

## Article

# Bootstrap Assessment of Crop Area Estimates Using Satellite Pixels Counting

Cristiano Ferraz <sup>1,\*</sup> , Jacques Delincé <sup>2</sup> , André Leite <sup>1</sup> and Raydonal Ospina <sup>1</sup>

<sup>1</sup> Computational Agricultural Statistics Laboratory—CASTLab, Federal University of Pernambuco, Recife 50740-540, Brazil; leite@castlab.org (A.L.); raydonal@castlab.org (R.O.)

<sup>2</sup> Independent Researcher, 1330 Rixensart, Belgium; jacques.delince@gmail.com

\* Correspondence: cferraz@castlab.org

**Abstract:** Crop area estimates based on counting pixels over classified satellite images are a promising application of remote sensing to agriculture. However, such area estimates are biased, and their variance is a function of the error rates of the classification rule. To redress the bias, estimators (direct and inverse) relying on the so-called confusion matrix have been proposed, but analytic estimators for variances can be tricky to derive. This article proposes a bootstrap method for assessing statistical properties of such estimators based on information from a sample confusion matrix. The proposed method can be applied to any other type of estimator that is built upon confusion matrix information. The resampling procedure is illustrated in a small study to assess the biases and variances of estimates using purely pixel counting and estimates provided by both direct and inverse estimators. The method has the advantage of being simple to implement even when the sample confusion matrix is generated under unequal probability sample design. The results show the limitations of estimates based solely on pixel counting as well as respective advantages and drawbacks of the direct and inverse estimators with respect to their feasibility, unbiasedness, and variance.

**Keywords:** error matrix; image classification; confusion matrix; resampling; inverse calibration



**Citation:** Ferraz, C.; Delincé, J.; Leite, A.; Ospina, R. Bootstrap Assessment of Crop Area Estimates Using Satellite Pixels Counting. *Stats* **2022**, *5*, 422–439. <https://doi.org/10.3390/stats5020025>

Academic Editor: Fulvia Mecatti

Received: 24 March 2022

Accepted: 22 April 2022

Published: 25 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Official agricultural statistics usually rely on field surveys based on well-designed sampling plans. Unfortunately, developing countries face financial and organizational problems in conducting such periodic inventories [1], and the current free access to satellite imagery offers an attractive complementary or even alternative solution. For many years, remote sensing has been advocated for boosting the precision of such censuses [2], using the so-called “regression” estimator to reduce the sampling variance of the field surveys. Unfortunately, the relative efficiency of the approach remained limited [3].

More recently, entire image classification at country level provided cheap crop maps, and temptation became high to just proceed to crop classes pixel counting to obtain crop areas [4]. Unfortunately, image classification is subject to errors (omissions and commissions [5] so that the results obtained are generally biased although exempt from sampling variance.

Instead of conceiving the use of imagery as a way to reduce the sampling variance of estimates derived from agricultural surveys, the idea proposed in this paper is to use the ground survey data firstly to correct the bias of an exhaustive image classification and secondly to propagate the errors of the classification rule to derive precision for the crop area estimates.

In 1982, the direct estimator based on image classification was proposed [6]. Later, in 1988, the inverse estimator was introduced [7], leading to a discussion about how direct and inverse estimators should be chosen [8]. Both can be used to redress the biased “pixel counting” estimators by using the so-called confusion matrix of the classification rule.

However, the problem of assessing their bias level as well as the variance of their estimates has not been addressed yet. In practice, no attempt to report variances is made because no variance formula is provided by [6–8]. The fact that the choice of the appropriate estimator depends on the sampling approach to the ground data collection has also not been discussed yet, representing another literature gap. In this paper, direct and inverse estimators are reviewed, and a discussion about how their feasibility in practice depends on the sampling strategy to collect data on the ground is provided. In addition, bootstrap is proposed as a statistical resampling method useful to assess both bias and variance of the direct and inverse estimators. A bootstrap algorithm based on information from a sample confusion matrix is introduced so as to properly consider the sampling strategy used to generate the sample data used by the estimators. The proposed method can be applied to any other type of estimator that is built upon confusion matrix information.

This paper is structured in six sections, including this introduction. Section 2 introduces types of errors and the direct and inverse calibration estimators. Section 3 discusses the feasibility of using the considered estimators in practice depending on how ground data are collected. Section 4 introduces a brief review on the bootstrap method and presents a bootstrap algorithm for assessing crop area estimates produced via confusion matrices. Section 5 illustrates the application of the proposed resampling method to assess statistical properties of estimates based only on pixel counting and estimates provided by the direct and inverse estimators. Section 6 includes concluding remarks.

## 2. Remote Sensing Estimates

Estimation of crop areas using pixel counting is subject to at least two sources of errors: mixed-borders pixels and misclassification of pure pixels. Considering the territory of interest is completely covered by satellite imagery, and assuming the effect of mixed border pixels can be neglected, the bias due to misclassification of pure pixel counting on crop area estimates can be defined with respect to an error matrix.

Consider the classification of images of the whole territory of interest leading to the identification of  $M$  classes of land covering so that  $R = (A_{+1}, A_{+2}, \dots, A_{+M})'$  is an  $M \times 1$  column vector with the total area of pixels classified in each type of class. The actual areas are represented by  $T = (A_{1+}, A_{2+}, \dots, A_{M+})'$ , an  $M \times 1$  column vector with the truth total area of classes found on the ground. The error matrix, also called confusion matrix, is an  $M \times M$  matrix  $A$  so that its elements are areas classified according to remote sensing image and ground truth:

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1M} \\ A_{21} & A_{22} & \cdots & A_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MM} \end{pmatrix}.$$

The elements of the error matrix  $A$ , denoted by  $A_{gc}$ , are areas of class  $g$  (ground), also classified as class  $c$  by pixel counting. All these components can be arranged into a table with the same structure as the partitioned matrix  $Q$ , which is given by:

$$Q = \begin{pmatrix} A & T \\ R' & A_{++} \end{pmatrix}_{(M+1) \times (M+1)}.$$

Table 1 illustrates an example of such a matrix in tabular form, with only three classes of land cover: wheat, corn, and soy. Such structure assumes the image classification algorithm is capable of correctly identifying the classes on the ground (absence of clouds).

**Table 1.** Example of a Q matrix in table form.

Crop Area Classification		Remote Sensing Classification			Total
		Wheat	Corn	Soy	
Ground truth classes	Wheat	$A_{11}$	$A_{12}$	$A_{13}$	$A_{1+}$
	Corn	$A_{21}$	$A_{22}$	$A_{23}$	$A_{2+}$
	Soy	$A_{31}$	$A_{32}$	$A_{33}$	$A_{3+}$
Total		$A_{+1}$	$A_{+2}$	$A_{+3}$	$A_{++}$

In Table 1,  $A_{11}$ ,  $A_{22}$ , and  $A_{33}$  represent area estimates of wheat, corn, and soy, respectively, using pixels correctly classified as the corresponding ground truth data. In addition,  $A_{21}$  and  $A_{31}$  represent area estimates of wheat based on pixel counting that are indeed areas of corn and soy, respectively. The summation  $(A_{21} + A_{31}) = \Phi_1$  is called the commission error related to the wheat area estimate. On the other hand,  $A_{12}$  and  $A_{13}$  are wheat ground area that were mistakenly estimated as corn and soy areas, respectively, using pixel counting. The summation  $(A_{12} + A_{13}) = \Psi_1$  is called the omission error related to wheat estimate. The vectors  $R'$  and  $T$  are, respectively, the row and the column marginals of the Table 1, while  $A_{++}$  represents the total number of pixels classified over the considered territory.

*Total Area and Bias Estimation*

Error matrix  $A$  and column vector  $T$  are unknown in practice.  $T$  is the parameter of interest. Estimating the total crop areas of  $T$  based only on  $R$  is subject to a bias given by the difference between the commission and the omission errors. Let  $R_c$  represent the crop  $c$  area estimator based solely on pixel counting and  $T_c$  be the ground truth area of the same crop. Define  $B_c = B_c(R_c, T_c)$  as the bias of the estimator  $R_c$  for the total area  $T_c$ . Let  $\Phi_c = A_{+c} - A_{cc}$  and  $\Psi_c = A_{c+} - A_{cc}$  be the commission and the omission error related to such estimate, respectively. Then,

$$B_c = \Phi_c - \Psi_c = A_{+c} - A_{c+}. \tag{1}$$

Denoting by  $B = (B_1, B_2, \dots, B_M)'$ , the column vector of biases for each of the  $M$  classes' estimates, one can write:

$$B = R - T. \tag{2}$$

Define  $D_R = \text{diag}(A_{+1}, A_{+2}, \dots, A_{+M})$  as the diagonal matrix with the  $R$  information in the main diagonal. Let the following relative error matrix be defined:

$$E_{g|c} = AD_R^{-1} = \begin{pmatrix} A_{11}/A_{+1} & A_{12}/A_{+2} & \dots & A_{1M}/A_{+M} \\ A_{21}/A_{+1} & A_{22}/A_{+2} & \dots & A_{2M}/A_{+M} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1}/A_{+1} & A_{M2}/A_{+2} & \dots & A_{MM}/A_{+M} \end{pmatrix};$$

$E_{g|c}$  is a matrix with conditional probabilities that a pixel is over the ground truth class  $g$  given the pixel is classified as  $c$ . Its column vectors are the relative columns frequencies of the table based on matrix  $Q$ . Hence,

$$T = E_{g|c}R. \tag{3}$$

Based on expression (3), if  $\hat{E}_{g|c}$  denotes an unbiased estimator for  $E_{g|c}$ , then one can use

$$\hat{T}_1 = \hat{E}_{g|c}R \tag{4}$$

as an unbiased estimator of  $T$ .

Following a similar reasoning, let  $D_T = \text{diag}(A_{1+}, A_{2+}, \dots, A_{M+})$  be the diagonal matrix with the  $T$  information in the main diagonal and  $E_{c|g}$  the relative error matrix given by:

$$E_{c|g} = A'D_T^{-1} = \begin{pmatrix} A_{11}/A_{1+} & A_{21}/A_{2+} & \dots & A_{M1}/A_{M+} \\ A_{12}/A_{1+} & A_{22}/A_{2+} & \dots & A_{M2}/A_{M+} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1M}/A_{1+} & A_{2M}/A_{2+} & \dots & A_{MM}/A_{M+} \end{pmatrix};$$

$E_{c|g}$  is a matrix with conditional probabilities that a pixel is classified as class  $c$  given the ground truth class is  $g$ . Its column vectors are relative row frequencies of the table based on matrix  $Q$ . Therefore,

$$R = E_{c|g}T. \tag{5}$$

If  $E_{c|g}$  is non-singular, then it is possible to write

$$T = E_{c|g}^{-1}R. \tag{6}$$

If  $\hat{E}_{c|g}^{-1}$  is an unbiased estimator for  $E_{c|g}^{-1}$ , one can use

$$\hat{T}_2 = \hat{E}_{c|g}^{-1}R \tag{7}$$

as an unbiased estimator for  $T$ .

Approximately unbiased estimators  $\hat{E}_{g|c}$  and  $\hat{E}_{c|g}^{-1}$  can be defined depending on the availability of further data. Suppose ground information can be observed for a sample of  $n$  test points, following Gallego's recommendations [5] (p. 252) so that a sample  $M \times M$  error matrix  $a$  is available:

$$a = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1M} \\ a_{21} & a_{22} & \dots & a_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \dots & a_{MM} \end{pmatrix}.$$

The elements  $a_{gc}$  of the sample error matrix  $a$  are areas of class  $g$  (ground) classified as class  $c$  by pixel counting over a set of sampling testing points.  $a$  provides information that can be used to estimate  $A$ .

Let  $r = (a_{+1}, a_{+2}, \dots, a_{+M})'$  be an  $M \times 1$  column vector with the total area of sample test pixels classified in each type of class. Let  $t = (a_{1+}, a_{2+}, \dots, a_{M+})'$  be an  $M \times 1$  column vector with the truth total areas of classes found on the sample ground points. All these components can be arranged into a table with the same structure as the partitioned matrix  $q$ , which is given by:

$$q = \begin{pmatrix} a & t \\ r' & a_{++} \end{pmatrix}_{(M+1) \times (M+1)},$$

with  $a_{++} = n$ .

Define the following diagonal matrices based on  $q$ :  $D_r = \text{diag}(a_{+1}, a_{+2}, \dots, a_{+M})$  and  $D_t = \text{diag}(a_{1+}, a_{2+}, \dots, a_{M+})$ . Then, one can define  $\hat{E}_{g|c} = e_{g|c}$ , and  $\hat{E}_{c|g}^{-1} = e_{c|g}^{-1}$  so that:

$$e_{g|c} = aD_r^{-1} = \begin{pmatrix} a_{11}/a_{+1} & a_{12}/a_{+2} & \dots & a_{1M}/a_{+M} \\ a_{21}/a_{+1} & a_{22}/a_{+2} & \dots & a_{2M}/a_{+M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}/a_{+1} & a_{M2}/a_{+2} & \dots & a_{MM}/a_{+M} \end{pmatrix},$$

and

$$e_{c|g} = a'D_t^{-1} = \begin{pmatrix} a_{11}/a_{1+} & a_{21}/a_{2+} & \dots & a_{M1}/a_{M+} \\ a_{12}/a_{1+} & a_{22}/a_{2+} & \dots & a_{M2}/a_{M+} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1M}/a_{1+} & a_{2M}/a_{2+} & \dots & a_{MM}/a_{M+} \end{pmatrix}.$$

The two estimators previously written as  $\hat{T}_1$  and  $\hat{T}_2$  assume the form of the known direct and inverse calibration type estimators, respectively, defined by:

The direct calibration estimator:

$$\hat{T}_{Direct} = e_{g|c}R = \begin{pmatrix} a_{11}/a_{+1} & a_{12}/a_{+2} & \dots & a_{1M}/a_{+M} \\ a_{21}/a_{+1} & a_{22}/a_{+2} & \dots & a_{2m}/a_{+M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}/a_{+1} & a_{M2}/a_{+2} & \dots & a_{MM}/a_{+M} \end{pmatrix} \begin{pmatrix} A_{+1} \\ A_{+2} \\ \vdots \\ A_{+M} \end{pmatrix}; \tag{8}$$

The inverse calibration estimator:

$$\hat{T}_{Inverse} = e_{c|g}^{-1}R = \begin{pmatrix} a_{11}/a_{1+} & a_{21}/a_{2+} & \dots & a_{M1}/a_{M+} \\ a_{12}/a_{1+} & a_{22}/a_{2+} & \dots & a_{M2}/a_{M+} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1M}/a_{1+} & a_{2M}/a_{2+} & \dots & a_{MM}/a_{M+} \end{pmatrix}^{-1} \begin{pmatrix} A_{+1} \\ A_{+2} \\ \vdots \\ A_{+M} \end{pmatrix}, \tag{9}$$

for nonsingular  $e_{c|g}$ .

The bias  $B$  for using only  $R$  as crop area estimates for  $T$  can then be assessed depending on the calibration estimator used:

$$\hat{B}_{Direct} = R - \hat{T}_{Direct}; \tag{10}$$

or

$$\hat{B}_{Inverse} = R - \hat{T}_{Inverse}. \tag{11}$$

$\hat{T}_{Direct}$  and  $\hat{T}_{Inverse}$  are nonlinear estimators. Analytical expressions for their variances depend on how the sample of test points that generated the matrix  $q$  was selected and may not be simple to derive. Although Taylor linearization technique could be applied to assess the direct estimator, assessment of the inverse estimator variance is not a simple task. In this paper, a bootstrap algorithm to estimate the variances of both estimators is introduced, which takes into account unequal probability sample designs. The same algorithm idea can be adapted to assess any estimator built upon confusion matrices' information.

### 3. Feasibility of Estimators in Practice

Direct and inverse calibration estimators rely on estimates of conditional probabilities  $\hat{E}_{g|c} = e_{g|c}$  and  $\hat{E}_{c|g}^{-1} = e_{c|g}^{-1}$ , respectively.

Let  $P(G = i, C = j)$ ,  $i = 1, 2, \dots, M, j = 1, 2, \dots, M$  be the joint probability of a pixel being on the ground of crop  $i$  and being classified as crop  $j$ . Let  $P(G = i)$  be the probability that a pixel is over a ground truth crop  $i$  and  $P(C = j)$  be the probability that a pixel is classified as crop  $j$ . Then, it is possible to write:

$$P(G = i) = \sum_{j=1}^M P(G = i, C = j) = \sum_{j=1}^M P(G = i|C = j)P(C = j); \tag{12}$$

$$P(C = j) = \sum_{i=1}^M P(G = i, C = j) = \sum_{i=1}^M P(C = j|G = i)P(G = i). \tag{13}$$

The direct calibration estimator defined in (8) relies on the relationship between the joint probabilities and the conditional probabilities expressed by (12), while the inverse calibration estimator defined in (9) relies on the relationship described in Equation (13).

The need to reach the joint distribution  $P(G = i, C = j)$  is addressed through the conditional probabilities  $P(G = i|C = j)$  and  $P(C = j|G = i)$ , estimated by a sample of test points. In practice, the following strategies to collect testing points could be considered:

*Strategy 1. Bivariate classification of points (Bivariate):* To randomly select a set of geographical coordinates in the region of interest and then to observe their category in the image (image classification) and over the field (ground truth).

The Bivariate strategy of classification provides direct information of the joint probabilities  $P(G = i, C = j)$ , allowing the choice of using either  $\hat{T}_{Direct}$  or  $\hat{T}_{Inverse}$ . However, field work costs involved with this procedure may lead to its non-feasibility in practice.

*Strategy 2. Classification by Remote Sensing (RS):* To use stratified sampling by randomly selecting a set of pixels in each classified image category and to later observe their ground truth class in the field.

The RS strategy of classification can provide information about the conditional probabilities  $P(G = i|C = j)$ , allowing the use of the direct estimator  $\hat{T}_{Direct}$ . However, such strategy may also represent a challenge to be implemented in practice due to field costs. If this strategy is used, it is not possible to estimate  $P(C = i|G = j)$ , and so the inverse estimator has no theoretical basis to be chosen.

*Strategy 3. Classification by Ground (G):* To use stratified sampling by randomly selecting a set of points over each ground truth category in the field and to later check about their image classification.

The G strategy of classification provides information about the conditional probabilities  $P(C = j|G = i)$  so that the inverse calibration estimator  $\hat{T}_{Inverse}$  can be used. If this strategy is implemented, it is not possible to estimate  $P(G = i|C = j)$ , so there is no theoretical justification to choose the direct estimator even though considering organizing the field work this way tends to be more cost-effective. Therefore, the choice of the G strategy of classification seems to be the one with most appeal to be used in practice.

It is possible that both direct and inverse estimators could suffer instability when the number of pixels observed is not large enough [8]. Agreement between the interpretation of the classification rules and the surveyed ground classes is essential. Inverse estimator is not feasible if nearly singular sample confusion matrices are observed. This can be a consequence, for example, of a poor classification algorithm. Inversibility of the sample confusion matrix is ensured when the classification rule is deemed minimally acceptable in practice [9] so that:

$$P(C = i|G = i) > 0.5 \text{ for all } i = 1, 2, \dots, M.$$

Further, caution is also needed when the image classification imposes the use of more categories (i.e., a class corresponding to cloudy areas) than the reality seen on the ground, leading to rectangular sample confusion matrices. Inverse estimators based on completing such matrices with zeroes are not guaranteed to work. In addition, an extra source of care is needed concerning the sample design used to select the testing points. Strategies 1 to 3 mention the use of random samples in the sense that probability sample designs are employed. Such designs may range from equal probability sampling to more complex selection methods. The bootstrap method proposed in this paper considers the possibility that an unequal probability sample design is used to generate the sample confusion matrix. In case of unequal probability sampling, the direct and inverse estimators also have to be weighted in function of inclusion probabilities. In this case, the sample confusion matrix  $a = \{\hat{a}_{gc}\}$  must be composed by design-consistent estimators using sampling weights defined as the inverse of inclusion probabilities  $\pi_k$ .

#### 4. Bootstrap Resampling

Bootstrap is a computer-intensive statistical methodology introduced by Efron in 1979 [10], which replaces complex analytical procedures by computer intensive empirical analysis. It relies on Monte Carlo method, where several random resamples are drawn from a given original sample. The bootstrap method has been applied in a variety of situations (e.g., [11–14]). Several authors provide comprehensive discussion of the bootstrap method. Beaumont [15], Efron and Tibshirani [16], Hersterber [17], and Shao and Tu [18] are some of them.

Consider a random sample  $\mathbf{y} = (y_1, \dots, y_n)^T$  of size  $n$ , where each element is a random draw from the random variable  $Y$ , which has the distribution function  $F = F(\theta)$ , and  $\theta$  is the parameter that indexes the distribution. Here,  $\theta$  is viewed as a functional of  $F$ , i.e.,  $\theta = T(F)$ . Let  $\hat{\theta}$  be an estimator of  $\theta$  based on  $\mathbf{y}$  so that it is possible to write  $\hat{\theta} = S(\mathbf{y})$ . The application of the bootstrap method consists in obtaining a large number of pseudo-samples  $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^T$  from the original sample  $\mathbf{y}$  and then extracting information from these pseudo-samples to improve inference.

In principle, there are two different ways of obtaining and evaluating bootstrap estimates: non-parametric bootstrap, which does not assume any distribution of the population, and parametric bootstrap, which assumes a particular distribution for the sample at hand [19]. In the parametric version, the bootstrap samples are obtained from  $F(\hat{\theta})$ , which is expressed here as  $F_{\hat{\theta}}$ , whereas in the nonparametric version, they are obtained from the empirical distribution function  $\hat{F}$  through sampling with replacement. The nonparametric bootstrap does not entail parametric assumptions.

Let  $B_F(\hat{\theta}, \theta)$  be the bias of the estimator  $\hat{\theta} = S(\mathbf{y})$ ; that is,

$$B_F(\hat{\theta}, \theta) = E_F(\hat{\theta} - \theta) = E_F[S(\mathbf{y})] - T(F), \quad (14)$$

where the subscript  $F$  indicates that expectation is taken with respect to  $F$ . The bootstrap estimators of the bias in the parametric and nonparametric versions are obtained by replacing the true distribution  $F$ , which generated the original sample, with  $F_{\hat{\theta}}$  and  $\hat{F}$ , respectively, in (13). Therefore, the parametric and nonparametric estimates of the bias are given, respectively, by:

$$B_{F_{\hat{\theta}}}(\hat{\theta}, \theta) = E_{F_{\hat{\theta}}}[S(\mathbf{y})] - T(F_{\hat{\theta}}), \quad (15)$$

and

$$B_{\hat{F}}(\hat{\theta}, \theta) = E_{\hat{F}}[S(\mathbf{y})] - T(\hat{F}). \quad (16)$$

If  $B$  bootstrap samples  $(\mathbf{y}^{*1}, \mathbf{y}^{*2}, \dots, \mathbf{y}^{*B})$  are generated independently from the original sample  $\mathbf{y}$ , and the respective bootstrap replications  $(\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*B})$  are calculated where  $\hat{\theta}^{*b} = S(\mathbf{y}^{*b})$ ,  $b = 1, 2, \dots, B$ , then it is possible to approximate the bootstrap expectations  $E_{F_{\hat{\theta}}}[S(\mathbf{y})]$  and  $E_{\hat{F}}[S(\mathbf{y})]$  by the average

$$\hat{\theta}^{*(\cdot)} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{*b}.$$

Therefore, the bootstrap bias estimates based on  $B$  replications of  $\hat{\theta}$  are:

$$\hat{B}_{F_{\hat{\theta}}}(\hat{\theta}, \theta) = \hat{\theta}^{*(\cdot)} - S(\mathbf{y}), \quad (17)$$

and

$$\hat{B}_{\hat{F}}(\hat{\theta}, \theta) = \hat{\theta}^{*(\cdot)} - S(\mathbf{y}), \quad (18)$$

for the parametric and nonparametric versions, respectively. By using the two bootstrap bias estimators above, it is possible to obtain estimates that are bias-corrected up to order  $O(n^{-1})$ . In addition, bootstrap variances can be assessed using:

$$\hat{V}(\hat{\theta}^{*(\cdot)}) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^{*b} - \hat{\theta}^{*(\cdot)})^2. \tag{19}$$

*A Bootstrap Algorithm for Crop Area Estimates' Assessment*

Monte Carlo simulation based on artificial data has been used to compare the performances of crop area estimators [20]. Bootstrap has been applied to assess accuracy of agricultural land classifications relying on resampling points on the entire territory of interest [21]. In this paper, the interest is to explore the potential for using bootstrap methods to assess crop estimates generated by estimators that are function of confusion (error) matrices, such as the direct and the inverse estimators. Hence, a bootstrap algorithm is proposed, built upon the use of confusion matrix sample information.

Define  $U$  as the set of all  $N$  pixel points needed to cover the entire territory of interest. The elements (pixel points) of such set need not be identifiable, but they are rather deemed to be used to build the matrix  $Q$ .

Direct and inverse estimators presented depend upon data provided by a sample  $S \subset U$  of  $n$  testing pixel points to compose a sample confusion matrix  $a$ . Consider sample  $S$  is selected from  $U$  using one of the three strategies listed on the last section by a probability sampling design  $p(\cdot)$ . Define  $I_k = 1$  if pixel  $k \in S$  and  $I_k = 0$  if pixel  $k \notin S$  so that  $\pi_k = P(I_k = 1)$  is the first order inclusion probability for  $k \in U$ .

Let  $\hat{Q}(S) = q_0$  be the sample matrix estimate for the matrix  $Q$  based on the probability sampling design  $p(\cdot)$  so that:

$$q_0 = \begin{pmatrix} a_0 & t_0 \\ r_0^T & a_{++}^0 \end{pmatrix}.$$

$a_0 = \{\hat{a}_{gc}\}$  is the sample confusion matrix built upon the classification of all  $n$  pixel points in  $S$ . Let  $\gamma_k(g, c) = 1$  if pixel  $k$  is classified as crop  $g$  on the ground and class  $c$  by the satellite image; otherwise,  $\gamma_k(g, c) = 0$ . Then, each element of  $a_0$  can be written as:

$$\hat{a}_{gc} = \sum_{k \in S} \frac{\gamma_k(g, c)}{\pi_k} = \sum_{k \in U} \frac{\gamma_k(g, c)}{\pi_k} I_k. \tag{20}$$

In 2020, Conti et al. provided the asymptotic theory needed to support selecting bootstrap samples in two phases where, in the first phase, a pseudo-population  $U^*$  is predicted and calibrated to the size of the original population  $U$ , and in the second phase, bootstrap resamples are selected conditionally from  $U^*$  based on the original sampling design  $p(\cdot)$  [22]. The following are proposed algorithms adapted from them to fit the crop area estimation scenario:

*Bootstrap Algorithm to be Used with a Bivariate Strategy of Point Classification:*

**Step 1.** *Build a multinomial pseudo-population:*

Let  $i = 1, 2, \dots, N$  be a sequence of independent trials where for each  $i$ , a pixel  $k \in S$  is selected for the pseudo-population with probability  $p_k = \pi_k / \sum_{j \in S} \pi_j$ . If pixel  $k$  is selected for the pseudo-population, information regarding the two-way classification of pixel  $k$  is also retained. Let  $\delta_{ik} = 1$  if pixel  $k \in S$  is selected at the  $i$ -th trial; otherwise,  $\delta_{ik} = 0$ . Recall that  $\gamma_k(g, c) = 1$  if pixel  $k$  is classified as crop  $g$  on the ground and class  $c$  by the satellite image; otherwise,  $\gamma_k(g, c) = 0$ . Then, the retained two-way classification of a pixel  $i \in U^*$ ,  $\gamma_i(g, c)$  can be written as:

$$\gamma_i(g, c) = \sum_{k \in S} \delta_{ik} \gamma_k(g, c). \tag{21}$$

The pseudo-population  $U^*$  is then composed by the set of  $N$  pixels selected from  $S$  and their respective classification  $\gamma_i(g, c)$ , for  $i = 1, 2, \dots, N$ .

**Step 2.** Select the  $b$ -th bootstrap sample:

Select a probability sample of  $n$  pixels from the pseudo-population  $S_b \subset U^*$  using the same sample design  $p(\cdot)$  used to generate the original sample  $S \subset U$ . This means to select  $S_b$  with inclusion probability  $\pi_i = \sum_{k \in S} \delta_{ik} \pi_k$ . Keep the values  $\gamma_i(g, c)$  for those  $i \in S_b$ .

**Step 3.** Build the  $b$ -th bootstrap sample confusion matrix  $a_b = \{a_{gc}^{(b)}\}$ :

$$a_{gc}^{(b)} = \sum_{i \in S_b} \frac{\gamma_i(g, c)}{\pi_i}, \tag{22}$$

and calculate:

$$r_b^T = (a_{+1}^{(b)}, a_{+2}^{(b)}, \dots, a_{+M}^{(b)}) = \sum_{g=1}^M (a_{g1}^{(b)}, a_{g2}^{(b)}, \dots, a_{gM}^{(b)}),$$

$$t_b = (a_{1+}^{(b)}, a_{2+}^{(b)}, \dots, a_{M+}^{(b)}) = \sum_{c=1}^M (a_{1c}^{(b)}, a_{2c}^{(b)}, \dots, a_{Mc}^{(b)}),$$

and  $a_{++}^{(b)} = \sum_{g=1}^M \sum_{c=1}^M a_{gc}^{(b)}$ .

**Step 4.** Calculate the  $b$ -th bootstrap sample conditional probability matrices:

$$e_{g|c}^b = \{p_{g|c}^{(b)}\} = \left\{ \frac{a_{gc}^{(b)}}{a_{+c}^{(b)}} \right\}, \tag{23}$$

$$e_{c|g}^b = \{p_{c|g}^{(b)}\} = \left\{ \frac{a_{gc}^{(b)}}{a_{g+}^{(b)}} \right\}, \tag{24}$$

for  $g = 1, 2, \dots, M$  and  $c = 1, 2, \dots, M$ .

**Step 5.** Calculate the  $b$ -th bootstrap estimates and keep their values:

Let  $R_0$  be the observed crop area estimates for the territory of interest based solely on pixel counting from satellite image.  $R_0$  is the only component of  $Q$  that is known:

$$Q = \begin{pmatrix} A & T \\ R_0' & A_{++} \end{pmatrix}.$$

Calculate the  $b$ -th bootstrap estimates using:

$$\hat{T}_{Direct}^{(b)} = e_{g|c}^b R_0, \tag{25}$$

and

$$\hat{T}_{Inverse}^{(b)} = e_{c|g}^b{}^{-1} R_0. \tag{26}$$

**Step 6.** Repeat steps 2 to 5 for  $b = 1, 2, \dots, B$ , where  $B$  is the desired number of bootstrap replicated samples.

**Step 7.** Calculate the bootstrap estimates and respective variances:

$$\hat{T}_{Direct}^{Bootstrap} = \frac{1}{B} \sum_{b=1}^B \hat{T}_{Direct}^{(b)}; \tag{27}$$

$$Var_{Bootstrap}(\hat{T}_{Direct}) = \frac{1}{B-1} \sum_{b=1}^B (\hat{T}_{Direct}^{(b)} - \hat{T}_{Direct}^{Bootstrap})^2. \tag{28}$$

$$\hat{T}_{Inverse}^{Bootstrap} = \frac{1}{B} \sum_{b=1}^B \hat{T}_{Inverse}^{(b)}; \tag{29}$$

$$\widehat{Var}_{Bootstrap}(\widehat{T}_{Inverse}) = \frac{1}{B-1} \sum_{b=1}^B \left( \widehat{T}_{Inverse}^{(b)} - \widehat{T}_{Inverse}^{Bootstrap} \right)^2. \tag{30}$$

One should note that the possibility of calculating the two bootstrap estimators, direct and inverse, relies on the fact that a Bivariate strategy of classification of testing points is used, as described in Section 3. However, as discussed in the same section, a G strategy of classification of points is the one that is feasible in practice. If the G strategy is adopted, the bootstrap algorithm steps 1 and 2 must be implemented for each ground category  $g = 1, 2, \dots, M$  independently. The modified steps 1 and 2 can be written as:

*Modified Bootstrap Algorithm Steps for Using with G strategy of Point Classification:*

**Step 1.** Build a multinomial-product pseudo-population for a G strategy:

Let  $S_g$  be the sample of  $n_g = n/M$  test points selected using the G strategy to compose the sample confusion matrix  $a$  for  $g = 1, \dots, M$  based on the probability sample design  $p_g(\cdot)$ . Let  $i = 1, 2, \dots, N/M$  be a sequence of independent trials, where for each  $i$ , a pixel  $k \in S_g$  is selected for the pseudo-population stratum  $U_g$  with probability  $p_k = \pi_k / \sum_{j \in S_g} \pi_j$ . If pixel  $k$  is selected for the pseudo-population stratum  $U_g$ , information regarding the two-way classification of pixel  $k$  is also retained. Let  $\delta_{ik} = 1$  if pixel  $k \in S_g$  is selected at the  $i$ -th trial, and  $\delta_{ik} = 0$  otherwise. Recall that  $\gamma_k(g, c) = 1$  if pixel  $k$  is classified as crop  $g$  on the ground and class  $c$  by the satellite image; and  $\gamma_k(g, c) = 0$  otherwise. Then, the retained two-way classification of a pixel  $i \in U_g^*$ ,  $\gamma_i(g, c)$  can be written as:

$$\gamma_i(g, c) = \sum_{k \in S_g} \delta_{ik} \gamma_k(g, c). \tag{31}$$

The pseudo-population stratum  $U_g^*$  is then composed by the set of  $N/M$  pixels selected from  $S_g$  and their respective classification  $\gamma_i(g, c)$  for  $i = 1, 2, \dots, N/M$ . The pseudo-population is so that  $U^* = \cup_{g=1}^{N/M} U_g^*$ , with all two-way classification information retained.

**Step 2.** Select the  $b$ -th bootstrap sample under the G strategy:

Let  $U_g$  and  $U_g^*$  be the set of pixels in the population stratum  $g$  and pseudo-population stratum  $g$ , respectively. For each  $g = 1, 2, \dots, M$ , select a probability sample of  $n/M$  pixels  $S_g^{(b)} \subset U_g^*$  using the same sample design  $p_g(\cdot)$  used to generate the original sample  $S_g \subset U_g$ . This means to select  $S_g^{(b)}$  from  $U_g^*$  with inclusion probability

$$\pi_{i|g} = \sum_{k \in S_g^{(b)}} \delta_{ik} \pi_k,$$

and keep the values  $\gamma_i(g, c)$  for those  $i \in S_g^{(b)}$ .

In this case, only the bootstrap inverse estimator makes sense to be calculated. Therefore, in step 4, only Equation (24) should be calculated, and in step 5, only the inverse estimator defined by Equation (26) should be calculated. In step 7, only Equations (29) and (30) apply.

If the RS strategy is used instead, then the bootstrap algorithm steps 1 and 2 must be implemented for each class category  $c = 1, 2, \dots, M$  independently. The modified steps 1 and 2 are given by:

*Modified Bootstrap Algorithm Steps for Using with RS strategy of Point Classification:*

**Step 1.** Build a multinomial-product pseudo-population for an RS strategy:

Let  $S_c$  be the sample of  $n_c = n/M$  test points selected using the RS strategy to compose the sample confusion matrix  $a$  for  $c = 1, \dots, M$  based on the probability sample design  $p_c(\cdot)$ . Let  $i = 1, 2, \dots, N/M$  be a sequence of independent trials, where for each  $i$ , a pixel  $k \in S_c$  is selected for the pseudo-population stratum  $U_c$  with probability  $p_k = \pi_k / \sum_{j \in S_c} \pi_j$ . If pixel  $k$  is selected for the pseudo-population stratum  $U_c$ , information regarding the two-way classification of pixel  $k$  is also retained. Let  $\delta_{ik} = 1$  if pixel  $k \in S_c$  is selected at the  $i$ -th trial, and  $\delta_{ik} = 0$  otherwise. Recall that  $\gamma_k(g, c) = 1$  if pixel  $k$  is classified as crop  $c$  on the ground

and class  $c$  by the satellite image; and  $\gamma_k(g, c) = 0$  otherwise. Then, the retained two-way classification of a pixel  $i \in U_c^*$ ,  $\gamma_i(g, c)$  can be written as:

$$\gamma_i(g, c) = \sum_{k \in S_c} \delta_{ik} \gamma_k(g, c). \tag{32}$$

The pseudo-population stratum  $U_c^*$  is then composed by the set of  $N/M$  pixels selected from  $S_c$  and their respective classification  $\gamma_i(g, c)$  for  $i = 1, 2, \dots, N/M$ . The pseudo-population is so that  $U^* = \cup_{c=1}^{N/M} U_c^*$ , with all two-way classification information retained.

**Step 2.** Select the  $b$ -th bootstrap sample under the RS strategy:

Let  $U_c$  and  $U_c^*$  be the set of pixels in the population and pseudo-population, respectively, with image classification  $c$ . For each  $c = 1, 2, \dots, M$ , select a probability sample of  $n/M$  pixels  $S_c^{(b)} \subset U_c^*$  using the same sample design  $p(\cdot)$  used to generate the original sample  $S_c \subset U_c$ . This means to select  $S_c^{(b)}$  from  $U_c^*$  with inclusion probability

$$\pi_{i|c} = \sum_{k \in S_c^{(b)}} \delta_{ik} \pi_k.$$

Keep the values  $\gamma_i(g, c)$  for those  $i \in S_c^{(b)}$ .

In this case, only the bootstrap direct estimator makes sense to be calculated. Therefore, in step 4, only Equation (23) should be calculated, and in step 5, only the inverse estimator defined by Equation (25) should be calculated. In step 7, only Equations (27) and (28) apply.

### 5. Application and Results

The proposed method is illustrated using data built upon the joint probabilities described in Table 2. The numbers correspond to a scenario where the crop areas of wheat, rapeseed, corn, sugar beet, and others are present in a given territory in the proportions of 25%, 5%, 10%, 20%, and 40%, respectively. The classification rule is such that  $0.2/0.25 = 0.8$  of the area with wheat is correctly classified. The proportion of correct classification for rapeseed, corn, sugar beet, and others are 0.6, 0.8, 0.6, and 0.6, respectively, all satisfying the condition to be practically minimally acceptable [9].

**Table 2.** Joint probabilities for the artificial data set.

Crop Area Classification		Remote Sensing Classification					Total
		Wheat	Rapeseed	Corn	Sugar Beet	Others	
Ground truth classes	Wheat	0.2	0.02	0.005	0.005	0.02	0.25
	Rapeseed	0.01	0.03	0	0	0.01	0.05
	Corn	0.001	0.01	0.08	0.005	0.004	0.1
	Sugar beet	0.005	0.015	0.04	0.12	0.02	0.2
	Others	0.1	0.02	0.01	0.03	0.24	0.4
Total		0.316	0.095	0.135	0.16	0.294	1

It is assumed the area of the territory of interest is covered by 1 million pixels ( $N$ ) that are classified generating an estimate of areas by pixel counting so that the proportion of area cultivated by wheat is 31.6%, by rapeseed 9.5%, by corn 13.5%, by sugar beet 16%, and by other crops (others) 29.4%. Such estimates reveal the classification rule overestimates wheat, rapeseed, and corn and underestimates sugar beet and others. The set of 1 million pixels is the population  $U$ , and the estimates based purely on pixel counting for this population is the  $R$  vector that corresponds to the marginal column of Table 2 multiplied by 1 million.

Assessment of purely pixel counting estimates and estimates generated by direct and inverse estimators was done for the three strategies of collecting test points, as described in Section 3.

Bivariate strategy was first investigated using simple random sampling of a thousand pixels that are classified based on the ground and based on the satellite image, generating Table 3.

**Table 3.** Test points classification using Bivariate strategy, with 1000 pixels selected by simple random sampling.

Crop Area Classification		Remote Sensing Classification				
		Wheat	Rapeseed	Corn	Sugar Beet	Others
Ground truth classes	Wheat	201	23	6	3	19
	Rapeseed	11	36	0	0	7
	Corn	4	8	82	6	5
	Sugar beet	4	17	38	117	19
	Others	108	31	7	29	219

Table 3 information is used to generate a sample confusion matrix  $a$  with crop area estimates. For example, the confusion matrix area for the classification  $g = 1, c = 1$ , is provided by:

$$\hat{a}_{11} = 1,000,000 \times \sum_{i \in S} \frac{\gamma_i(1,1)}{1000} = 1000 \times 201 = 201,000.$$

The elements of the confusion matrix can also be represented by the estimated proportion of each crop area classification. In such a case,

$$\hat{a}_{11} = \sum_{i \in S} \frac{\gamma_i(1,1)}{1000} = 0.201.$$

The confusion matrix is then used to provide input for the direct and inverse calibration estimators.

The proposed bootstrap algorithm was used to first generate a pseudo-population  $U^*$  of size 1 million, built upon the set of test points classified in Table 3 and then, at each replicate  $b$ , to select a new set of 1000 points using strategy 1 and simple random sampling to generate  $a_b$  and to compose the direct and inverse estimates.

RS strategy of classification of selected test points was then investigated using independent simple random sampling of two hundred pixels within each class of crop classified by remote sensing. This corresponds to use the columns of Table 2 as strata to select the test points. The selected points are then classified based on the ground, generating Table 4.

Table 4 information can only be used to generate a sample confusion matrix  $a$  with crop area estimates based on probabilities conditioned to the remote sensing classification (columns). For example, the confusion matrix proportion area for the classification  $g = 1$ , given a pixel is classified as  $c = 1$ , is provided by:

$$\hat{a}_{11} = \sum_{i \in S_1} \frac{\gamma_i(1,1)}{200} = \frac{127}{200} = 0.635.$$

Therefore, it only makes sense to use such information to calculate estimates by the direct estimator.

**Table 4.** Test points classification using RS strategy, with 200 pixels selected by simple random sampling from each set of remote sensing class independently.

Crop Area Classification		Remote Sensing Classification				
		Wheat	Rapeseed	Corn	Sugar Beet	Others
Ground truth classes	Wheat	127	44	11	4	10
	Rapeseed	4	49	0	0	9
	Corn	1	27	119	5	4
	Sugar beet	4	29	56	160	17
	Others	64	51	14	31	160
Total		200	200	200	200	200

The proposed bootstrap algorithm, with modified steps 1 and 2 to strategy 2, was used to first generate a pseudo-population  $U^*$  of size 1 million, built upon the set of test points classified in Table 4 and then, at each replicate  $b$ , to select independently a new set of 200 points for each class of remote sensing using simple random sampling. The data generate  $a_b$ , and estimates are calculated using the direct estimator. Although inappropriate, the inverse estimator was also calculated for each bootstrap replicate for the sake of illustration.

G strategy was studied using independent simple random sampling of two hundred pixels within each class of crop classified on the ground. This corresponds to use the rows of Table 2 as strata to select the test points. The selected points are then classified based on remote sensing, generating Table 5.

**Table 5.** Test points classification using G strategy, with 200 pixels selected by simple random sampling from each set of ground class independently.

Crop Area Classification		Remote Sensing Classification					Total
		Wheat	Rapeseed	Corn	Sugar Beet	Others	
Ground truth classes	Wheat	163	14	3	6	14	200
	Rapeseed	37	124	0	0	39	200
	Corn	3	25	150	15	7	200
	Sugar beet	7	15	37	117	24	200
	Others	57	15	3	12	113	200

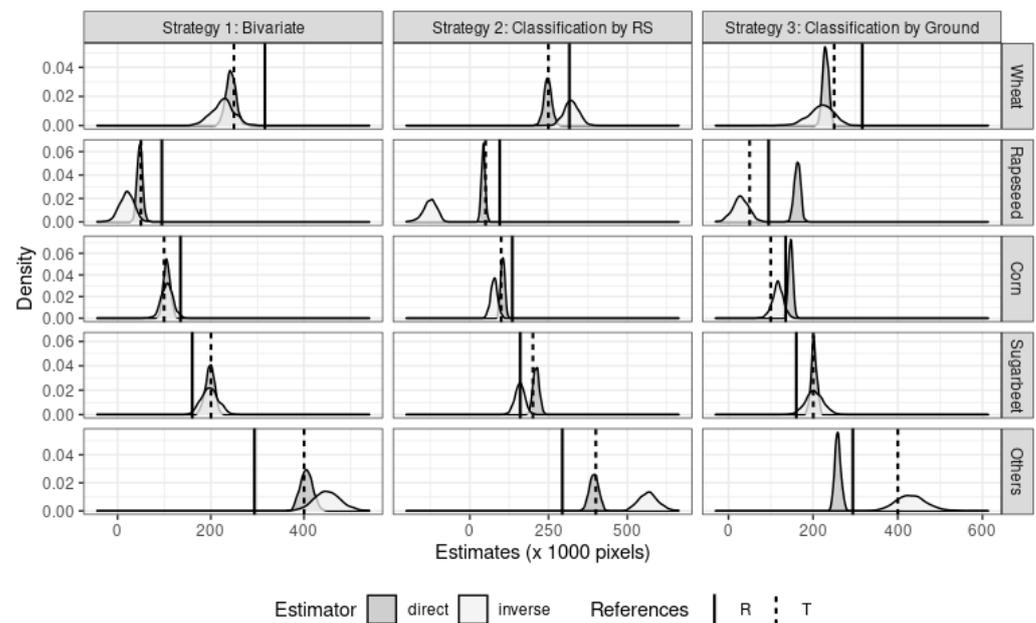
Table 5 information can only be used to generate a sample confusion matrix  $a$  with crop area estimates based on probabilities conditioned to the ground classification (rows). For example, the confusion matrix proportion area for the classification  $c = 1$ , given a pixel is on the ground class  $g = 1$ , is provided by:

$$\hat{a}_{11} = \sum_{i \in S_1} \frac{\gamma_i(1,1)}{200} = \frac{163}{200} = 0.815.$$

Therefore, it only makes sense to use such information to calculate estimates by the inverse estimator.

The proposed bootstrap algorithm, with steps 1 and 2 modified to strategy 3, was used to first generate a pseudo-population  $U^*$  of size 1 million, built upon the set of test points classified in Table 5 and then, at each replicate  $b$ , to select independently a new set of 200 points for each ground class using simple random sampling. The data generate  $a_b$ , and estimates are calculated using the inverse estimator. Although inappropriate, the direct estimator was also calculated for each bootstrap replicate for the sake of illustration.

All the strategies were evaluated using 1000 bootstrap replicates. Figure 1 summarizes the results for each one, describing the bootstrap distribution of the direct and inverse estimators for each crop.



**Figure 1.** Summary of bootstrap estimates per crop using a thousand replicates.

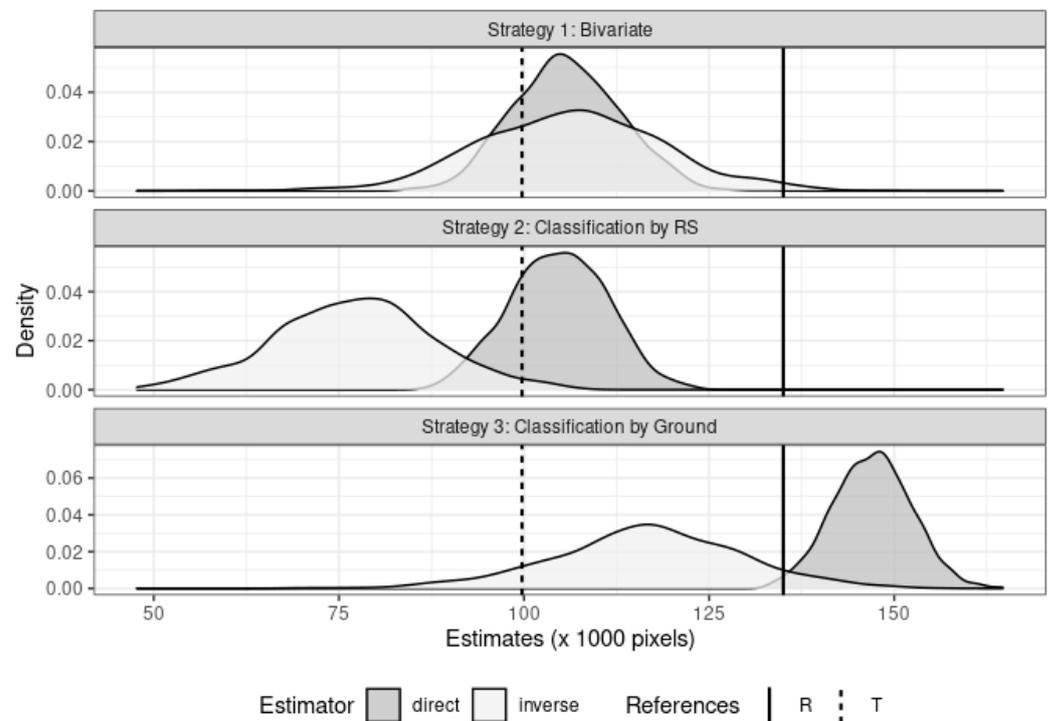
In Figure 1, the dotted line represents the parameter of the population (ground truth,  $T$ ), and the solid line represents the estimate based purely on pixel counting ( $R$ ). Strategy 1 uses a multinomial selection to compose the confusion matrix and hence allows for the use of either direct or inverse estimators. Focusing on the results of strategy 1, it is possible to see the bias of the estimates provided by  $R$ , based on pixel counting. Clearly, using direct or inverse estimates results in improvement over the ones from  $R$ . Considering crop areas of sugar beet, both estimators (direct and inverse) provide practically unbiased estimates, with the direct estimator showing smaller variance. For corn, a small bias is noted for both direct and inverse estimators, with direct estimator providing smaller variance. For rapeseed, wheat, and others, one can see that the direct estimator performs better than the inverse, showing smaller variance and bias.

The analysis of the results under the RS strategy must consider the fact that the confusion matrix was built based on a multinomial product selection of points per image classification. Hence, when analyzing the results in Figure 1, only the direct estimator makes sense to be calculated. For all crops, the estimates provided by the direct estimator represent an improvement over estimates provided by  $R$ , based on pixel counting in the sense that it redresses considerably the bias. For wheat, rapeseed, and others, the direct estimator shows smaller bias than for corn and sugar beet. The use of the inverse estimator is not appropriate in this scenario, and one can see that its performance does not represent improvement over the  $R$  estimates.

The last analysis refers to the G strategy to classify selected testing points. Under this strategy, confusion matrix is built based on a multinomial product selection of points per ground classification. Therefore, only the inverse estimator makes sense to be calculated. Focusing on the results for the G strategy in Figure 1, one can see that for all the crops, using the inverse estimator represents improvements over the use of  $R$  in the sense of redressing the bias of using only pixel counting estimates. The use of the direct estimator is not appropriate in this case, and one can see that for rapeseed, corn, and others, the direct estimator is not acting to diminishing bias. Although for sugar beet and wheat, the direct estimator has shown a similar effect of bias reduction as the inverse estimator and with smaller variance, its composition has no support under this scenario and should be avoided.

Figure 2 allows for an analysis emphasizing the effect of the different strategies for selecting test points over the crop area estimates of corn. In this case, when the Bivariate strategy is adopted, direct and inverse estimators offer improved estimates in the sense

that they show smaller bias than the estimate based on pixel counting alone. The use of either estimator is theoretically justified, and the direct estimator shows smaller variance than the inverse one.



**Figure 2.** Bootstrap estimation for corn per strategy of selection of test points.

Continuing to analyze Figure 2, if the RS strategy is adopted, then only the direct estimator is theoretically justified. Indeed, the bootstrap distributions show that only the direct estimator was able to redress considerably the bias for corn area estimates compared to the estimate based solely on pixel counting ( $R$ ). On the other hand, if the G strategy is adopted, only the use of the inverse estimator can be justified. Indeed, one can see that for the G strategy, the performance of the estimators shows that only the inverse estimator was able to diminish bias for corn area estimates. In such case, the inappropriate use of the direct estimator leads to an increase in bias. It should be noted, however, that only the G strategy has an appeal to be used in practice, as argued in Section 3.

The type of analysis presented for corn in Figure 2 can also be done for the remaining considered crops. Although they are omitted due to space constraints, they are available through a GitHub directory. Tables with the summary of statistical properties for each estimator under each strategy are presented in the Appendix A. When analyzing the summaries of Tables A1–A3, one should keep in mind that offices of official statistics look for CVs around 5–10% at the regional level (10,000 km<sup>2</sup>). Considering that 1 million pixels of Sentinel 1 (10 × 10 m<sup>2</sup>) correspond roughly to only 100 km<sup>2</sup> and the fact that, very often, the obtained CVs are below 10%, it is possible to conclude that the obtained precision is well in line with the needs.

## 6. Concluding Remarks

Remote sensing has several potential uses in agricultural statistics [23–25], including estimating crop areas based on pixel counting over satellite images. Such estimates, based purely on pixel counting, are known to be biased [5,20], and their variance depends upon the error rates of the classification rule in use even though several studies still use them with no further assessment of their statistical properties. Keeping the accuracy of the estimates under some control [26] may depend on factors such as landscape and image resolution. Direct and inverse calibration estimators reviewed in this paper are built upon

sample confusion matrix information and intend to redress the bias of estimates based on pixel counting. Several studies comparing both estimators describe the direct estimator as presenting the best performance [26,27]. There are instances where the inverse estimator shows smaller variance [20]. In this paper, we emphasized the fact that the feasibility of each estimator in practice depends upon the chosen strategy to collect and classify testing points on the field. Three strategies were discussed in this paper: the Bivariate, the RS and the G strategy. The G strategy is presented as the one with more practical appeal. If the G strategy is carried out in practice, the only estimator that is theoretically supported is the inverse estimator. Even considering the appropriate sampling strategy, assessing their variance may not be a simple task, as it also depends on the complexity of the sample design used to build the confusion matrix. In order to cope with this problem, a bootstrap algorithm was introduced for each sampling strategy, based on information provided by confusion matrices, that considers unequal inclusion probabilities. A small simulation study was presented where the statistical properties of the considered estimators were assessed based on the proposed bootstrap algorithm. The results illustrate the effectiveness of the bootstrap resampling method to assess direct and inverse calibration estimators under appropriate strategies. The performances shown are in line with the theoretical expectations and with the results from other studies [20,21,26–28]. The codes and main results of the simulation can be found in [29].

**Author Contributions:** Conceptualization, J.D. and C.F.; Methodology, C.F. and J.D.; software, A.L.; validation, R.O.; writing—original draft preparation, C.F.; writing—review and editing, J.D., C.F. and R.O.; supervision, C.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors would like to thank the editors and three anonymous reviewers for their comments. They helped to improve the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

Tables A1–A3 show the summaries of the performances of each estimator of crop area, assessed by the proposed bootstrap algorithm, with 1000 replicates. Each Table corresponds to a strategy of sampling testing points in the ground. Estimates of area need to be multiplied by 1000 to represent the estimated total area in hectares. Estimators are assessed with respect to standard deviation (Std.Dev.), CV expressed in percentage, and absolute bias (Bias).

**Table A1.** Statistical performances per crop for the Bivariate strategy.

Crop	Estimator	Estimate	Std.Dev.	CV (%)	Bias
Wheat	Direct	243.1	10.45	4.3	−6.5
	Inverse	225.6	23.33	10.3	−24.0
Rapeseed	Direct	47.7	5.81	12.2	−2.4
	Inverse	19.9	15.66	78.7	−30.2
Corn	Direct	105.8	7.23	6.8	6.0
	Inverse	106.4	12.46	11.7	6.6
Sugar beet	Direct	198.2	9.54	4.8	−2.1
	Inverse	198.0	17.76	9.0	−2.4
Others	Direct	405.3	13.01	3.2	5.1
	Inverse	450.2	28.10	6.2	50.0

**Table A2.** Statistical performances per crop for the RS strategy.

Crop	Estimator	Estimate	Std.Dev.	CV (%)	Bias
Wheat	Direct	247.5	12.14	4.9	−2.2
	Inverse	322.8	22.76	7.1	73.2
Rapeseed	Direct	43.0	6.02	14.0	−7.1
	Inverse	−126.4	20.81	−16.5	−176.5
Corn	Direct	104.6	6.52	6.2	4.9
	Inverse	77.0	10.64	13.8	−22.7
Sugar beet	Direct	210.7	9.76	4.6	10.3
	Inverse	158.9	15.38	9.7	−41.5
Others	Direct	394.3	14.26	3.6	−6.0
	Inverse	567.8	30.22	5.3	167.5

**Table A3.** Statistical performances per crop for the G strategy.

Crop	Estimator	Estimate	Std.Dev.	CV (%)	Bias
Wheat	Direct	229.6	6.99	3.0	−20.0
	Inverse	219.1	28.82	13.2	−30.6
Rapeseed	Direct	162.8	7.34	4.5	112.7
	Inverse	29.0	18.46	63.6	−21.1
Corn	Direct	147.0	5.22	3.6	47.3
	Inverse	116.6	12.77	11.0	16.9
Sugar beet	Direct	202.1	6.78	3.4	1.8
	Inverse	203.0	20.67	10.2	2.6
Others	Direct	258.5	7.11	2.7	−141.8
	Inverse	432.3	36.06	8.3	32.1

## References

- Fonteneau, F.; Delincé, J. Surveying Farms in the 21st Century. In *Handbook on the Agricultural Integrated Survey (AGRIS); Global Strategy to Improve Agricultural and Rural Statistics*; FAO: Rome, Italy, 2017; pp. 1–6.
- Taylor, J.C.; Sannier, C.; Delincé, J.; Gallego, J. Regional Crop Inventories in Europe assisted by Remote Sensing: 1988–1983. In *Synthesis Report of the MARS Project Action 1*; Technical JRC Report EUR 17319 EN; European Commission: Luxembourg, 1997; 67p.
- Delincé, J. Cost-Effectiveness of Remote Sensing for Agricultural Statistics in Developing and Emerging Economies. In *GSARS Technical Report GO09-2015*; FAO: Rome, Italy, 2015.
- D’Andrimont, R.; Verhegghen, A.; Lemoine, G.; Kempeneers, P.; Meroni, M.; Van der Velde, M. From parcel to continental scale—A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations. *Remote Sens. Environ.* **2021**, *266*, 112708. [[CrossRef](#)]
- Gallego, J. Estimating and correcting the bias of pixel counting. In *Handbook on Remote Sensing for Agricultural Statistics*; Delincé, J., Ed.; GSARS, FAO: Rome, Italy, 2017; pp. 249–261.
- Card, D.H. Using known map categorical marginal frequencies to improve estimates of thematic map accuracy. *Photogramm. Eng. Remote Sens.* **1982**, *48*, 431–439.
- Hay, A.M. The derivation of global estimates from confusion matrices. *Int. J. Remote Sens.* **1988**, *9*, 1395–1398. [[CrossRef](#)]
- Jupp, D.L.B. The stability of global estimates from confusion matrices. *Int. J. Remote Sens.* **1989**, *10*, 1563–1569. [[CrossRef](#)]
- Yuan, D. Natural constraints for inverse area estimate corrections. *Photogramm. Eng. Remote Sens.* **1996**, *62*, 413–417.
- Efron, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **1979**, *7*, 569–593. [[CrossRef](#)]
- Ospina, R.; Cribari-Neto, F.; Vasconcellos, K.L. Improved point and interval estimation for a beta regression model. *Comput. Stat. Data Anal.* **2006**, *51*, 960–981. [[CrossRef](#)]
- Brick, J.M. *Bootstrap Methods for Finite Population Sampling*; American University: Washington, DC, USA, 1984.
- Hall, P. *The Bootstrap and Edgeworth Expansion*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.

14. Liu, B.; Diallo, M. *Parametric Bootstrap Confidence Intervals for Survey-Weighted Small Area Proportions*; JSM Central: New York, NY, USA, 2013; pp. 109–121.
15. Beaumont, F. The analysis of survey data using bootstrap. In *Contributions to Sampling Statistics*; Mecatti, F., Conti, P.L., Ranalli, M.G., Eds.; Springer: New York, NY, USA, 2014; pp. 53–63.
16. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; CRC Press: Boca Raton, FL, USA, 1994.
17. Shao, J.; Tu, D. *The Jackknife and Bootstrap*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
18. Hesterberg, T. Bootstrap. *WIREs Comput. Stat.* **2011**, *3*, 497–526. [[CrossRef](#)]
19. Davison, A.C.; Hinkley, D.V.; Young, G.A. Recent developments in bootstrap methodology. *Stat. Sci.* **2003**, *18*, 141–157. [[CrossRef](#)]
20. Yuan, D. A simulation comparison of three marginal area estimators for image classification. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 385–392.
21. Champagne, C.; McNairn, H.; Daneshfar, B.; Shang, J. A bootstrap method for assessing classification accuracy and confidence for agricultural land use mapping in Canada. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *29*, 44–52. [[CrossRef](#)]
22. Conti, P.L.; Marella, D.; Mecatti, F.; Andreis, F. A unified principled framework for resampling based on pseudo-populations: Asymptotic theory. *Bernoulli* **2020**, *26*, 1044–1069. [[CrossRef](#)]
23. Carfagna, E.; Gallego, F.J. Using remote sensing for agricultural statistics. *Int. Stat. Rev.* **2005**, *73*, 384–404. [[CrossRef](#)]
24. GEOSS. *Best Practices for Crop Area Estimation with Remote Sensing*; Gallego, J., Craig, M., Michaelsen, J., Bossyns, B., Fritz, S., Eds.; Joint Research Center: Ispra, Italy, 2008.
25. Ndao, B.; Leroux, L.; Gaetano, R.; Aziz Diouf, A.; Soti, V.; Bégué, A.; Mbow, C.; Sambou, B. Landscape heterogeneity analysis using geospatial techniques and a priori knowledge in Sahelian agroforestry systems of Senegal. *Ecol. Indic.* **2021**, *125*, 107481. [[CrossRef](#)]
26. Waldner, F.; Defourny, P. Where can pixel counting meet user-defined accuracy requirements? *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *60*, 1–10. [[CrossRef](#)]
27. Czaplewski, R.L.; Catts, G.P. Calibrating area estimates for classification error using confusion matrices. In Proceedings of the 56th Annual Meeting of The American Society of Photogrammetry and Remote Sensing, Denver, CO, USA, 18–23 March 1990; Volume 4, pp. 431–440.
28. Walsh, T.A.; Burk, T.E. Calibration of satellite classifications of land area. *Remote Sens. Environ.* **1993**, *46*, 281–290. [[CrossRef](#)]
29. CastLaboratory/Croparea. Available online: <https://github.com/castlaboratory/croparea> (accessed on 23 March 2022).