not use direct measurements of crop production and/or area by farm size, which is also a limitation of the approaches followed by the authors.

This difficulty in addressing smallholder agriculture is fundamentally related to the scarcity of data on the distribution, activity and productivity of small farms (Labarthe and Laurent, 2013; Lesiv et al., 2019; Ricciardi et al., 2018; Samberg et al., 2016), implying the use of estimates based on data generalization, the combination of multiple databases and of data for inferring productivity that cannot be validated and that it is not always adjusted for each particular geographic context. Such efforts are commendable, but they deal with a number of issues that generate errors both in crop yield and land size estimates, and given the skewed distribution, small imprecisions in the baseline data can propagate and generate much more significant inaccuracies. Moreover, some of these problems may be amplified by biased measures of farm area and productivity resulting from self-reports (Carletto et al., 2013; de Groote and Traoré, 2005), which support an old and broad debate around the inverse farm size-productivity relationship (e.g., Barrett et al., 2010; Carletto et al., 2013; Julien et al., 2019; Muyanga and Jayne, 2019; Rada and Fuglie, 2019; Sheng et al., 2019; Wassie et al., 2019).

de Groote and Traoré (2005) and Carletto et al. (2015) found systematic discrepancies between direct measurements and self-reported plot areas, showing that smallholders tend to overestimate land size while large-scale farmers tend to do the opposite. Despite corrections made using Global Positioning System (GPS) assessments, several authors found that the inverse relationship persists, although in some cases it is reduced (Carletto et al. 2013, 2015; 2015; Julien et al., 2019). However, GPS measurements are not error free as reported by Cohen (2019), who developed an approach using GPS-based and self-reported areas to solve bias from errors in both estimates.

Other papers have analysed the effects of errors in self-reported crop outputs on the size-productivity relationship using direct measurements of crop-cuts (e.g., Desiere and Jolliffe, 2018; Gourlay et al., 2019). These studies also described over-reporting among small farmers and under-reporting in larger farms, and found that differences in the distribution of yields by farm sizes vanished when physical measures of crop outputs were used. Abay et al. (2019) found strong correlations between the errors in self-reported areas and crop outputs and, therefore, correcting only the errors of one of the estimates does not solve bias in the farm size-productivity relationship, and can even increase it.

The review conducted by Rada and Fugley (2019) indicates that there is no optimal agrarian structure and both small and large farms can be equally productive. In fact, the variability in productivity per unit area or in the overall production of small farms is a function of multiple factors. Land use intensity, crop diversity and type of products, and biophysical characteristics can explain differences on yields, regardless of farm size (e.g., Ali and Deininger, 2015; Assunção and Braido, 2007; Barrett et al., 2010; Benjamin, 1995; Bevis and Barrett, 2020; Iizumi and Ramankutty, 2015; Lamb, 2003). Moving forward to improve knowledge on the potentialities and weaknesses of small-scale farming systems in each specific context implies exploring different data acquisition, integration and analytical procedures (Jiménez et al., 2019; Ricciardi et al., 2018).

The advantages of direct measurements of production and area are unquestionable (Carletto et al., 2015; Ricciardi et al., 2018), but this is a time-consuming and costly process. Kilic et al. (2017) advocates that the number of measurements can be limited to reduce costs and missing values can be estimated by imputation techniques. However, this requires a deep knowledge of each context under analysis to support the definition of robust sampling rules to avoid bias resulting from sample design. Recent advances in satellite remote sensing applied to agricultural land use systems suggest that part of the above mentioned difficulties can be suppressed or minimized through the use of remotely obtained data, as it allows the spatial delineation of field sizes (García-Pedrero et al., 2017; Graesser and Ramankutty, 2017; Kuemmerle et al., 2009), but also spatially-explicit assessments of crop types and yields (Azzari et al., 2017; Belgiu and Csillik, 2018; Kenduiywo et al., 2018) and the establishment of relationships between its spatial variability and co-variates (e.g., Jin et al., 2019), particularly relevant in areas where small-scale farming is prevalent (Burke and Lobell, 2017; Delrue et al., 2013; Neigh et al., 2018).

The main objective of this paper is twofold: first, to demonstrate a new and integrated methodological approach to assess the contribution of small farms to regional food production in 17 European regions, combining data on agricultural crop types and field sizes estimated from remote sensing imagery (see Godinho et al., 2019) with self-reported yield data obtained through a field survey (Guarín et al., 2020; Rivera et al., 2020); second, to provide the assessment of this contribution, in the regions studied, which are representative of the diversity of regions in Europe concerning the importance of small farms (Guiomar et al., 2018).

Our study was developed in four fundamental phases: (1) the first was an exploratory analysis of the distribution of the selected crop types in the reference regions (RR) and respective yield values based on the field surveys; (2) in the second phase we conducted a comparison between these self-reported yields and reference values published in the national agricultural statistics to assess the deviations and discuss differences between products and regions; (3) the third part results from the integration of the self-reported yields (a robust metric of the central tendency of the yield distributions was used to minimize the effect of over- and under-estimates mentioned in the literature) with two datasets derived from remote sensing data (plot size and composition) to estimate the total production of each crop type in each region that can be affected to the small-scale farming systems and to compare to the overall production; (4) and finally determine the potential capacity of small farms to cover the regional consumption of selected key products.

2. Material and methods

2.1. Study areas and data sources

This study was conducted in 17 NUTS-3 regions of 8 European countries (Table 1), which were previously selected by an expert panel to cover a variety of the different types of small farms found in Europe (Guarín et al., 2020). The selection of the reference regions was based on the clustering process conducted by Guiomar et al. (2018) using the distribution of structural and economic farm sizes and considering the relative importance of agriculture of each region.

In our analysis we used data from 541 questionnaires to farmers that carried out face-to-face in the reference regions between May and August 2017. The average number of farmers interviewed through the reference regions was 32. Farms' sample aimed to cover the largest diversity of histories, strategies, resources, activities and challenges for small farm households in each region. Sampling was purposive and build up following a snowball process, comprising farms identified by field teams below the thresholds used for statistical and policy purposes within the European Union (5 ha or less in size and/or below 8 Economic Size Units; EC, 2011). To capture the broadest possible diversity, the sample also included farms that had different degrees of market integration and self-provisioning, and that covered a wide range of geographical locations within the region (Guarín et al., 2020).

We selected the questions related to the farms' crops produced (area covered in square meters and total annual production in mass (weight) units) using all the cases with valid responses from the complete survey available in https://doi.org/10.6084/m9.figshare.12888350.v1. Crop yield estimation was performed for a set of key-crops in each reference region, that have been selected considering their production, revenue, consumption and cultural significance (Rivera et al., 2020). The mentioned values were converted to ton.ha^{-1}. It is important to note, therefore, that the crop yields used are self-reported estimates, and not values resulting from measurements. These self-reported yields for each product were compared with national productivity data published by EUROSTAT and, for cereals and potatoes, also with regional productivity data (at NUTS-2 level), considering both the mean productivity of the last ten years and the yields established in 2017 (the year the surveys were conducted). This comparison aimed to evaluate marked deviations from average productivity.

The crop yields determined through the data collected in the field surveys were then combined with the spatial distribution of crop types obtained by Godinho et al. (2019). The authors mapped the selected key-crop products using Sentinel-1 and Sentinel-2 data and accomplished image segmentation to select small plots below 5 ha as proxies of small farms (see details in Godinho et al., 2019), given the high correlation between field crop size and farm size already highlighted in other studies (e.g., Fritz et al., 2015; Levin et al., 2006). Since we aim to relate crop yields with the spatial distribution of crop types produced in each region and because spatially scattered crop types and with small covered area are sources of noise and increase misclassification, our focus was only in the key target-products selected based on criteria related to the spatial representativeness of the crops within the region (Godinho et al., 2019). A descriptive synthesis of the methods used to produce the crop type maps is provided as supplementary material. Due to the lack of systematic data on the total production of each key product at NUTS-3 level, we used values provided by key regional informants (experts) to establish the percentage of production that in each region can be linked to small-scale farming systems.

Finally, it was also our objective to compare the estimated production achieved through our approach with consumption indicators, allowing the assessment of the proportions of consumption that could potentially be covered by small-scale agriculture. Therefore, we used the estimated consumption quantities per head (for age class) per year, provided in the "Comprehensive European Food Consumption Database" (https://www.efsa.europa.eu/en/data/food-consumption-data), by EFSA (European Food Safety Authority). To complete the dataset, we also used official statistics from the Hellenic Statistical Authority (Household Budget Survey: https://www.statistics.gr/en/statistics/-/publication/SFA05/2014); Portuguese Institute for Statistics (Food Balance Surveys: https://www.ine.pt/); Romanian Institute of Statistics (Food Balance Sheets: http://statistici.insse.ro/shop/?page=catD&lang=en&category_id=24); Spanish Ministry of Agriculture, Fishery, Food and Environment (Household Consumption Database:

Table 1
Percentage of the agricultural area covered by small farms (PAASF), percentage of small farms (PSF) and Utilized Agricultural Area in the study areas.

Code	Reference regions (NUTS III)	*n*	PAASF (%)	PSF (%)	UAA (ha)
RR01	Imathia (GR)	38	42.61	81.76	58910
RR02	Larissa (GR)	35	16.67	57.51	207550
RR03	Ileia (GR)	42	36.19	79.61	109230
RR04	Lucca (IT)	32	36.79	87.47	26310
RR05	Pisa (IT)	20	9.47	66.63	101490
RR06	Latgale (LV)	29	10.47	59.82	453200
RR07	Pierīga (LV)	24	4.60	51.41	248000
RR08	Vilniaus Apskritis (LT)	10	25.69	69.00	224820
RR09	Rzeszowski (PL)	33	59.84	91.38	194058
RR10	Nowosadecki (PL)	48	63.88	89.92	179817
RR11	Nowotarski (PL)	36	52.33	90.17	127639
RR12	Alentejo Central (PT)	36	1.43	49.24	575576
RR13	Oeste (PT)	36	27.68	75.63	64204
RR14	Bistrita-Nasaud (RO)	49	39.39	89.21	285510
RR15	Giurgiu (RO)	15	38.35	95.81	271100
RR16	Castellón (ES)	26	17.77	75.28	188364
RR17	Córdoba (ES)	32	5.33	50.71	844019

Source: EUROSTAT, National Farm Surveys and Agricultural Censuses.
n: number of questionnaires; PAASF: Percentage of agricultural area covered by farms below 5 ha; PSF: Percentage of the farms below 5 ha in relation to the total number of farms.
GR: Greece; IT: Italy; LV: Latvia; PL: Poland; PT: Portugal; RO: Romania; ES: Spain.

https://www.mapa.gob.es/app/consumo-en-hogares/consulta.asp). The estimated consumption quantities by person were multiplied by the population number in the region, by age class. This provides a rough indication of the consumption in each reference region.

2.2. Data analysis and crop production estimation by small-scale farms

We used a *t*-test for a single mean to compare the self-reported yields by small farmers obtained through the field surveys in each NUTS-3 region with the published yield data at national and regional levels in EUROSTAT (in the latter case only for potatoes and cereals). This test is indicated for small sample sizes and/or when the variance of the sampled population is unknown. In these *t*-tests, the yields published in the official statistics were considered as reference values with which the observed means of the self-reported yields for the different key products were compared.

To assess the crop production in small farms we used the crop area estimates obtained by Godinho et al. (2019) through Sentinel-1 and Sentinel-2 data. To reduce uncertainty in crop production assessments, a direct calibration estimator of the area was used (Gallego, 2004; Lambert et al., 2018) since crop area estimates can be biased as a result of errors in crop types classification (Canters, 1997). The unbiased area was determined through the following equation (Lambert et al., 2018):

$$A_j = \sum_{i=1}^{n} (A_i * P(j|i))$$

where A_j is the unbiased area of crop *j*, A_i is the total area classified as *i*, and $P(j|i)$ is the conditional probability to be *j* when knowing *i*.

The unbiased crop area computation and the crop production estimates was performed only for the highly accurate key crop products (*Fscore* > 75%; Godinho et al., 2019). After the unbiased area estimation, the crop production in small farms in each region was determined by using the estimated self-reported crop yields multiplied by the corresponding crop area of the small plots (below 5 ha). To minimize the effect of potential outliers in the self-reported yields, we did not use the average yields, but a robust measure of central tendency based on Tukey's biweight function (Huber, 1981), reducing error propagation resulting the errors in yield estimates provided by the farmers.

3. Results

3.1. Self-reported average yields

Potato was selected in different regions of our sample as a very important crop for small farmers; it is, therefore, the most frequently cultivated agricultural product among the farmers interviewed (about 39.19% of the sample), followed by cereals (34.57%), fruits (29.39%), fresh vegetables (19.22%), olives (19.04%) and grapes (18.30%). We analysed small-scale producers of potatoes in 9 out of the 17 regions considered. About 21.70% are located in the region of Bistrița-Năsăud (RR14), 13.68% in Rzeszowski (RR09), 12.74% in Nowosadecki (RR10) and 12.26% in Nowotarski (RR11) and Oeste (RR13). The average yield among small farmers is 17.71 ton/ha (Table 2), and the highest values were found in Alentejo Central (RR12; 19.99 ton/ha), Nowotarski (RR11; 19.15 ton/ha), Oeste (RR13; 18.64 ton/ha) and Rzeszowski (RR09; 18.53 ton/ha). In contrast, the lowest yields were found in the north-eastern regions of Latgale (RR06; 9.39 ton/ha) and Pierīga (RR07; 13.51 ton/ha).

More than 50% of the small-scale producers of cereals included in the sample are in the three Polish regions: 20.32% in Nowosadecki (RR10), 17.11% in Nowotarski (RR11), and 15.51% in Rzeszowski (RR09). The mean yield of cereals within the sampled small farms is 3.28 ton/ha (Table 2), and the variability is lower than we found in potatoes production (the coefficient of variation is 43.60% for cereals and 64.71% for potatoes). Only in 4 regions, out of the 12 with small cereal-producing farms, yields were higher than average. Mean yields above 4 ton/ha were found in Rzeszowski (RR09; 4.14 ton/ha), Pisa (RR05; 4.06 ton/ha) and Giurgiu (RR15; 4.01 ton/ha). The average yield in Bistrița-Năsăud (RR14) and Pierīga (RR07) is approximately half of the highest recorded values, approximately 2.13 ton/ha and 2.09 ton/ha, respectively.

The regions of Imathia (RR01), Larissa (RR02), Bistrița-Năsăud (RR14), Oeste (RR13) and Castellón (RR16) are the ones who have more small-farmers producing fresh fruits (except citrus, which was assessed separately), with 20.75%, 15.09%, 13.21 and 10.06% (both for Oeste and Castellón) of the total fruit producers, respectively. Citrus production is mainly concentrated in the regions of Ileia (RR03) and Castellón (RR16), with 51.28% and 30.77% of the citrus producers included in our sample, respectively. In Ileia the mean yield is 40.13 ton/ha and in Castellón is 33.82 ton/ha, and deviances from the mean are similar, of 30.68% and 43.61% respectively.

Regarding the production of fresh vegetables, it was possible to establish an average yield value in ~47% of the regions analysed in our study (Table 2). However, almost half of the small farmers producing fresh vegetables in our sample are in the regions of Lucca (RR04) and Alentejo Central (RR12) (48.08%), while 14.42% are located in Pierīga (RR07) and just over 10.58% in Pisa (RR05). Yield variability across regions is very high ranging from 30.83 ton/ha in Ileia (RR03) to 2.54 ton/ha in Larissa (RR02), both in Greece. However, this class of crop types includes a large variety of leafy vegetables, legumes and root crops.

The production of grapes and olives is mainly concentrated in the Southern European regions. Small-scale producers of grapes were found in Oeste (RR13; 21.21%), Alentejo Central (RR12; 19.19%), Lucca and Imathia (RR04 and RR01; 13.13%), and Ileia (RR03; 12.12%). The highest yields were found in Larissa (RR02; 14.63 ton/ha), Imathia (RR01; 9.25 ton/ha) and Córdoba (RR17; 8.68 ton/ha), and the lowest ones in Alentejo Central (RR12; 5.69 ton/ha) and Pisa (RR05; 5.60 ton/ha) (Table 2).

Concerning the farms producing olives, 27.18% are located in Ileia (RR03), 20.39% in Alentejo Central (RR12), 17.48% in Lucca (RR04), 12.62% in Castellón (RR16), and 11.65% in Córdoba (RR17). However, the mean olive yields in Lucca, in Italy, stands out from all others at 4.23 tons/ha (Table 2). In the Alentejo Central and Castellón, which follows Lucca in the productivity ranking, the self-reported yields were also much lower, of 2.77 ton/ha and 2.34 ton/ha, while in the remaining regions the average productivity was below 2 ton/ha. In Córdoba, the mean self-reported yield of olives was very low (1.03 ton/ha) when compared with the values obtained for other regions.

3.2. Deviation of self-reported yields from reference values

The comparison carried out using the *t*-test for single means between the yield estimates obtained through the surveys and the national values show high variability (Table 3), both within crops and at regional levels.

In general, small-scale farms producing potato, cereals and vegetables showed lower yields than national and regional average values registered in official statistics (regional yields are only available for cereals and potatoes). An opposite sign can be observed in fresh fruits and citrus productions. The statistically significant differences are all positive, and in general the yields declared by the farmers in our survey are higher than the average national values of the last 10 years and also higher than those established in the 2017 reference year.

However, there are relevant regional variations not only in average values but also in trends. For example, in the NUTS-2 region of Małopolskie (where Nowosadecki (RR10) and Nowotarski (RR11) NUTS-3 regions are located) the mean potato yield over the last 10 years has been lower (22.89 ton/ha) than the one registered at national level (24.46 ton/ha), while in Podkarpackie (where the Rzeszowski (RR09) region is located) was observed the opposite (25.58 ton/ha). In the reference year of 2017, the mean yield of potato (27.88 ton/ha) was

Table 2
Self-reported mean productivity values for different key crops. In the left column are the mean yield values (μ) and the respective standard deviations (σ), and in the right column are the location (T) and scale (s) parameters of Tukey's bi-weighted function.

		Potatoes		Cereals		Fruits		Peaches		Apples		Pears		Citrus		Vegetables		Olives		Grapes	
		μ(σ)	T(s)	μ(σ)	T(s)	μ(σ)	T(s)	μ(σ)	T(s)	μ(σ)	T(s)	μ(σ)	T(s)	μ(σ)	T(s)	μ(σ)	T(s)	μ(σ)	T(s)	μ(σ)	T(s)
	All regions	17.71 (11.46)	15.69 (8.70)	3.28 (1.43)	3.16 (1.37)	15.52 (13.92)	13.32 (13.81)	30.20 (9.91)	30.38 (10.17)	22.89 (17.98)	21.93 (18.45)	15.81 (13.02)	12.14 (12.66)	32.91 (16.94)	34.64 (17.64)	20.52 (14.08)	19.25 (13.93)	2.19 (1.70)	1.50 (1.88)	8.10 (4.31)	7.52 (4.32)
RR01	Imathia (GR)			3.21 (1.87)	3.19 (1.94)	24.89 (11.41)	24.64 (11.91)	29.86 (10.09)	29.99 (10.39)	31.00 (13.87)	30.57 (14.61)							1.00 (0.00)	1.00 (0.00)	9.25 (2.57)	10.00 (0.00)
RR02	Larissa (GR)			2.94 (1.49)	2.74 (1.36)	19.25 (16.63)	18.21 (16.51)	35.00 (7.07)	35.00 (7.45)	39.38 (16.86)	39.12 (17.61)	32.50 (10.61)	32.50 (11.18)			2.54 (2.09)	1.48 (0.86)	1.47 (0.57)	1.48 (0.59)		
RR03	Ileia (GR)													40.13 (12.31)	40.41 (12.30)	30.83 (4.68)	30.00 (0.00)	1.30 (0.45)	1.31 (0.46)	14.63 (3.63)	14.68 (3.73)
RR04	Lucca (IT)					9.79 (6.30)	8.23 (4.00)									21.28 (5.32)	22.92 (6.54)	4.23 (1.43)	3.98 (1.36)	7.35 (1.81)	6.88 (1.47)
RR05	Pisa (IT)			4.06 (1.56)	4.06 (1.65)											20.45 (16.34)	15.02 (12.04)			5.60 (2.37)	5.55 (2.45)
RR06	Latgale (LV)	9.39 (6.29)	9.08 (6.49)	2.90 (1.65)	2.65 (1.56)																
RR07	Pierīga (LV)	13.51 (7.87)	13.38 (8.00)	2.09 (1.05)	2.01 (1.06)	3.74 (2.89)	3.62 (3.05)			3.74 (2.89)	3.62 (3.05)					23.53 (16.92)	21.00 (18.15)				
RR08	Vilniaus Apskritis (LT)	15.20 (7.98)	15.42 (8.44)	3.28 (1.08)	3.28 (1.14)	2.90 (2.16)	3.05 (0.19)			3.20 (0.28)	3.20 (0.30)					5.92 (9.17)	2.91 (1.83)				
RR09	Rzeszowski (PL)	18.53 (9.14)	18.41 (8.69)	4.14 (1.26)	4.11 (1.28)																
RR10	Nowosadecki (PL)	17.17 (6.62)	18.76 (6.54)	3.45 (1.48)	3.22 (1.03)	26.54 (10.56)	24.45 (10.32)			31.05 (10.90)	31.24 (11.17)	13.50 (2.12)	13.50 (2.24)			25.60 (14.21)	35.00 (0.00)				
RR11	Nowotarski (PL)	19.15 (3.13)	19.18 (3.07)	2.84 (1.01)	2.75 (0.94)																
RR12	Alentejo Central (PT)	19.99 (12.22)	18.68 (13.09)			10.35 (8.84)	10.08 (9.15)							10.71 (13.78)	5.19 (5.70)	22.09 (16.51)	17.95 (16.07)	2.34 (2.24)	1.47 (1.12)	5.69 (3.20)	5.63 (3.35)
RR13	Oeste (PT)	18.64 (15.34)	15.47 (12.85)			14.01 (12.87)	9.23 (10.58)					14.01 (12.87)	9.23 (10.58)							7.26 (4.86)	6.66 (5.02)
RR14	Bistrita-Nasaud (RO)	17.49 (12.20)	14.42 (7.97)	2.13 (0.97)	2.16 (0.99)	15.90 (14.93)	10.81 (12.54)			15.90 (14.93)	10.81 (12.54)										
RR15	Giurgiu (RO)			4.01 (1.40)	3.99 (1.42)																
RR16	Castellón (ES)					1.80 (1.44)	1.35 (0.72)							33.82 (14.75)	35.46 (14.28)			2.77 (1.08)	2.63 (1.12)		
RR17	Córdoba (ES)			2.95 (1.29)	2.94 (1.33)													1.03 (1.01)	0.76 (0.45)	8.62 (3.75)	8.68 (3.19)

T: Tukey's bi-weight location estimate; s: Tukey's bi-weight scale estimate.

Table 3
T-test for single mean against the mean yield values of the last ten years and the reference value in 2017 (in parenthesis).

Reference Regions		Potatoes		Cereals		Fruits	Peaches	Apples	Pears	Citrus	Vegetables	Olives	Grapes
		Nat.	Reg.	Nat.	Reg.	Nat.	Nat.	Nat.	Nat.	Nat.	Nat.	Nat.	Nat.
RR01	Imathia (GR)			−1.15 (−0.50)	−0.83 (1.15)	6.35*** (5.39***)	5.79*** (3.49**)	1.16 (0.25)				a)	−0.82 (−0.99)
RR02	Larissa (GR)			−2.76* (−1.57)	−4.34*** (−2.56*)	2.05 (1.49)	3.24 (2.36)	3.19** (2.04)	2.31 (1.93)		−34.14*** (−33.40***)	−1.28 (2.73*)	
RR03	Ileia (GR)									6.22*** (4.18***)	−2.34* (−1.90)	−4.76 (4.11***)	4.59*** (4.47***)
RR04	Lucca (IT)					−2.63*					−9.19*** (−8.47***)	5.38*** (5.80***)	−6.89*** (−7.11***)
RR05	Pisa (IT)			−1.66 (−1.42)	0.46						−2.15 (−2.00)		−7.27*** (−7.42***)
RR06	Latgale (LV)	−6.38***		−1.66 (−3.56***)									
RR07	Pierīga (LV)	−2.36*		−3.93** (−5.90***)		1.06 (1.83)		0.08 (1.35)			−0.22 (−0.41)		
RR08	Vilniaus Apskritis (LT)	−0.05 (0.82)		−0.92 (−2.15)		−0.40 (−1.20)		−11.58 (−21.35*)			−4.48** (−3.75**)		
RR09	Rzeszowski (PL)	−3.49** (−5.50***)	−4,15***	1.48 (−0.24)	2.94**								
RR10	Nowosadecki (PL)	−5.72*** (−8.41***)	−4.49*** (−3.80***)	−1.45 (−3.14**)	−0.87 (−2.54*)	5.26*** (5.93***)		3.96** (4.65**)	3.72 (3.94)		0.00 (−0.67)		
RR11	Nowotarski (PL)	−8.12*** (−13.35***)	−5.72*** (−4.36***)	−5.44*** (−7.61***)	−4.57*** (−6.81***)								
RR12	Alentejo Central (PT)	0.12 (−0.86)	−1.54			2.19 (2.01)				−0.90 (−1.45)	−6.74*** (−7.80***)	1.35 (−0.21)	1.33 (0.93)
RR13	Oeste (PT)	0.01 (−1.01)	0.72			3.26** (3.09**)			−0.25 (−0.65)				2.39* (2.12*)
RR14	Bistrita-Nasaud (RO)	1.21 (−0.39)		−5.49** (−8.41***)	−4.38** (−7.63***)	2.11* (2.68*)		2.11* (3.01**)					
RR15	Giurgiu (RO)			−0.37 (−3.34**)	−0.09 (−4.39***)								
RR16	Castellón (ES)					−8.29*** (−8.68***)				2.90* (2.89*)		−0.90 (0.71)	
RR17	Córdoba (ES)			−1.18 (0.44)	−0.15 (−0.13)							−6.90*** (−5.24***)	1.80 (2.42*)

a) Insufficient number of cases to compute the *t*-test.
Nat.: Comparison with national yields; Reg.: Comparison with regional yields (NUTS-2 level).

higher than the mean value of the last 10 years, revealing that 2017 was, in general, more favorable for potato production. Nevertheless, in the Małopolskie region this value was lower (22.00 ton/ha) than the same mean established at regional level.

Regarding the productivity of olives and grapes, differences were found both between regions and between the mean value of the last ten years and the reference year, for the same region. Despite de non-significant results of some tests for olive groves, we identified changes in signal bias when performed using as reference value the mean yield of the last ten years or the yield of olives in 2017.

Overall, the results suggest that signal bias is more related to the characteristics of the key product than to the characteristics of the region itself or the prevalence of smaller farms in each region.

3.3. Crop production estimation in small farms

Combining unbiased crop area estimations with the self-reported yields of the key crops we estimated a total production of 1,524,167.94 tonnes of agricultural crops by small farms over the 17 reference regions here reported (Table 4). Considering the five main groups of crop products, the results show a total production of 994,157.23 tonnes of fruits (apples, pears, peaches, citrus and other orchards), 214,315.76 tonnes of cereals (wheat, barley, oats, and rye), 166,269.66 tonnes of vegetables (including potatoes), 89,834.51 tonnes of grapes (mainly for wine production), and 59,590.78 tonnes of olives (mainly for oil production). This means that these small farms may potentially produce an average of 28.64 ton/ha/year of fruits, 3.52 ton/ha/year of cereals, 14.95 ton/ha/year of vegetables (including potatoes), 1.10 ton/ha/year of olives, and 8.61 ton/ha of grapes.

There are considerable differences in crop area estimations (e.g., cereal: min = 1304.62 ha; max = 17,416.00 ha) and in the self-reported crop yields (e.g. cereal: min = 2.01 ton/ha; max = 4.11 ton/ha), and as well the crop production estimates differ significantly

Table 4
Unbiased crop area for small plots (<5 ha) and production estimations for each key crop product.

Reference region	Crop types	Fscore (%)	Estimated crop area (ha)	Self-reported yields (ton/ha)	Estimated annual production (ton/year)
RR01	Peaches	80.4	8782.06	29.99	263,373.98
RR02	Vegetables	73.2	814.55	1.48	1205.53
RR03	Olive groves	85.5	20,618.20	1.31	27,009.84
	Vineyards	77.3	2289.35	14.68	33,607.66
RR04	Olive groves	87.2	2180.83	3.98	8679.70
	Vineyards	81.9	789.90	6.88	5434.51
RR05	Cereals	75.4	1304.62	4.06	5296.76
RR06	Cereals	68.1	6596.26	2.65	17,480.09
RR07	Cereals	88.8	3389.91	2.01	6813.72
RR08	Vegetables	76.5	1592.14	2.91	4633.13
RR09	Cereals	91.7	15,603.10	4.11	64,128.74
	Potatoes	86.9	8714.34	18.41	160,431.00
RR10	Cereals	70.8	10,779.93	3.22	34,711.37
	Apples	81.4	1705.50	31.24	53,279.82
RR11	Cereals	70.8	3020.26	2.75	8305.72
RR12	Vineyards	87.5	1867.13	5.63	10,511.94
RR13	Pears	90.9	2407.64	9.23	22,222.52
	Vineyards	83.4	3627.34	6.66	24,158.08
RR14	Orchards	98.2	4799.07	10.81	51,877.95
RR15	Cereals	98.2	17,416.00	3.99	69,489.84
RR16	Citrus	88.1	17,016.44	35.46	603,402.96
RR17	Cereals	87.9	2742.21	2.95[a]	8089.52
	Olive groves	87.2	31,449.00	0.76	23,901.24
	Vineyards	85.5	1857.41	8.68	16,122.32

[a] Due to the absence of field-level wheat information the mean national wheat yield was used.

within key crops and across reference regions. In absolute values, the highest production levels were obtained for citrus (603,402.96 ton) in Castellón (RR16), peaches (263,373.98 ton) in Imathia (RR01), and potatoes (160,431.00 ton) in Rzeszowski (RR09) (Table 4). The lowest production estimates were obtained for vegetables in Larissa (RR02; 1205.53 ton) and in Vilniaus Apskritis (RR08; 4633.13 ton), cereals (5296.76 ton) in Pisa (RR05), vineyards (5434.51 ton) in Lucca (RR04), and cereals (6813.75 ton) in Pierīga (RR07) (Table 4).

The contribution of the small farms to the overall production of cereals is, in general, low. Exceptions in our sample occur in regions where the area covered by farms below 5 ha is prevalent, such as Rzeszowski (RR09; 33.99%), Nowosadecki (RR10; 42.61%) and Nowotarski (RR11; 62.28%).

Concerning the permanent crops, particularly for orchards, the contribution of small farms is substantially higher than the one evidenced for annual crops. These are citrus in Castellón (RR16; 112.04%), unspecified orchards (mainly apples according to the surveys) in Bistrița-Năsăud (RR14; 110.83%), apples in Nowosadecki (RR10; 80.94%), and peaches in Imathia (RR01; 68.95%).

The percentage of total regional production higher than 100% found for citrus in Castellón and for fresh fruits in Bistrița-Năsăud, can be explained by overestimation of the total area covered by fruit orchards in small scale farms, overestimation of the self-reported yields, or underestimation of the total production related to the proportion of food produced on small farms that are used for self-consumption.

Concerning olives production, the contribution of the small farms in Ileia (RR03) and Lucca (RR04) is much higher than in Córdoba (RR17) where olives are mainly cultivated in large-scale farms. In Ileia and Lucca the area covered by small-scale farming systems is 36.19% and 36.79% while in Córdoba is only 5.33%.

The same occurs in Rzeszowski (RR09), where the area covered by small farms is even higher (59.84%) and where the relevance of the small farms to the overall production of potato is also very high (75.3%). However, for the remaining leafy vegetables, legumes and root vegetables our results are inconclusive. A considerable part of their distribution within the small farms are associated with mixed crops farming systems, and this complexity in its spatial distribution is difficult to approach.

These differences can also be addressed from another perspective, in relation to the potential of small farms in the reference regions to cover the regional consumption demands of the key products analysed (Fig. 2). The estimation of the average consumption for each crop, in each region, based on the EFSA data (complemented by national surveys for Greece, Portugal, Romania and Spain), provided the indicators on consumption demand. The difference between the potential of permanent agricultural crops, which are generally in surplus, is clear from the annual crops which, with the exception of potatoes, show a deficit in production compared to the consumption of these products.

4. Discussion

Our results show that small farms produce ~29% of the overall production of the selected key products (considering the total regional production of these crops provided by the key informants). We acknowledge this is a broadly estimated value, encompassing considerable differences between the products and the regions considered, as can be observed in Fig. 1. However, it shows an order of magnitude which should not be underestimated. Furthermore, it is very much in line with the results of Ricciardi et al. (2018), what remarkably contributes to reinforcing the interest of both studies. These authors, based on agricultural censuses covering 55 countries and 154 crop types, showed that small farms play a central role in terms of crop production, contributing with 28–31% of the total global crop production.

We did not find inconsistencies in the self-reported yields of cereals by small farmers. For example, Schils et al. (2018) determined, in some climatic zones in Latvia and Romania, yields below the 10th percentile for wheat (2.2 ton/ha), and between 4.0 and 4.8 ton/ha in regions with

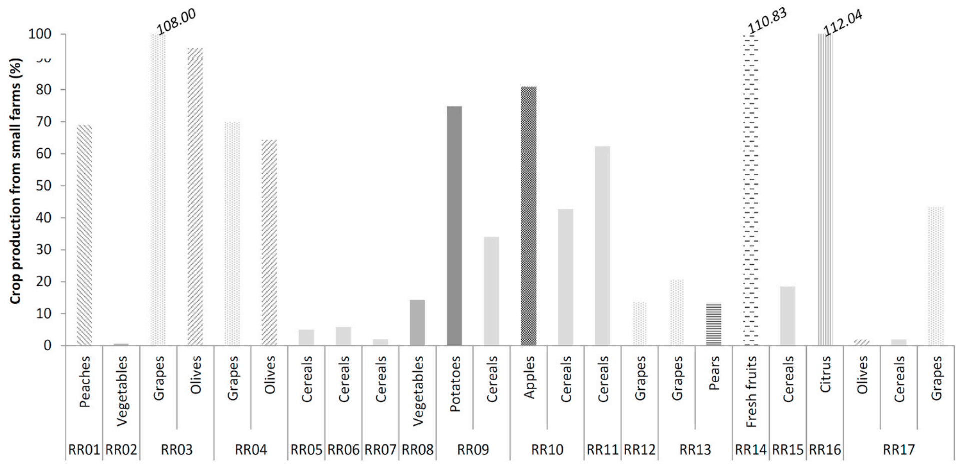

Fig. 1. Production of small farms in each region as a percentage of the total crop production for the selected key crop products.

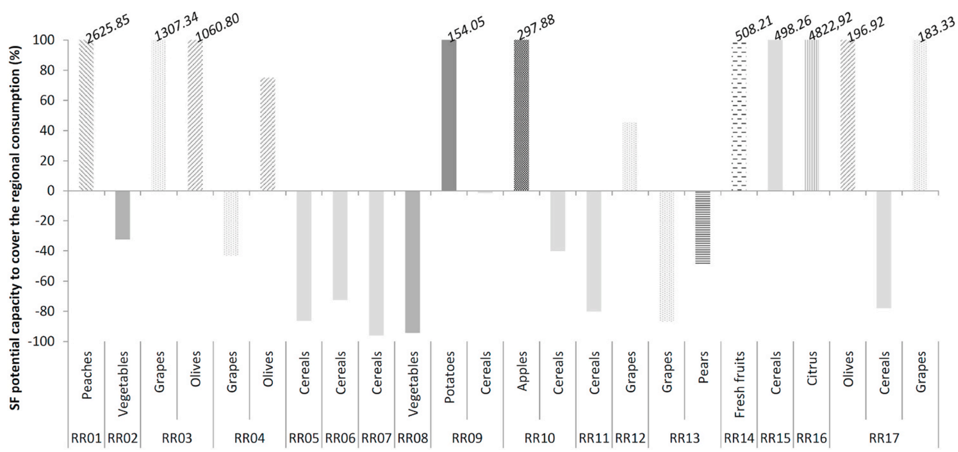

Fig. 2. Potential capacity of small farms to cover the regional consumption of the selected key crop products. The percentages above 100% (values of production that exceed those of consumption) are shown in the graph.

highest yields, varying in a range close to that determined in our study. The lower contribution of small farms to the overall production of cereals in the reference regions analysed can be more properly explained by the farm size typically related to these production models, than by any discrepancies in the self-reported yields. According to the most important farming systems existing in Europe obtained by Andersen (2017), the production of cereals is more dominant in large-scale farming systems with low to medium land use intensity. These are low input but highly mechanized land use systems that require large areas to be economically sustainable (Fereres and Villalobos, 2016).

The regional variability found in potato yield estimates, and the established mean values, are also in accordance with spatially explicit studies conducted on a global scale (*e.g.*, Monfreda et al., 2008; Haverkort et al., 2014), but in this case the contribution of small farms to the overall production seems to be much higher than the one observed for cereal farms. Also, the permanent and specialized orchards seem to benefit from management at small scales. The production in small orchards accounts for a considerable proportion of the total production of fresh fruits. It is important to stress out here that some of these large productions of fruits are not related with mixed crop farming or inter-crop management systems, but rather with spatially clustered, specialized and export-oriented productions of both post-processed and fresh fruits. Greece and Spain are among the largest world exporters of citrus fresh fruits, and both regions are highlighted as important producers of citrus (Navarro, 2015; Ordoudi et al., 2018). Spain is the fifth largest citrus producer and the world's largest exporter of fresh citrus fruit, exporting 50% of the overall production, while 20% is consumed as fresh fruit in the country and 18% is processed (Navarro, 2015), while in Greece the proportion that is channelled to processing and to the internal market strongly varies between species and varieties (Ordoudi et al., 2018).

However, in some regions, such as Castellón and Bistrița-Năsăud, we estimated a small farming's contribution to the total output of more than 100%. In Castellón, these values may result from an overestimation of the total area covered by fruit orchards in small scale farms. Small farms from the citrus key sector of Castellón are export oriented and highly productive. However, their distribution is spatially clustered and farmland is highly fragmented (in the 27 surveyed farms in Castellón, the mean number of plots per farm is 10.20). These particular characteristics may underlie the differences found in crop area estimates (in small farms) obtained from Sentinel satellite data and official statistics. The total area covered by citrus in small scale plots estimated using the Sentinel crop map was 17,016.74 ha, while the area registered in the official statistics is 14,583.70 ha, resulting in a difference of 2433.04 ha. The above mentioned excess can also be related to an overestimation in the self-reported yields given the noticeable disparity between yield estimates from different sources. The self-reported citrus yield for the region of Castellón was estimated as 35.46 ton/ha, but the value resulting from the official statistics is substantial lower, only 15.30 ton/ha for this reference region. According to Navarro (2015) the national average citrus yield is 19.10 ton/ha and, therefore, none of the above figures seem to be correct, yet it is more likely to be above the national average than below given the high specialization of these farms and their market orientation. Yet it cannot be ruled out the hypothesis of this difference results from underestimation of the total production related to the proportion of food produced on small farms that are used for self-consumption. In Bistrița-Năsăud 19% of total production of the surveyed farms are retained for self-consumption and 21% of the farms kept fruits, and remains unclear if this amount of food is properly registered in the official statistics.

We also found discrepancies in the self-reported yields of olives, particularly in Lucca where the mean value was substantially higher than the one reported by Maselli et al. (2012) for this region. According to the authors, olive groves in Tuscany are predominantly distributed by small plots below 2 ha (~43% of the total) and yields range from 0.88 ton/ha in Siena to 2.32 ton/ha in Lucca and Livorno. Areal and Riesgo (2014) also reported significant differences in productivity levels among the sub-regions analysed in Andalusia, and particularly between three important areas in the region of Córdoba, one in the central part of the region (La Sierra) and two in the southern part (Penibética and Campiña Alta). In this region medium to large-scale farming systems prevail, and the estimated yields by these authors were much higher (2.2 ton/ha in La Sierra, 4.0 ton/ha in Penibética and 5.5 ton/ha in Campiña Alta) than the ones obtained through the survey carried out in our study. In fact, the spatial variability in the distribution of productivity is high, and the differences found can be related to other factors than to estimation errors in reporting. In the traditional rainfed systems, olive tree density ranges between 30 and 173 trees ha^{-1}, while in drip irrigated super-intensive olive orchards tree density range is much higher, between 1700 and 3000 trees ha^{-1} (Vossen, 2007). The Spanish region of Andalusia, where Córdoba is located, is the major olive producer in the world, but the contribution of small farms for the total regional production in these reference regions was the lowest one, since olives are mainly cultivated in large-scale farms (Areal and Riesgo, 2014) but also due to the low area covered by small-scale farming systems (5.33%). In Greece the olive groves are distributed throughout the country, but Peloponnese (Ileia is situated in the western part) and Crete have the major revenues from olive oil production (Ordoudi et al., 2018), and the area covered by small farms is substantially higher than in Córdoba (36.19%). Yield also varies substantially depending on the olive cultivars used (Silveira et al., 2018) and the irrigation regime (Gómez-Rico et al., 2007; Patumi et al., 2002). Areal and Riesgo (2014) showed significant differences between irrigated and non-irrigated olive farms in Andalusia, but also between traditional olive groves located in plains and mountain areas. In the Alentejo region marked changes operated in the olive sector since 2006, with the expansion of the irrigated intensive and super-intensive olive groves in large-scale farms which allowed the increase of the national production of olives from 252,247.50 ton to 421, 386.42 ton and of olive oil from 390,493.62 hl to 677,249.14 hl between the periods 2001–2008 and 2008–2014 (Guiomar and Pinto-Correia, 2016).

There are also significant differences in yields between different horticultural crops that are accentuated by different management practices (e.g., Pieper et al., 2015) and between table grapes and wine grapes (Permanhani et al., 2016; Teixeira et al., 2007). Spatial and temporal variations in grape productivity are high even on short temporal and spatial scales. In a study conducted on a 1.4 ha table grape vineyard plot in southern Greece, Anastasiou et al. (2017) estimated variations in the average yield in the 36 plots analysed from 27.72 ton/ha in 2015 to 22.44 ton/ha in 2016. Moreover, the authors also identified high spatial variation, recording minimum values of 15.49 ton/ha and 5.71 ton/ha and maximum values of 42.28 ton/ha and 34.07 ton/ha in 2015 and 2016, respectively, which were attributed to differences in the weather conditions.

5. Conclusions

In this study we developed an approach to assess the production of a set of key agricultural products in small-scale farming in 17 European regions. With this, we aimed to contribute to the assessment of small farms contribution to food availability, a key dimension of food security. Considering the variability in yields estimates, both between crop types and regions, the inherent difficulty in establishing production estimates that can be assigned to small-scale farms becomes evident, particularly if this analysis is carried out for large spatial units. The level of uncertainty in the estimates can, however, be minimized through the use of more suitable statistical metrics and through the integration of spatially explicit data of the size and composition of the agricultural plots – and by this, the shortcomings imposed by the existing statistical data basis, be overcome.

To achieve our goals, we combined data derived from remote sensing to obtain the area covered by small farms and by each crop type, with self-reported data on crop yields. Considering the errors that can emerge from farmer's self-reported yields, which are widely discussed in the literature devoted to the research on the farm size-productivity relationship, and those that normally follow land cover classification processes, to estimate the crop production at regional level we have used an unbiased area estimate for each crop type based on the accuracy of the image classification, and also a robust measure of the central tendency in the distribution of yields obtained through the surveys to reduce the effect of outliers. Moreover, we only used the best data to produce our estimates, selecting regions and products with sufficient responses to establish average productivity, and crop maps with higher accuracy. The productivity was also compared with data published in the official statistics, to determine the deviation from the regional averages and to support the discussion, not only of the performance of small farms, but also of the values that are published in scientific literature on the production and supply of food from small production systems.

The results highlight that small farms have an important contribution to food availability in the regional food system in terms of crop production for the selected crops. This is particularly relevant for the permanent crops (e.g., orchards). Our estimates indicate that small farms produce ~29% of the total regional production stated on the data provided by small farmers and regional experts. This relevance confirms what has been demonstrated in the few other recent studies assessing small farm production, as by Ricciardi et al. (2018).

The data production capacity we have today as well as the capacity of remote sensing to reduce bias, do not justify the prevailing unawareness about what is produced in small scale farming. The high temporal and spectral resolution of images provided by satellites such as the Sentinel-1 and Sentinel-2 allows effective knowledge formalization about farm structure and farming systems, monitoring their dynamics, but also assessing food productivity possibilities and influence on food availability.

Our study has showed the viability of using such novel technologies and estimation methods to obtain quality data on small scale production. Due to the growing precision and availability of remote sensing imagery, the approach we used can now be further generalized to calculations on the production quantities by small farms, for other regions or for future monitoring of changes in the studied regions. Average yields can be obtained by direct surveys, as in the present study, or by technical expertise, easier to collect. To the best of our knowledge, this is a step further in the evidence that can be produced scientifically, on small farms contribution to food availability and therefore to regional food systems. However, our study also shows there is an incredible differentiation among small farms, across regional contexts and across crops and products. For example, while there are known high productivity levels in specialized productions of export oriented products as citrus and olives, there are many other productivity variations in other production systems for the same crops, or other types of crops. The information we have hereby generated is key in informing public policies. With the European Green Deal, and its related Farm to Fork Strategy and Biodiversity Strategy, we are entering an era in Europe, where reducing the global footprint and enhancing circular economy is key. In this renewed setting, food conservation and transport will need to be reduced, and regional food systems will need to be strengthened and transformed to meet its demands. Knowing what the contribution of small farms is and being able to monitor changes in this contribution, is crucial to design and progressively adapt targeted policy tools to achieve the sustainability goals imposed by the Green Deal. We are confident our method will be able to inform progress in this sense.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the SALSA project (Small farms, small food businesses and sustainable food and nutrition security) (Project ID: 677,363) funded under H2020-EU.3.2. – Societal Challenges - Food security, sustainable agriculture and forestry, marine, maritime and inland water research and the bioeconomy; and by FEDER Funds through the Operational Programme for Competitiveness Factors - COMPETE and National Funds through FCT - Foundation for Science and Technology.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gfs.2021.100555.

References

Abay, K.A., Abate, G.T., Barrett, C.B., Bernard, T., 2019. Correlated non-classical measurement errors, "Second best" policy inference, and the inverse size-productivity relationship in agriculture. J. Dev. Econ. 139, 171–184. https://doi.org/10.1016/j.jdeveco.2019.03.008.
Ali, D.A., Deininger, K., 2015. Is there a farm size-productivity relationship in African agriculture? Evidence from Rwanda. Land Econ. 9, 317–343.
Alonso, E.B., Cockx, L., Swinnen, J., 2018. Culture and food security. Glob. Food Sec. 17, 113–127. https://doi.org/10.1016/j.gfs.2018.02.002.
Altieri, M.A., Nicholls, C.I., 2020. Agroecology and the reconstruction of a post-COVID-19 agriculture. J. Peasant Stud. https://doi.org/10.1080/03066150.2020.1782891.
Anastasiou, E., Tsiropoulos, Z., Balafoutis, T., Fountas, S., Templalexis, C., Lentzou, D., Xanthopoulos, G., 2017. Spatiotemporal stability of management zones in a table grapes vineyard in Greece. Adv. Anim. Biosci. Precis. Agric. 8, 510–514. https://doi.org/10.1017/S2040470017000632.
Andersen, E., 2017. The farming system component of European agricultural landscapes. Eur. J. Agron. 82, 282–291. https://doi.org/10.1016/j.eja.2016.09.011.
Areal, F.J., Riesgo, L., 2014. Farmers' views on the future of olive farming in Andalusia, Spain. Land Use Pol. 36, 543–553. https://doi.org/10.1016/j.landusepol.2013.10.005.
Assunção, J.J., Braido, L.H.B., 2007. Testing household-specific explanations for the inverse productivity relationship. Am. J. Agric. Econ. 89, 980–990. https://doi.org/10.1111/j.1467-8276.2007.01032.x.
Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: testing multiple methods and satellites in three countries. Remote Sens. Environ. 202, 129–141. https://doi.org/10.1016/j.rse.2017.04.014.
Barrett, C.B., 2010. Measuring food insecurity. Science 327, 825–828. https://doi.org/10.1126/science.1182768.
Barrett, C.B., Bellemare, M.F., Hou, J.Y., 2010. Reconsidering conventional explanations of the inverse productivity-size relationship. World Dev. 38, 88–97. https://doi.org/10.1016/j.worlddev.2009.06.002.
Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time- weighted dynamic time warping analysis. Remote Sens. Environ. 204, 509–523. https://doi.org/10.1016/j.rse.2017.10.005.
Benjamin, D., 1995. Can unobserved land quality explain the inverse productivity relationship? J. Dev. Econ. 46, 51–84.
Bevis, L.E.M., Barrett, C.B., 2020. Close to the edge: high productivity at plot peripheries and the inverse size-productivity relationship. J. Dev. Econ. 143 https://doi.org/10.1016/j.jdeveco.2019.102377, 102377.
Blacksell, M., 2010. Agriculture and landscape in 21st century Europe: the post-communist transition. Eur. Countrys. 1, 13–24. https://doi.org/10.2478/v10091-010-0002-8.
Burchi, F., de Muro, P., 2016. From food availability to nutritional capabilities: advancing food security analysis. Food Pol. 60, 10–19. https://doi.org/10.1016/j.foodpol.2015.03.008.
Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proc. Natl. Acad. Sci. Unit. States Am. 114, 2189–2194. https://doi.org/10.1073/pnas.1616919114.
Canters, F., 1997. Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification. Photogramm. Eng. Rem. Sens. 63, 403–414.
Carletto, C., Gourlay, S., Winters, P., 2015. From Guesstimates to GPStimates: land area measurement and implications for agricultural analysis. J. Afr. Econ. 24, 593–628. https://doi.org/10.1093/jae/ejv011.
Carletto, C., Savastano, S., Zezza, A., 2013. Fact or artifact: the impact of measurement errors on the farm size-productivity relationship. J. Dev. Econ. 103, 254–261. https://doi.org/10.1016/j.jdeveco.2013.03.004.
Cohen, A., 2019. Estimating farm production parameters with measurement error in land area. Econ. Dev. Cult. Change 68, 305–334. https://doi.org/10.1086/700557.
Davidova, S., Fredriksson, L., Gorton, M., Mishev, P., Petrovici, D., 2012. Subsistence farming, incomes, and agricultural livelihoods in the new member states of the European Union. Environ. Plann. C Govern. Pol. 30, 209–227. https://doi.org/10.1068/c1195r.
de Groote, H., Traoré, O., 2005. The cost of accuracy in crop area estimation. Agric. Syst. 84, 21–38. https://doi.org/10.1016/j.agsy.2004.06.008.
Deaton, B.J., Deaton, B.J., 2020. Food security and Canada's agricultural system challenged by COVID-19. Can. J. Agric. Econ. https://doi.org/10.1111/cjag.12227.
Delrue, J., Bydekerke, L., Eerens, H., Gilliams, S., Piccard, I., Swinnen, E., 2013. Crop mapping in countries with small- scale farming: a case study for West. Int. J. Rem. Sens. 34, 37–41. https://doi.org/10.1080/01431161.2012.747016.
Desiere, S., Jolliffe, D., 2018. Land productivity and plot size: is measurement error driving the inverse relationship? J. Dev. Econ. 130, 84–98. https://doi.org/10.1016/j.jdeveco.2017.10.002.
Devereaux, S., Béné, S., Hoddinott, J., 2020. Conceptualising COVID-19's impacts on household food security. Food Secur. https://doi.org/10.1007/s12571-020-01085-0.
Ec, 2011. What Is a Small Farm? EU Agricultural Economics Brief, vol. 2. EuropeanCommission Agriculture and Rural Development, Brussels.
Fan, S., Brzeska, J., 2016. Sustainable food security and nutrition: demystifying conventional beliefs. Glob. Food Sec. 11, 11–16. https://doi.org/10.1016/j.gfs.2016.03.005.
FAO, 2014. The State of Food Insecurity in the World 2014. Food and Agricultural Organization of the United Nations, Rome.
Fereres, E., Villalobos, F.J., 2016. Agriculture and agricultural systems. In: Villalobos, F. J., Fereres, E. (Eds.), Principles of Agronomy for Sustainable Agriculture. Springer, Cham, pp. 1–12. https://doi.org/10.1007/978-3-319-46116-8_1.
Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., Aziz, S. A., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D.R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., van der Velde, M., Dunwoody, A., Kraxner, F., Obersteiner, M., 2015. Mapping global cropland and field size. Global Change Biol. 21, 1980–1992. https://doi.org/10.1111/gcb.12838.
Gallego, F.J., 2004. Remote sensing and land cover area estimation. Int. J. Rem. Sens. 25, 3019–3047.
García-Pedrero, A., Gonzalo-Martín, C., Lillo-Saavedra, M., 2017. A machine learning approach for agricultural parcel delineation through agglomerative segmentation. Int. J. Rem. Sens. 38, 1809–1819. https://doi.org/10.1080/01431161.2016.1278312.
Godinho, S., Guiomar, N., Machado, R., Pinto-Correia, T., Tsiligiridis, T., Ainali, K., 2019. Methodological Report and Guidance for the Use of SENTINEL-2 Data in Assessing and Monitoring of Small Farms and Crop Production. Deliverable 2.3 of the EU-H2020 Project SALSA – Small Farms, Small Food Businesses and Sustainable Food Security.
Gómez-Rico, A., Salvador, M.D., Moriana, A., Pérez, D., Olmedilla, N., Ribas, F., Fregapane, G., 2007. Influence of different irrigation strategies in a traditional Cornicabra cv. olive orchard on virgin olive oil composition and quality. Food Chem. 100, 568–578. https://doi.org/10.1016/j.foodchem.2005.09.075.
Gourlay, S., Kilic, T., Lobell, D.B., 2019. A new spin on an old debate: errors in farmer-reported production and their implications for inverse scale - productivity relationship in Uganda. J. Dev. Econ. 141 https://doi.org/10.1016/j.jdeveco.2019.102376, 102376.
Graesser, J., Ramankutty, N., 2017. Detection of cropland field parcels from Landsat imagery. Remote Sens. Environ. 201, 165–180. https://doi.org/10.1016/j.rse.2017.08.027.
Guarín, A., Rivera, M., Pinto-Correia, T., Guiomar, N., Šūmane, S., Moreno-Pérez, O.M., 2020. A new typology of small farms in Europe. Glob. Food Sec. 26 https://doi.org/10.1016/j.gfs.2020.100389, 100389.
Guiomar, N., Godinho, S., Pinto-Correia, T., Almeida, M., Bartolini, F., Bezák, P., Biró, M., Bjørkhaug, H., Bojnec, Brunori, G., Corazzin, M., Czekaj, M., Davidova, S., Kania, J., Kristensen, S., Marraccini, E., Molnár, Z., Niedermayr, J., O'Rourke, E., Ortiz-Miranda, D., Redman, M., Sipiläinen, T., Sooväli-Sepping, H., Šūmane, S., Surová, D., Sutherland, L.A., Tcherkezova, E., Tisenkopfs, T., Tsiligiridis, T., Tudor, M.M., Wagner, K., Wästfelt, A., 2018. Typology and distribution of small farms in Europe: towards a better picture. Land Use Pol. 75, 784–798. https://doi.org/10.1016/j.landusepol.2018.04.012.
Guiomar, N., Pinto-Correia, T., 2016. Portugal. In: Mantino, F. (Ed.), Ten Country Reports on Socio-Political, Economic and Institutional Drivers. Deliverable 3.1 of the EU-H2020 Project PEGASUS Public Ecosystem Goods and Services from Land Management – Unlocking the Synergies.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics. The Approach Based on Influence Functions. John Wiley & Sons, New York.
Haverkort, A.J., de Ruijter, F.J., van Evert, F.K., Conijn, J.G., Rutgers, B., 2014. Worldwide Sustainability Hotspots in Potato Cultivation, vol. 1, pp. 343–353. https://doi.org/10.1007/s11540-013-9247-8. Identification and mapping.
Herrero, M., Havlík, P., Valin, H., Notenbaert, A., Rufino, M.C., Thornton, P.K., Blümmel, M., Weiss, F., Grace, D., Obersteiner, M., 2013. Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. Proc. Natl. Acad. Sci. U. S. A 110, 20888–20893. https://doi.org/10.1073/pnas.1308149110.
Herrero, M., Thornton, P.K., Power, B., Bogard, J.R., Remans, R., Fritz, S., Gerber, J.S., Nelson, G., See, L., Waha, K., Watson, R.A., West, P.C., Samberg, L.H., van de Steeg, J., Stephenson, E., van Wijk, M., Havlík, P., 2017. Farming and the geography of nutrient production for human use: a transdisciplinary analysis. Lancet Planet Heal 1, e33–e42. https://doi.org/10.1016/S2542-5196(17)30007-4.
Huber, P.J., 1981. Robust Statistics. John Wiley & Sons, New York.
Iizumi, T., Ramankutty, N., 2015. How do weather and climate influence cropping area and intensity? Glob. Food Sec. 4, 46–50. https://doi.org/10.1016/j.gfs.2014.11.003.
Jiménez, D., Delerce, S., Dorado, H., Cock, J., Muñoz, L.A., Agamez, A., Jarvis, A., 2019. A scalable scheme to implement data-driven agriculture for small-scale farmers. Glob. Food Sec. 23, 256–266. https://doi.org/10.1016/j.gfs.2019.08.004.
Jin, Z., Azzari, G., You, C., di Tommaso, S., Aston, S., Burke, M., Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. Remote Sens. Environ. 228, 115–128. https://doi.org/10.1016/j.rse.2019.04.016.
Julien, J.C., Bravo-Ureta, B.E., Rada, N.E., 2019. Assessing farm performance by size in Malawi, Tanzania, and Uganda. Food Pol. 84, 153–164. https://doi.org/10.1016/j.foodpol.2018.03.016.
Kenduiywo, B.K., Bargiel, D., Soergel, U., 2018. Crop-type mapping from a sequence of Sentinel 1 images. Int. J. Rem. Sens. 39, 6383–6404. https://doi.org/10.1080/01431161.2018.1460503.
Kilic, T., Zezza, A., Carletto, C., Savastano, S., 2017. Missing(ness) in action: selectivity bias in GPS-based land area measurements. World Dev. 92, 143–157. https://doi.org/10.1016/j.worlddev.2016.11.018.
Killeen, P.R., 1985. The bimean: a measure of central tendency that accomodates outliers. Behav. Res. Methods Instrum. Comput. 17, 526–528.
Kuemmerle, T., Hostert, P., St-Louis, V., Radeloff, V.C., 2009. Using image texture to map farmland field size: a case study in Eastern Europe. J. Land Use Sci. 4, 85–107. https://doi.org/10.1080/17474230802648786.
Labarthe, P., Laurent, C., 2013. Privatization of agricultural extension services in the EU: towards a lack of adequate knowledge for small-scale farms? Food Pol. 38, 240–252. https://doi.org/10.1016/j.foodpol.2012.10.005.
Lamb, R.L., 2003. Inverse productivity: land quality, labor markets, and measurement error. J. Dev. Econ. 71, 71–95. https://doi.org/10.1016/S0304-3878(02)00134-7.
Lambert, M.J., Traoré, P.C.S., Blaes, X., Baret, P., Defourny, P., 2018. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. Remote Sens. Environ. 216, 647–657. https://doi.org/10.1016/j.rse.2018.06.036.
Lesiv, M., Carlos, J., Bayas, L., See, L., Duerauer, M., Bilous, A., Dahlia, D., Durando, N., Hazarika, R., Kumar, P., Volodymyr, V., Perez-Hoyos, A., Gengler, S., Prestele, R., Bilous, S., Karner, M., Narzary, W., Sturn, T., Mccallum, I., Schepaschenko, D., Moltchanova, E., Fraisl, D., Moorthy, I., Fritz, S., 2019. Estimating the global distribution of field size using crowdsourcing. Global Change Biol. 25, 174–186. https://doi.org/10.1111/gcb.14492.
Levin, G., 2006. Farm size and landscape composition in relation to landscape changes in Denmark. Dan. J. Geogr. 106, 45–59.
Lowder, S.K., Skoet, J., Raney, T., 2016. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. World Dev. 87, 16–29. https://doi.org/10.1016/j.worlddev.2015.10.041.
Maselli, F., Chiesi, M., Brilli, L., Moriondo, M., 2012. Simulation of olive fruit yield in Tuscany through the integration of remote sensing and ground data. Ecol. Model. 244, 1–12. https://doi.org/10.1016/j.ecolmodel.2012.06.028.
Monfreda, C., Ramankutty, N., Foley, J.A., 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochem. Cycles 22. https://doi.org/10.1029/2007GB002947. GB1022.
Muyanga, M., Jayne, T.S., 2019. Revisiting the farm size-productivity relationship based on a relatively wide range of farm sizes: evidence from Kenya. Am. J. Agric. Econ. 101, 1140–1163. https://doi.org/10.1093/ajae/aaz003.
Navarro, L., 2015. The Spanish citrus industry. Acta Hortic. 1065, 41–48. https://doi.org/10.17660/ActaHortic.2015.1065.1.
Neigh, C.S.R., Carroll, M.L., Wooten, M.R., McCarty, J.L., Powell, B.F., Husak, G.J., Enenkel, M., Hain, C.R., 2018. Smallholder crop area mapped with wall-to-wall WorldView sub-meter panchromatic image texture: a test case for Tigray, Ethiopia. Remote Sens. Environ. 212, 8–20. https://doi.org/10.1016/j.rse.2018.04.025.
Ordoudi, S.A., Bakirtzi, C., Tsimidou, M.Z., 2018. The potential of tree fruit stone and seed wastes in Greece as sources of bioactive ingredients. Recycling 3, 9. https://doi.org/10.3390/recycling3010009.
Patumi, M., D'Andria, R., Marsilio, V., Fontanazza, G., Morelli, G., Lanza, B., 2002. Olive and olive oil quality after intensive monocone olive growing (Olea europaea L. cv. Kalamata) in different irrigation regimes. Food Chem. 77, 27–34. https://doi.org/10.1016/S0308-8146(01)00317-X.
Permanhani, M., Costa, J.M., Conceição, M.A.F., de Souza, R.T., Vasconcellos, M.A.S., Chaves, M.M., 2016. Deficit irrigation in table grape: eco-physiological basis and potential use to save water and improve quality. Theor. Exp. Plant Physiol. 28, 85–108. https://doi.org/10.1007/s40626-016-0063-9.
Pieper, J.R., Brown, R.N., Amador, J.A., 2015. Effects of three conservation tillage strategies on yields and soil health in a mixed vegetable production system. Hortscience 50, 1770–1776.
Rada, N.E., Fuglie, K.O., 2019. New perspectives on farm size and productivity. Food Pol. 84, 147–152. https://doi.org/10.1016/j.foodpol.2018.03.015.
Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A., 2013. Yield trends are insufficient to double global crop production by 2050. PloS One 8. https://doi.org/10.1371/journal.pone.0066428 e66428.
Ricciardi, V., Ramankutty, N., Mehrabi, Z., Jarvis, L., Chookolingo, B., 2018. How much of our world's food do smallholders produce? Glob. Food Sec 17, 64–72. https://doi.org/10.1016/j.gfs.2018.05.002.
Rivera, M., Guarín, A., Pinto-Correia, T., Almaas, H., Arnalte-Mur, L., Burns, V., Czekaj, M., Ellis, R., Galli, F., Grivinis, M., Hernández, P., Karanikolas, P., Prosperi, P., Zamora, P.S., 2020. Assessing the role of small farms in regional food systems in Europe: evidence from a comparative study. Glob. Food Sec. 26 https://doi.org/10.1016/j.gfs.2020.100417, 100417.
Samberg, L.H., Gerber, J.S., Ramankutty, N., Herrero, M., West, P.C., 2016. Subnational distribution of average farm size and smallholder contributions to global food production. Environ. Res. Lett. 11 https://doi.org/10.1088/1748-9326/11/12/124010, 124010.
Schils, R., Olesen, J.E., Kersebaum, K.-C., Rijk, B., Oberforster, M., Kalyada, V., Khitrykau, M., Gobin, A., Kirchev, H., Manolova, V., Manolov, I., Trnka, M., Hlavinka, P., Palosuo, T., Peltonen-Sainio, P., Jauhiainen, L., Lorgeou, J., Marrou, H., Danalatos, N., Archontoulis, S., Fodor, N., Spink, J., Paolo, P., Bassu, S., Pulina, A., Seehusen, T., Kjersti, A., Żyłowska, K., Nieróbca, A., Kozyra, J., Silva, J. V., Maçãо, B.M., Coutinho, J., Ion, V., Takáč, J., Mínguez, M.I., Eckersten, H., Levy, L., Herrera, J.M., Hiltbrunner, J., Kryvobok, O., Kryvoshein, O., Sylvester-Bradley, R., Kindred, D., Topp, C.F.E., Boogaard, H., de Groot, H., Lesschen, J.P., van Bussel, L., Wolf, J., Zijlstra, M., van Loon, M.P., van Ittersum, M.K., 2018. Cereal yield gaps across Europe. Eur. J. Agron. 101, 109–120. https://doi.org/10.1016/j.eja.2018.09.003.
Sheng, Y., Ding, J., Huang, J., 2019. The relationship between farm size and productivity in agriculture: evidence from maize production in Northern China. Am. J. Agric. Econ. 101, 790–806. https://doi.org/10.1093/ajae/aay104.
Silveira, A., Ferrão, J., Muñoz-Rojas, J., Pinto-Correia, T., Guimarães, M.H., Schmidt, L., 2018. The sustainability of agricultural intensification in the early 21st century: insights from the olive oil production in Alentejo (Southern Portugal). In: Delicado, A., Domingos, N., de Sousa, L. (Eds.), Changing Societies: Legacies and Challenges. Vol. Iii. The Diverse Worlds of Sustainability. Imprensa de Ciências Sociais da Universidade de Lisboa, Lisboa, Portugal, pp. 247–275.
Stringer, R., 2016. Food security: global overview. In: Caraher, M., Coveney, J. (Eds.), Food Poverty and Insecurity: International Food Inequalities. Springer, Cham, pp. 11–18. https://doi.org/10.1007/978-3-319-23859-3_2.
Teixeira, A.H.C., Bastiaanssen, W.G.M., Bassoi, L.H., 2007. Crop water parameters of irrigated wine and table grapes to support water productivity analysis in the São Francisco river basin. Brazil. Agric. Water Manag. 94, 31–42. https://doi.org/10.1016/j.agwat.2007.08.001.
Tocco, B., Davidova, S., Bailey, A., 2013. Determinants to leave agriculture in the EU. In: Swinnen, J., Knops, L. (Eds.), Land, Labour and Capital Markets in European Agriculture: Diversity under a Common Policy. Centre for European Policy Studies, Brussels, pp. 154–170.
Tudor, M.M., 2014. Utilization of land resources in agriculture – opportunity or risk for Romanian agri-food sector competitiveness. Procedia Econ. Financ. 8, 720–728. https://doi.org/10.1016/S2212-5671(14)00150-6.
Vossen, O., 2007. Olive oil: history, production, and characteristics of the world's classic oils. Hortscience 42, 1093–1100. https://doi.org/10.21273/HORTSCI.42.5.1093.
Wassie, S.B., Abate, G.T., Bernard, T., 2019. Revisiting farm size-productivity relationship: new empirical evidence from Ethiopia. Agrekon 58, 180–199. https://doi.org/10.1080/03031853.2019.1586554.
Watson, R.A., 2017. A database of global marine commercial, small-scale, illegal and unreported fisheries catch 1950-2014. Sci. Data 4. https://doi.org/10.1038/sdata.2017.39, 170039.