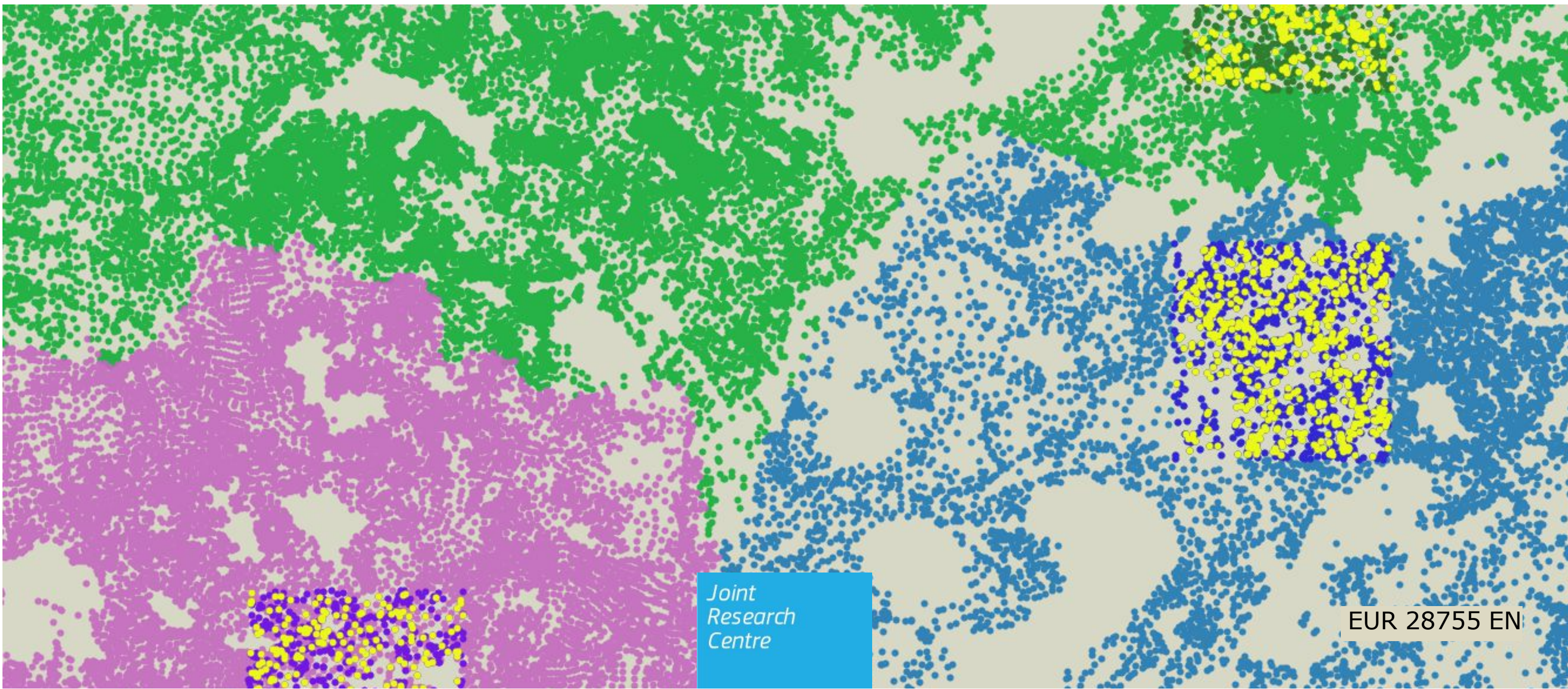
JRC TECHNICAL REPORTS

Analysis of the sampling representativeness for the Land Parcel Identification System Quality Assurance

Dominique Fasbender, Wim Devos,
Slavko Lemajic

2017



EUR 28755 EN

Contact information
Name: Wim Devos
Address: JRC, Via Fermi, 2749, I-21027 Ispra (VA), Italy
Email: wim.devos@ec.europa.eu
Tel.: +39 0332 785895

JRC Science Hub
https://ec.europa.eu/jrc

JRC107675

EUR 28755 EN

PDF ISBN 978-92-79-73181-5 ISSN 1831-9424 doi:10.2760/458367

Luxembourg: Publications Office of the European Union, 2017

How to cite this report: Fasbender, D., Devos, W. and Lemajic, S., *Analysis of the sampling representativeness for the Land Parcel Identification System Quality Assurance*, EUR 28755 EN, Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-73181-5, doi:10.2760/458367, JRC107675.

Contents

Executive summary

After its 2016 performance audit on the Land Parcel Identification System (LPIS), the European Court of Auditors recommended that the Commission should carry out a cost-benefit analysis in order to determine whether the representativeness of QA samples could be improved so that a better coverage of the population of parcels in LPIS be achieved. This report holds the results of that analysis.

An appropriate indicator for representativeness was developed and benchmarked. That indicator, "percentage of the population in the 95% central probability interval (PCPI)" was then applied to the actual samples of the past to measure their performance. Additionally, the simulation of several sampling scenarios allowed to assess whether sub-optimal conditions could be remediated by the appropriate mitigation measures.

The results reconfirmed that the initial approach of spatial stratification does not interfere with the representativeness in se, but it also found that the current implementation could leave a few territories with some residual effect from the implied spatial clustering. This potential weakness was addressed by improving the resolution of the stratification control layer, ensuring a minimum set of clusters as well as introducing stratified random sampling. Simulations demonstrated these combined remedies were effective for nearly all Member States and regions. Cyprus and Luxemburg missed the target, but only just.

The revised methodology led to a 15 percent increase of the number of control zones over Europe, but did not require any procedural modification or other additional inspection workload for the Member States. This revised sampling methodology thus considers the Court's cost-benefit concerns and has been fully implemented for the 2017 LPIS QA campaign.

1 Introduction

1.1 Context

The LPIS Quality Assurance (QA) is a yearly exercise that is jointly organised since 2010 by the European Commission (EC, in particular by DG JRC) and the member states (MS) administrations. It consists in an evaluation of the quality of the LPIS systems based on a sample of the reference parcels (RP) and a recent photo-interpretation of the parcel boundaries and interior. A set of quality elements (QE) were defined by the EC and compared to limiting quality thresholds. The LPIS QA sample requirements are thus driven by the ISO standard 2859/2.

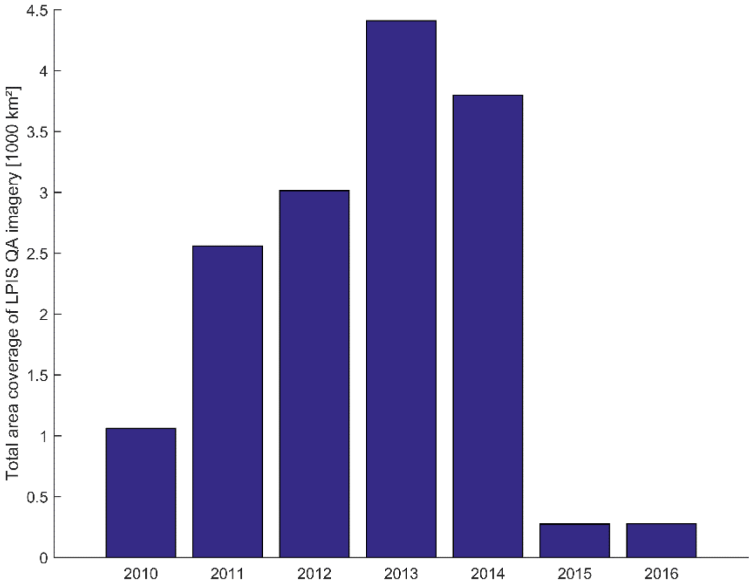
The LPIS QA exercise is performed at LPIS system-level. It means that the MS inspect what they consider as homogeneous population of their RPs. This follows the hypothesis that the same rules were applied for the creation of each RP in the system. Should there be different local rules, then the MS is invited to consider and send one lot for each sub-population. In the past, PT and RO sent their population in 2 lots reflecting update projects. BE, DE and UK adopted a systematic and enduring split of their respective national territories, following their administrative regional boundaries: BE is split in 2 regions, DE in 13 Bundesländer and UK in 4 countries. MS such as ES, FR, IT or PL are sending their RP population in a unique lot, even if several paying agencies (PA) are operating regionally (e.g. ES has 17 PA and IT has 9). However, the homogeneity hypothesis is currently not challenged and the generation of the samples for the LPIS QA is assuming that this hypothesis is valid.

Until 2014 included, the samples were taken from a sub-set of the control with remote sensing (CwRS) zones from the simultaneous OTS control period to minimize the inspection burden for the Member States. Historically, it was asked to the MS administration to identify at least 3 unbiased zones that could be subject to LPIS QA and communicate these to DG JRC who would ensure appropriate image acquisition. The latter was also in charge of the random selection of the RP within the designated zones. The selection of the zones was obviously known beforehand by the MS administration. Even though the LPIS QA is an exercise aiming to help the MS administrations to improve their own systems, this situation could create biased estimations of the different QEs. The first potential cause of bias could be that, since the CwRS zones are originally selected for random or risk On-The-Spot (OTS) checks, the RPs within these zones are already getting much attention and update priorities from the authorities. In this configuration, the CwRS zones (and the RPs that are covered by these zones) might be of higher quality than the rest of the RP population. The second and more worrying potential cause of bias is that the MS administrations could be tempted to pre-clean the known LPIS QA zones anticipatively in order to perform better during the exercise. While the first cause could be explained by innocent prioritization of the work, the second reason is intentional and both reasons lead to the same consequence of a biased assessment of the LPIS system quality. To mitigate that consequence, the EC had from the start requested uploading the full population to be able to perform both ex-ante and ex-post analyses and had been monitoring the assumptions, costs and benefits.

After an evaluation in 2014, DG JRC set up a new image acquisition procedure in which the LPIS QA zones are not known by the MS administrations before the beginning of the LPIS QA measurements. In addition, the new procedure had the objectives to (i) rely on images with better quality for the inspections and (ii) to reduce the logistic burdens of control zone selection and RP sample generation and as a result (iii) facilitate complementary field observations. Indeed, by concentrating the QA inspections (see

Figure **1**), the potential Rapid Field Visits are also easier to organize. In practice, DG JRC orders the images directly from the image provider. The images are stored and kept secret until the MS administration upload the full population subject to the LPIS QA exercise. After several quality checks on the population sent, DG JRC crops the submitted population to the zone delimited by the images and proceeds with a random selection of the so-called *sample pre-selection* that consist of 3 times the target sample size (i.e. 500/800/1250 depending on the population size).

Figure 1. Total area coverage of LPIS QA imagery between 2010 and 2016



The number of LPIS QA zones by LPIS system was determined based on the proper characteristics of the corresponding LPIS system. 1, 2 and 3 zones were automatically assigned to LPIS systems requiring samples of 500, 800 and 1250 RPs respectively and extra zones were allocated to larger LPIS systems (+1 when the difference "sin(lat_max)-sin(lat_min)" is between 0.05 and 0.07 and +2 when this difference is larger than 0.07, where lat_min and lat_max are respectively the minimum and maximum latitudes of the corresponding LPIS system). The latitude parameter is chosen because it is closely linked the biogeographical zoning. Each of the 44 LPIS systems is thus attributed from 1 to 5 zones. For the systems getting extra zones, a sub-division of the system was created as a spatial aggregation of the NUTS2 regions with the constraint that each aggregation would contain roughly the same number of RPs. Finally, the zones with a density of RPs smaller than 2 parcels per km² where regrouped into a single GIS mask for which 2 LPIS QA zones were requested in order to guarantee the chance of selection for reference parcels inside these low density areas; the zones with no RP were obviously completely excluded for the acquisition of the images. In this practice, the European territory was split in 77 different planning polygons in which 1, 2 or 3 images are expected to be acquired for the purpose of the LPIS QA.

In 2015 and 2016, the European Court of Auditors (ECA) performed an audit to "assess the reliability, effectiveness and impact of LPIS/GIS across the EU" (see ECA Special report 25/2016). After visiting both the MS administrations (6 different LPIS systems in 5 different MS) and the EC (DG AGRI and DG JRC), the auditors came up with 6 recommendations of which Recommendation 5 is mostly concerning actions from DG JRC:

"*The Commission should carry out a cost-benefit analysis in order to determine whether the representativeness of QA samples could be improved so that a better coverage of the population of parcels in LPIS be achieved.*" This recommendation was acknowledged and accepted by the EC as "*The monitoring of sample representativeness is part of a continuous process.*" In substance, it was argued by ECA that the number of zones was too low for some MS (e.g. Poland) in order to represent the different landscapes.

This document contains the analysis results of this request and the sub-optimal representativeness such request implies.

The analyses will therefore proceed on two objectives:

1. Evaluate the JRC's initial assumption that the stratification does not invalidate the randomness and so does not interfere with the representativeness.
2. Investigate if and where there are LPIS territories that do not receive enough zones to reliably counter any residual effect of spatial clustering. The initial working hypothesis of the JRC was that a minimum of three independent zones per territory provides enough variability to neutralize (or at least detect) any residual spatial clustering effect. This hypothesis was no longer enforced upon smaller systems when the dedicated LPIS images were introduced.

To deal with the two objectives above, an 8 step workplan was adopted:

1. Compile lessons learned from the 2015 and 2016 LPIS QA image acquisition campaigns (section 1.2)
2. Define indexes for assessment of the representativeness (section 2.4)
3. Evaluate the representativeness under the current situation using the different indexes (sections 3.1 and 3.2)
4. Make a first estimation of the potential increases of images for each MS (section 3.2)
5. Revise the sampling procedures and algorithms to ensure randomness (section 2.3 and IT developments not presented here)
6. Prepare the data for the sampling in 2017 (using up-to-date LPIS population data of 2016) (collection of the RP population from the 44 LPIS systems; not presented here)
7. Update subdivisions of larger MS and of the low density European zone (sections 3.3 to 3.5)
8. Re-evaluate the representativeness under the updated situation using the different indexes (section 3.6)

1.2 Lesson learned

The general methodology of the LPIS QA works and the positive effects of the LPIS QA exercises are acknowledged by the whole community since its first implementation in 2010. Additionally, the random image selection that was initiated in 2015 has been well accepted by the MS. Its advantages were recently highlighted by some MS, with a suggestion to extend in the future the random selection procedure to the generation of the random control zones for OTSC. Such a suggestion is a clear evidence of the efficiency of the procedure and the trust of the MS in DG JRC's handling.

The methodology proved to be particularly efficient in several cases. As an example, it occurred in a particular MS that the QEs assessing the correct implementation of LPIS update (i.e. QE4: ”Categorization of non-conformities” and QE6: ”Rate of non-conforming reference parcels due to undetected and unaccounted land cover change, as observed in ETS, accumulated over the years”) was particularly poor in 2015. The MS complained that the zone was not representative of the whole system and thus not reflecting the correct state of update. The MS even disputed the randomness of the zone selection from DG JRC and the image provider insinuating that the choice of the control zones have been driven by an external cause (i.e. train infrastructure network). DG JRC considers that, even if the selected zones were allegedly “ unluckily “ located and have an error rate above the system as a whole, the methodology (and in particular the use of control zone clusters) and results permitted to detect this weakness in the update process as significant; a significance that would not have been achieved with a full random selection of RPs over the whole territory because the clustered effect of these lacked updates would then be completely diluted in the corresponding QEs. This effect is also an argument for introducing a multistage sampling (i.e. a random selection of the control zones followed by a random selection of the RPs within these zones) because it has the tendency to reflect the spatial trends throughout the LPIS QA exercise.

On the other hand, the LPIS QA image selection is quite recent and, as for every kind of procedures, is in constant monitoring for improvements. In particular, DG JRC observed other potential for further improvement and addressed this for the 2017 LPIS QA exercise:

- the spatial resolution of the regular grid at the basis of the zonings for image acquisition was relatively coarse (i.e. 10 km x 10km);
- the MS administrative boundaries were too coarse (derived from a scale 1:1,000,000);
- in particular, the spatial quality of the coastline representation was poor and
- as a consequence, some RPs (0.045% of the EU RPs) were observed out of the zoning polygons used for image acquisition (by consequence, they could not be sampled);
- the LPIS populations have evolved (because of updates, upgrades but also because of some evolution of the LPIS QA scope). The low-density mask needed continuous update;
- some MS experience difficult image acquisition conditions (e.g. cloud and snow cover) which could jeopardize the representativeness of the sample by “technical exclusion” of some parts of their territory.

1.3 Costs

1.3.1 Costs for the EC budget

At the EC’s side this could affect the image acquisition costs. Though the FWC (389911 (720383)), JRC purchases imagery at a fixed price for a fixed minimum zone size. If the increased sample representativeness would require more area on a larger spread, this would increase the number of zones, and a fixed purchase cost will be added for each additional zone. Reducing either the number of zones or the total area would free some budget, but from the ECA argumentation follows that ECA presumes that the number of zones could be increased.

For the 2016 LPIS QA image acquisition campaign, the numbers and costs were:

	Profile 50cm	Profile 40cm	Total
Number of zones	115	9	124
Average area [km^2]	224.82	169.00	220.77
Total area [km^2]	25,854	1,521	27,375
Unit cost [€/km²]	16	22	n/a
zone cost [€/km²]	3597	3718	3606
Total cost [€]	413,664	33,462	447,126

The selection of profile is based on local condition and sensor availability.

On top of the purchasing cost, there would also be a proportional increase of the management overhead estimated at 1-2 hours/zone, take 100 euro/zone which is considered negligible in the overall process.

1.3.2 Costs for the MS budgets

On the MS's side, an increase of zones would entail two extra costs:

- The cost for one extra orthorectification: this ranges from 500€-1500€ (depending on in-house capacities and external tenders) so can be estimated at an average 1000€ per extra zone.
- The logistic cost for the extra geographical spread in inspection activities (i.e. extra transfer to the additional zones and possibly interfacing with extra local administration levels in those additional zones. The exact overhead is certainly varied and more difficult to estimate, but a rough cost of 1000€ per zone is considered realistic.

1.3.3 Total

In short, the marginal cost per extra zone can be estimated at 5700€, two thirds of which would be borne by the budget of the EC services.

2 Data and methods for analysing representativeness

2.1 Data

For this analysis, data are available in the form of the RP population of the LPIS systems that are yearly sent by the MS administrations. Each of the RPs is represented by a point with coordinates and the corresponding reference area. From 2016 on, the respective areas for the 3 main land cover types (i.e. arable land, permanent grassland and permanent crop) is also known.

The total number of RPs containing CAP-related agricultural area in EU28 sums up to some 74 million. The number of RPs in the LPIS system is obviously driven by the size of the corresponding MS/region but also by the RP type, which may or may not fragmentate the landscape. Typically, the systems based on cadastral parcels (e.g. DE-BW, ES, IT, PL) have more sub-divisions of the agricultural land than those based on physical blocks (e.g. RO) or topological blocks (i.e. the 4 systems in UK). This also influences the RP area distribution.

Figure 2 shows the boxplots of the RP area evaluated for each of the 44 systems and regrouped according to the RP type. Systems based on cadastral parcels or agricultural parcels tend to have smaller RPs than the other RP type systems. However, regardless of the RP type, the RP area distributions are generally highly skewed to the right (i.e. a majority of RPs with small or medium area and few RPs with very large area). As an example, Figure 3 shows the histogram of the RP area in Sweden. One can clearly see the asymmetry of the distribution that is bounded by 0 on the left and with a long tail on the right.

Figure 2. Boxplots of reference parcel area per LPIS systems and regrouped by reference parcel type (AP=agricultural parcel, CP=cadastral parcel, FB=farmer block, PB=physical block and TB=topological block). The red lines represent the median of the distributions, the boxes show the 1st and 3rd quartile and the dashed lines show the minimum and maximum values within that are within 1,5 times the interquartile range (i.e. the size of the boxes)

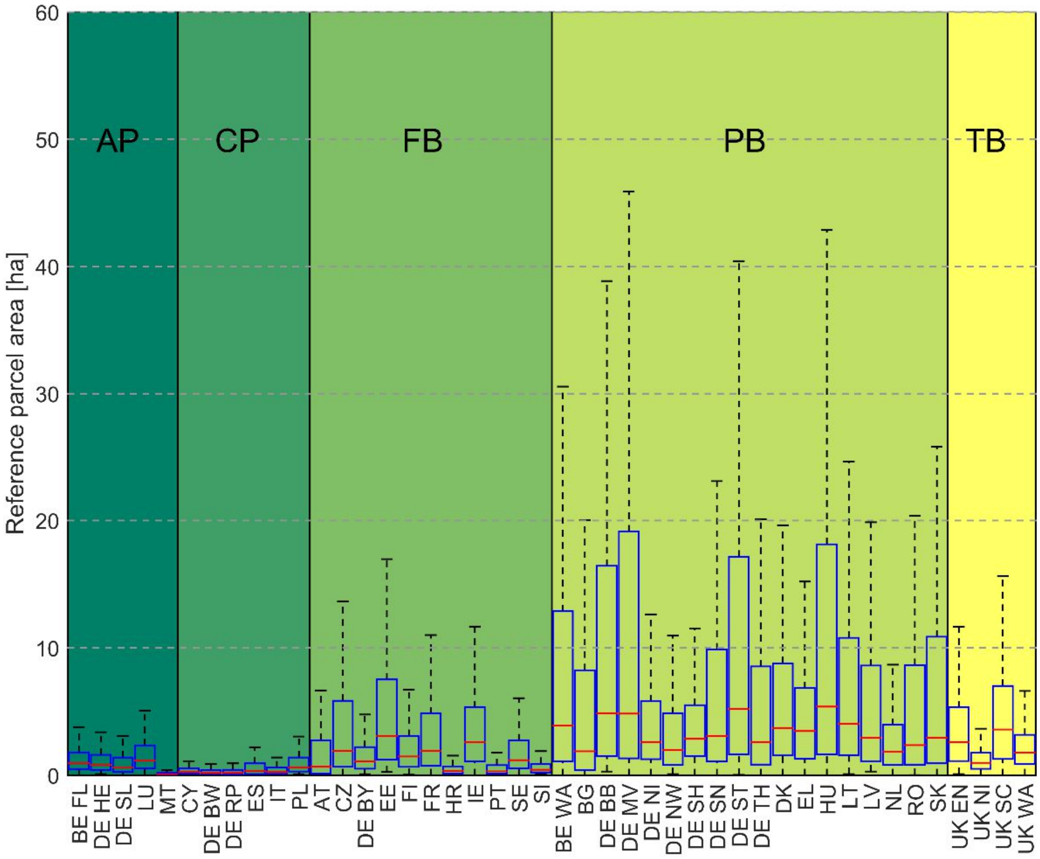
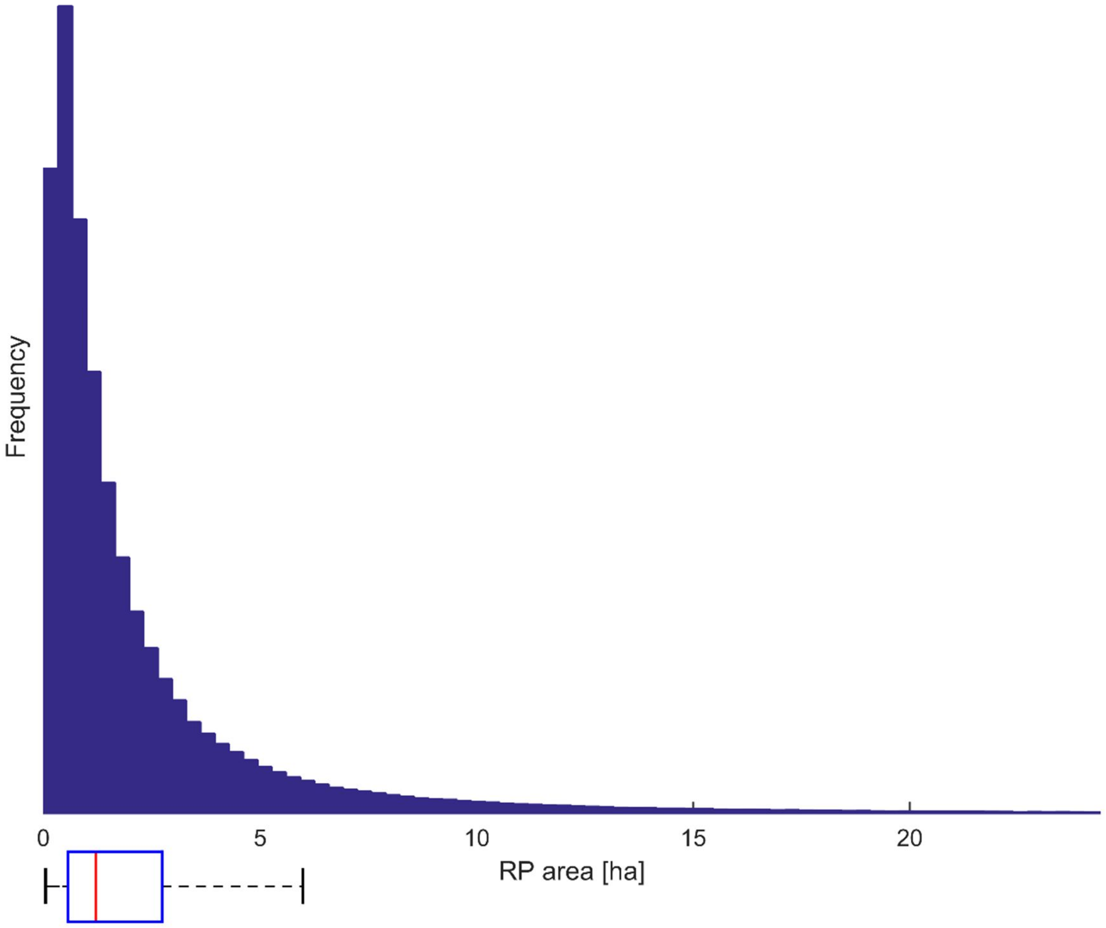


Figure 3. Histogram and boxplot of the reference parcel area in SE



There is no direct information regarding the different landscapes within the LPIS systems. The mere parameter is the area of the RPs. Any analysis on the representativeness of the samples is therefore bound to that sole parameter. Another important characteristic of the LPIS data is its spatial facet. The purpose of the LPIS system is to manage (inform and crosscheck) information on agricultural parcels in a Geographic Information System (GIS). For this, the agricultural areas of the MS's territory are divided into the RPs. Because of their spatial nature, imprecisions on one RP are propagated to the surrounding RPs (i.e. "what is wrongly allocated to one RP cannot be allocated to the next RP and will be missing from it").

2.2 Using simulated QA campaigns for evaluating the impact of the number of zones on the sample (spatial clustering)

What is a simulation?

In order to assess the representativeness of the sample, simulations were run. For each of the LPIS systems, the LPIS QA sampling procedure is applied using randomly simulated control zones of 15-by-15 km² (i.e. the same size as the actual LPIS QA images). These simulations were run both for the 2015-2016 conditions and for testing the amended sampling procedures for 2017.

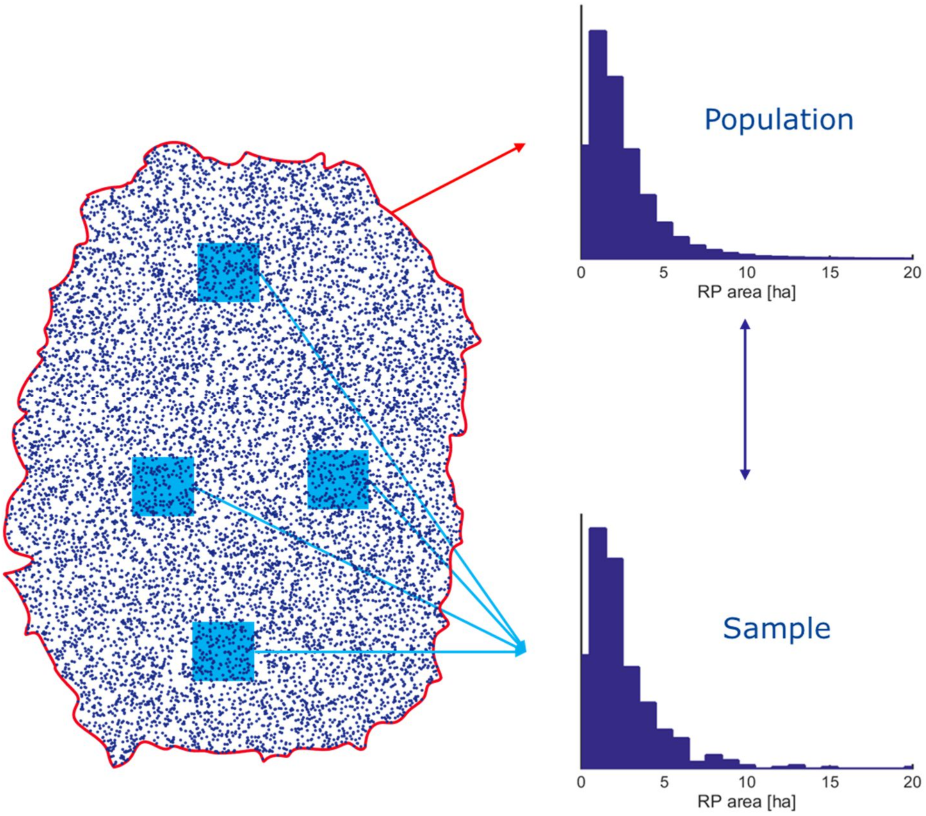
What is a scenario?

For testing the 2015-2016 conditions, different scenarios were considered from 1 to 5 QA control zones. It was thus possible to compare the effects of an increase of control zones on the different indexes. For larger LPIS systems where sub-delimitations were created, the same approach was applied for each of the sub-zonings. Consequently, the number of simulated QA zones ranged thus from 1 to 5 for the LPIS systems with a unique delimitation and from 1 to 25 (for the largest LPIS systems like France, Italy, Spain and Greece) for other systems. Based on the results of this first analysis, a new number of control zones is proposed for each LPIS system to be the new sampling conditions for 2017.

For each of the simulated sets of QA zones, the population data are cropped on the zones and the adequate number of RPs (i.e. 500/800/1250) are randomly selected from the cropped population data, simulating thus the multistage sampling that is applied during the generation of the sample selection. The procedure for the actual selection of the RPs within each control zones is presented later on (section 2.3) are computed.

Figure 4 illustrates how the samples are generated. This procedure can be repeated many times for each of the scenarios and each of the LPIS systems and after some preliminary tests on a selection of LPIS systems, the number of repetitions in this study was set to 100 because it already guarantees the convergence of the different indexes. For each simulation, the indexes of representativeness (see their definition in section 2.4) are computed.

Figure 4. Principles of the LPIS campaign simulations. Step 1: random selection of control zones (in bright blue). Step 2: Random selection of the reference parcels within the random control zones (dark dots within the control zones)



2.3 Stratifying the random selection of the reference parcels within the control zones

For the 2010-2016 conditions, the random selection was taken directly from the cropped population data without additional constraints. This step is thus a simple random sampling on the subpopulation covered by the control zones. This sampling method has the disadvantage that it completely misses the factor of control zones. Indeed, there is a risk that some control zones will not cover at least one inspected RP. This risk increases with the number of control zones because the number of selected RPs is fixed (thus the average number of RPs per control zones is getting smaller with the increase of control zones). This situation actually occurred in the simulations on the 2010-2014 period because the number of images was very high. For the 2015-2016 period, however, because the density of the RPs is not homogeneous within each MS, the probability of such missed control zones still exists. This issue is more important now that the LPIS QA imagery is purchased only for the LPIS QA purposes contrarily to the 2010-2014 period where the images were recycled from the CwRS imagery where a minimum density criterion applied at the time. In the current LPIS QA image acquisition, missing a control zone is not anymore acceptable because it would mean that an image was purchased for no use.

In order to address this remaining risk, the random selection for the new 2017 sampling procedure follows a stratified sampling procedure. In this alternative selection procedure, the different control zones are considered as strata. For each control zone, a simple random sample is generated. Consequently, each control zone is covered by the sample (i.e. no risk of unused images) and the number of selected RPs by control zone is proportional to the number of RPs covered by each zones. For instance, if 1250 RPs must be selected over 3 control zones respectively covering 70k, 20k and 10k RPs, then the sample will be split in 875, 250 and 125. This example illustrates that the stratified sampling procedure is actually equivalent to defining, for each of the 3 zones, a fixed sampling rate equal to 1,25% (i.e. the ratio of 1250 on 100k).

2.4 Representativeness and how to assess it

The representativeness of a sample extracted from a given population can be measured in different ways. The approach selected in this document is to rely on statistical tests for the assessment of the representativeness. Any statistical test is by definition characterized by its null and its alternative hypotheses. Often, there are different possible alternative hypothesis while the test's null hypothesis is always unique. In the statistical test, the null hypothesis is assumed true, unless data (observed or simulated) provide evidence to the contrary. The critical region (i.e. the range of values for the statistic that will lead to the rejection of the null hypothesis) depends on the choice of alternative hypothesis. Once the critical region is defined, the statistic of the test is evaluated using the available data and a decision is taken by checking whether the statistic is in or out of the critical region.

Five indexes, four of which are taken from statistical literature, were selected. For each of them, a statistical test is proposed and tested in the specific context of the LPIS QA and the available data.

It is however important emphasizing that, by nature, random samples are always different from one run to the other. Even if the same sampling algorithm is applied twice, the corresponding two samples are not expected to have the exact same properties and in particular to always achieve that same level of representativeness. In that sense, one must distinguish between two concepts of representativeness:

- the representativeness of a sampling procedure;
- the representativeness of a specific sample.

The representativeness of a sampling procedure is the averaged representativeness that is expected when applying the sampling procedure. It is thus a long-term quality of sampling procedure. It reflects the trend of a procedure to produce representative samples.

The representativeness of a specific sample is an individual evaluation of the representativeness between this particular sample and its original population. This evaluation is really specific to the sample and thus can potentially deviate from the expected representativeness of the corresponding sampling procedure.

Thus, while a sampling procedure tends to produce representative samples, it may happen that some individual samples fail to pass the representativeness test because they are not reflecting some of the population characteristics. Inversely, the fact that a specific sample is not a perfect reflection of the population for a given year does not prove itself that the sampling procedure is wrong. It is only by evaluating the representativeness of the samples on the long-term that the representativeness of the procedure can be confirmed or disputed. This long-term evaluation is performed here thanks to the simulated scenarios that are described in section 2.2. The objective of the JRC methodology and this analysis is to guarantee the representativeness of the sampling procedure, it is not to guarantee the representativeness of every single sample that procedure has generated or will generate.

2.4.1 The Z-test (or "one sample location test")

The Z-test is a statistical procedure for testing the hypothesis whether the sample mean $\bar{X}$ is likely to be equal to a given value. In our context, the test is applied using the true population mean μ and the true population standard deviation σ (see "Student", 1908). The statistic of the test is

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

The distribution of this statistic Z tends asymptotically towards the standard normal distribution.

For the alternative hypothesis, i.e. the mean from the sample and the true population mean are different, the critical region (i.e. the range of values for which the test is rejected) is defined as

$$|z| > 1.96$$

Where $|z|$ is the absolute value of the observed statistic and 1.96 is the quantile of the standard normal distribution at 97.5%. Figure 5 schematically shows how the Z-test is performed.

Under the conditions of the LPIS QA sample size and populations, the effect size (i.e. the ratio between the maximum allowed difference that is undetected by the test with 10% probability and the standard deviation of the population) is generally between 0.1 and 0.2 (Figure 6). This is considered as "small" to "very small" in the literature (see Sawilowsky, 2009). MT has a higher effect size but is not considered as an issue since its image coverage amounts to almost 50% of its territory.

Figure 5. Schematic explanation of the Z-test

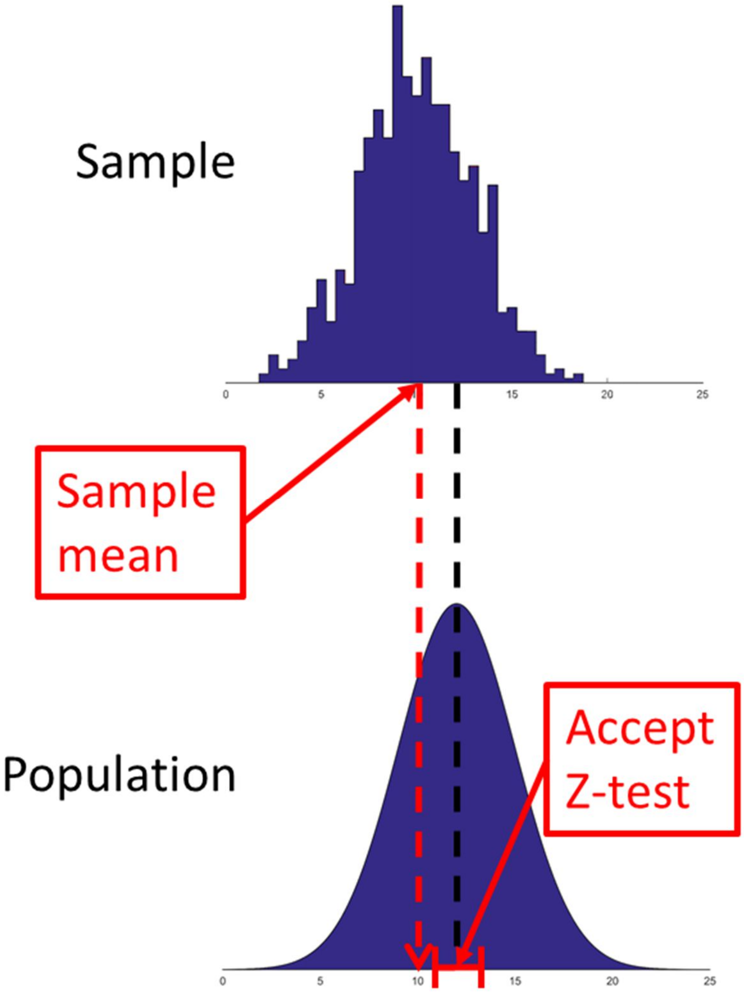
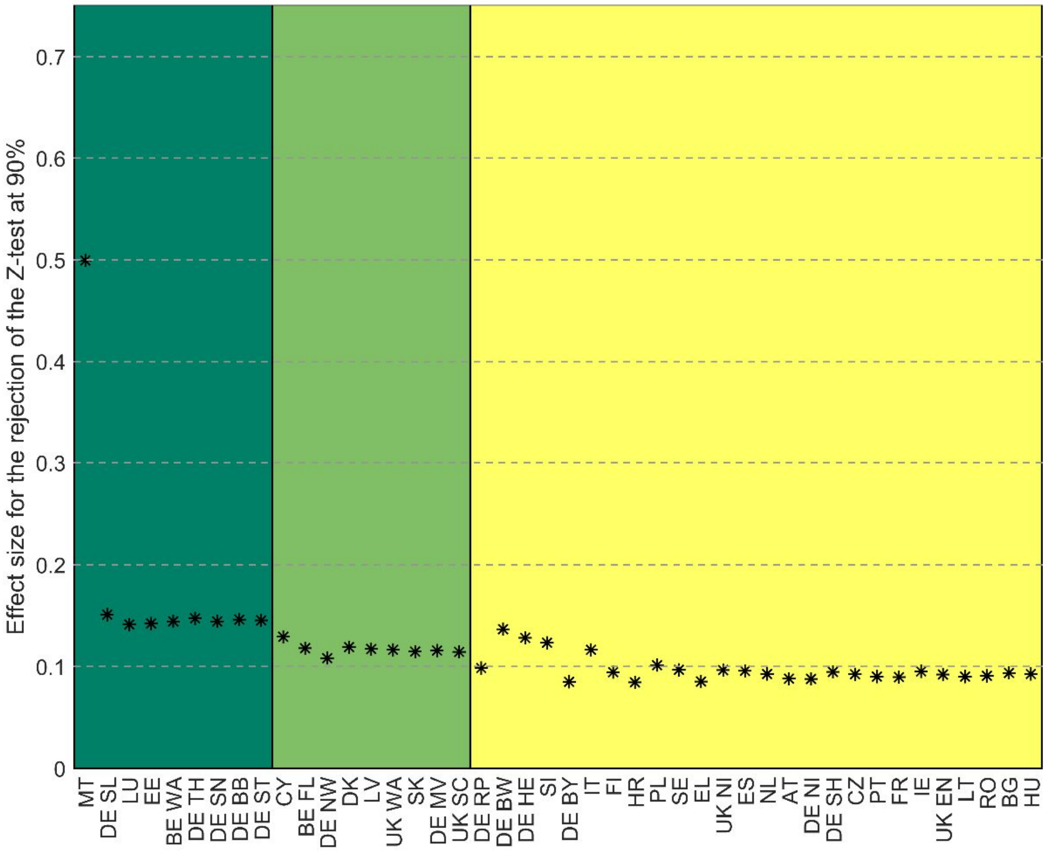


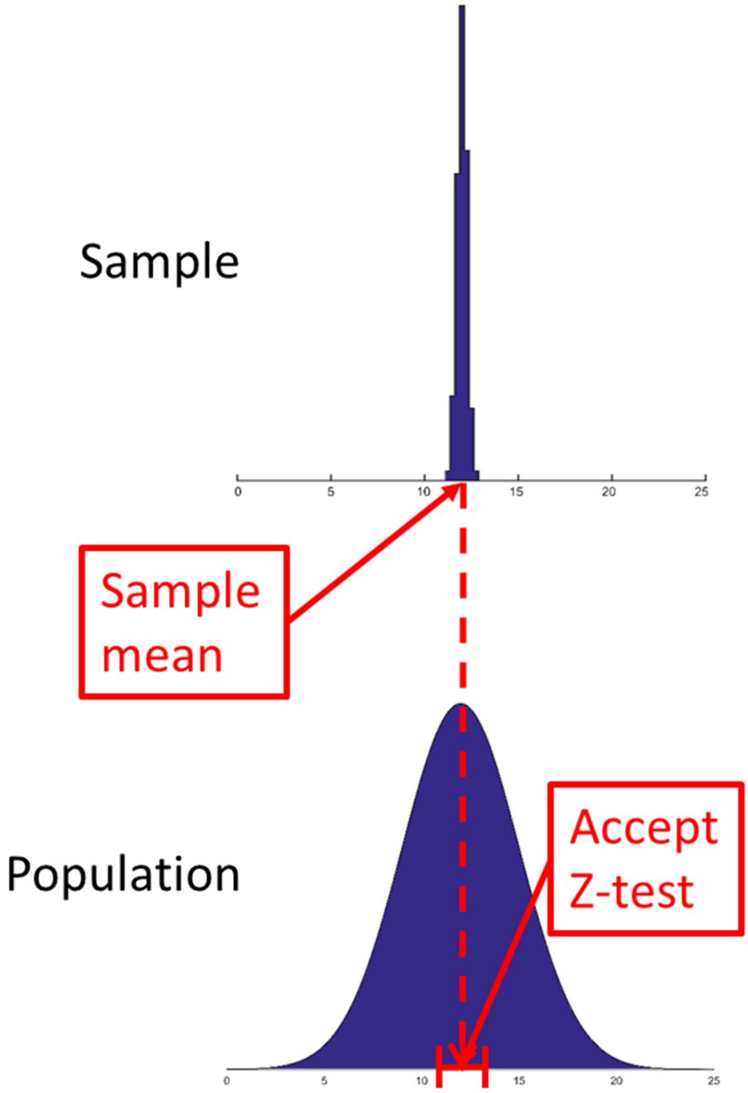
Figure 6. Effect size for the rejection of the Z-test at 90%. The values are compute as the ratio between the maximum allowed difference that is undetected by the test with 10% probability on the standard deviation of the population. The LPIS systems are sorted according to their LPIS QA sample size (dark green for 500, light green for 800 and yellow for 1250) and their standard deviation (decreasing order)



The Z-test is known to be optimal for normally distributed data and is a good approximation for other distributions. However, as a single value, it is not completely depicting the representativeness quality of the sample.

Figure 7 shows an example for which the test does not reject the agreement between the sample and the population means, while the distribution of the sample is clearly different from the original population. In that example, the sample covers the central part of the distribution only, missing both the small and the large elements of the population.

Figure 7. Counter-example of the inefficiency of the Z-test. The sample mean is in perfect agreement with the population mean but the two distributions are drastically different



2.4.2 The Chi-square test (or "one sample variance test")

The Chi-square test for a one sample variance is a statistical procedure for testing whether the sample variance S^2 is equal to a given value (see Kenney and Keeping, 1951). In our context, the test is applied using the true population variance σ^2. The statistic of the test is

$$X^2 = \frac{(n-1)S^2}{\sigma^2}$$

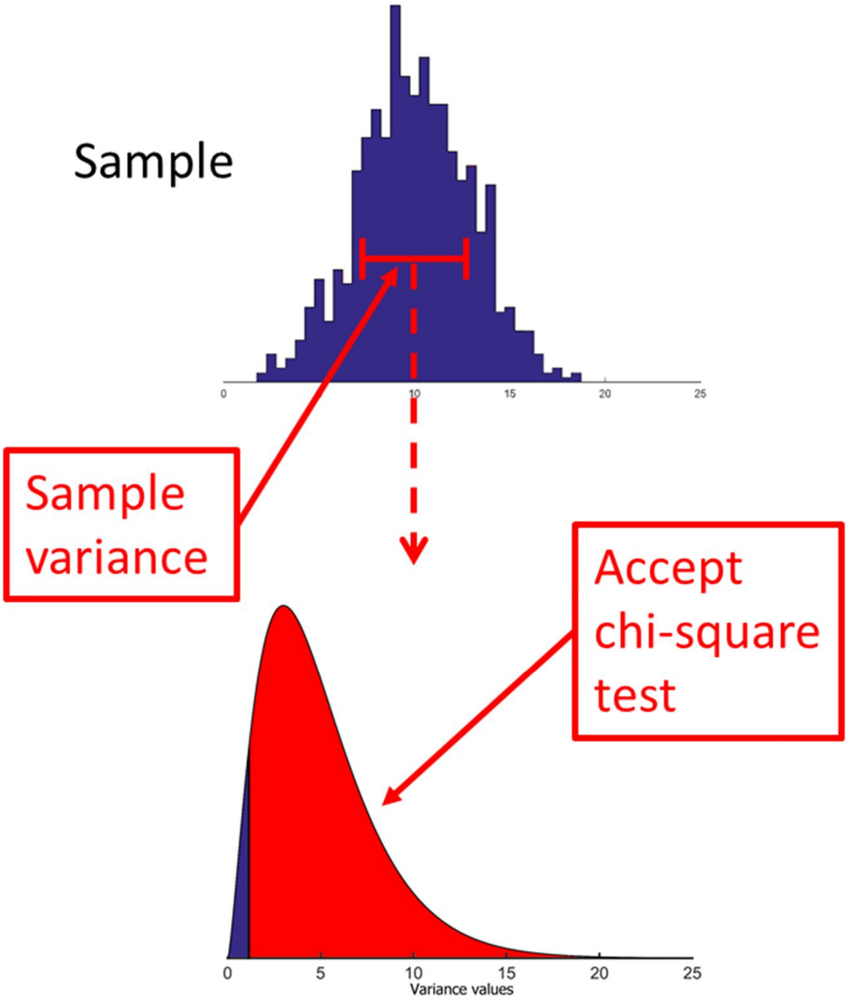
In our context, it is expected that a lack of representativeness of the sample would lead to a smaller sample variance (because the sample is taken from the known distribution). Moreover, if the sample variance is larger than the true population variance, it indicates that the sample covers a wider range of the population, which again can be seen as less dramatic. Consequently, the alternative hypothesis for this test is set to "the variance coming for the sample is smaller than the true population variance". Consequently, the critical region for our Chi-square test is defined as

$$\chi^2 < \chi^2_{0.05,n-1}$$

Where $\chi^2_{0.05,n-1}$ is the quantile of the chi-square distribution with $n-1$ degree of liberty at 5%. Depending on the LPIS QA sample size, the Chi-square test is rejected when the sample variance is 10% (for the 500 samples), 8% (for the 800 samples) or 6.5% (for the 1250 samples) smaller than the population variance. Figure 8 schematically shows how the chi-square test is performed.

Note that here too, passing the chi-square test is not sufficient for ensuring the sample representativeness. Indeed, a shifted sample may pass the test but would not be representative of the population. In addition, the chi-square test was designed for normally distributed data and is known to be sensitive to skewness (i.e. non-symmetric distributions). This is generally the case for the LPIS RP populations (see section 2.1 below).

Figure 8. Schematic explanation of the chi-square test



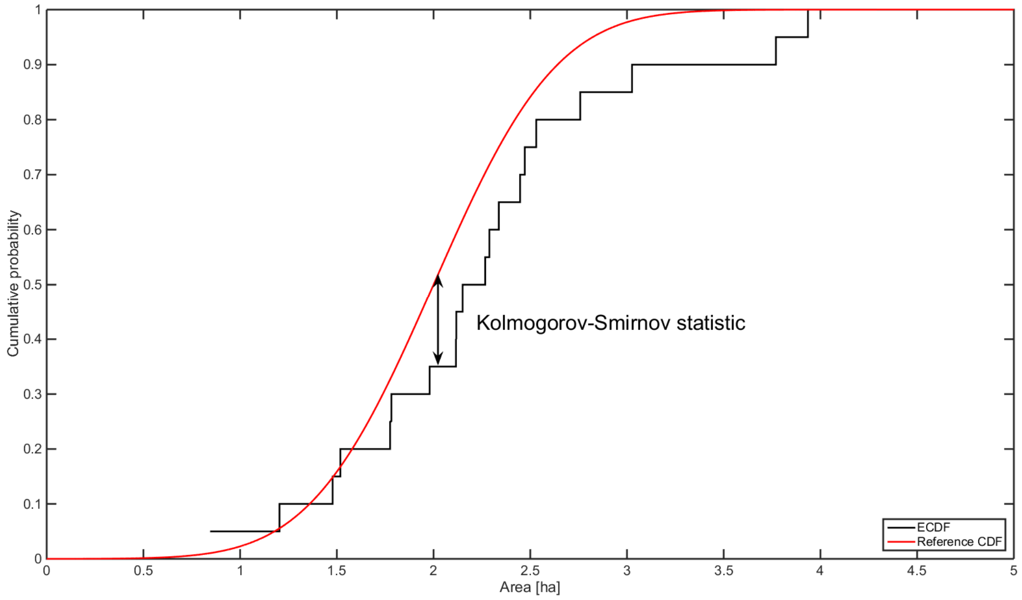
2.4.3 The Kolmogorov-Smirnov test

The Kolmogorov-Smirnov (KS) test is a statistical procedure specifically for testing whether the sample distribution is equal to a given distribution (see Kolmogorov, 1933). By definition, it is the most comprehensive and demanding test for a distribution. Indeed, the mean and the variance are merely distinct properties of that distribution. Thus, if the sample and the population distributions are in agreement, then in principle, all other properties should follow as well, which is what representativeness is all about.

The statistic of the test is defined as the largest absolute difference over the domain of the distributions between the empirical distribution function (i.e. the estimated cumulative distribution function; ECDF) and the tested cumulative distribution function (i.e. the null hypothesis of the test; CDF). Figure 9 shows a graphical representation of this statistic. For the KS test, the alternative hypothesis is always "the distributions are not equal".

The equality of the distributions is rejected if the observed statistic of the test is larger than the quantile at 95% of the Kolmogorov distribution that is provided in a table (see Smirnov, 1948). In the LPIS QA context, the quantiles are 6% for 500 samples, 4.8% for 800 samples and 3.8% for 1250 samples.

Figure 9. Illustration showing rational of the Kolmogorov-Smirnov statistic



2.4.4 Percentage of the population in the 95% central probability interval (PCPI)

Because the KS test is so comprehensive, the small sample sizes from LPIS QA make it virtually impossible to pass. Increasing these sample sizes would be incompatible with the desire to retain cost-effectiveness. Hence a PCPI is introduced to look only at the part of the distribution that really matters.

For a given distribution, a defined central probability interval (CPI) of level (1-α)% as the interval containing (1-α)% of the population will have the α% remaining units equally located in both sides of the population (i.e. exclusion of the α/2% smallest and the α/2% largest). It is assumed that if a sample is representative of the population, the percentage of the population that is included in the (1- α)% CPI of the sample should be close to (1-α)% (and *vice versa*).

In the particular LPIS QA context, it was decided to define α equal to 5% and to define the index as the percentage of the population that belongs to the 95% CPI of the sample; boundaries that reflect the consumer and producer risk of the ISO2859 series. This percentage is henceforth called PCPI.

Figure 10 illustrates how the PCPI is computed. Simulations based on a full random selection show that the expected absolute difference between the objective of 95% and the PCPI is generally equal to 2% (for the smaller LPIS systems and some larger systems) and 1.25% (for most of the larger systems). These results are presented in Figure 11. However, a marked clustering effect is expected for the larger systems, with a likely increase of the corresponding absolute difference. Following these results, it was decided to fix a 2% tolerance on the PCPI. As a consequence, a sample was considered as representative if the PCPI is contained in the [93% ; 97%] interval.

Figure 10. Visual description of the Central Probability Interval (CPI) and the Percentage in Central Probability Interval (PCPI)

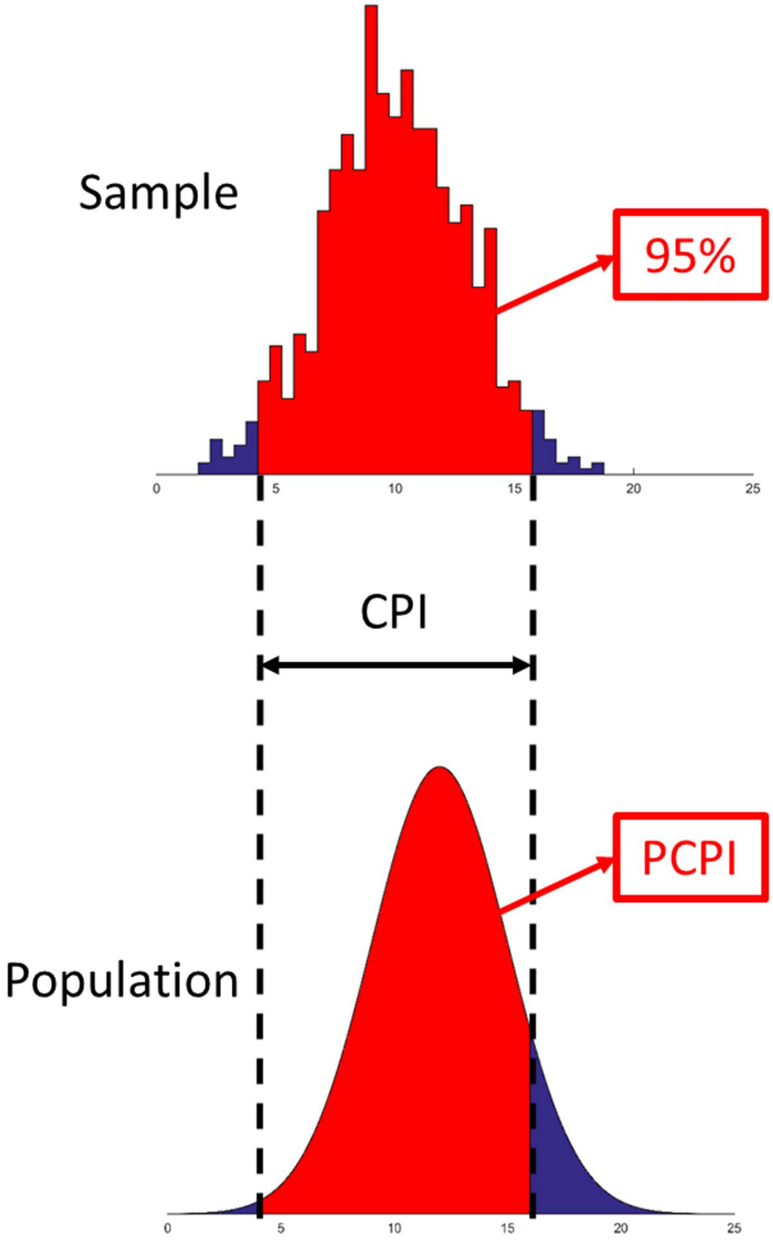
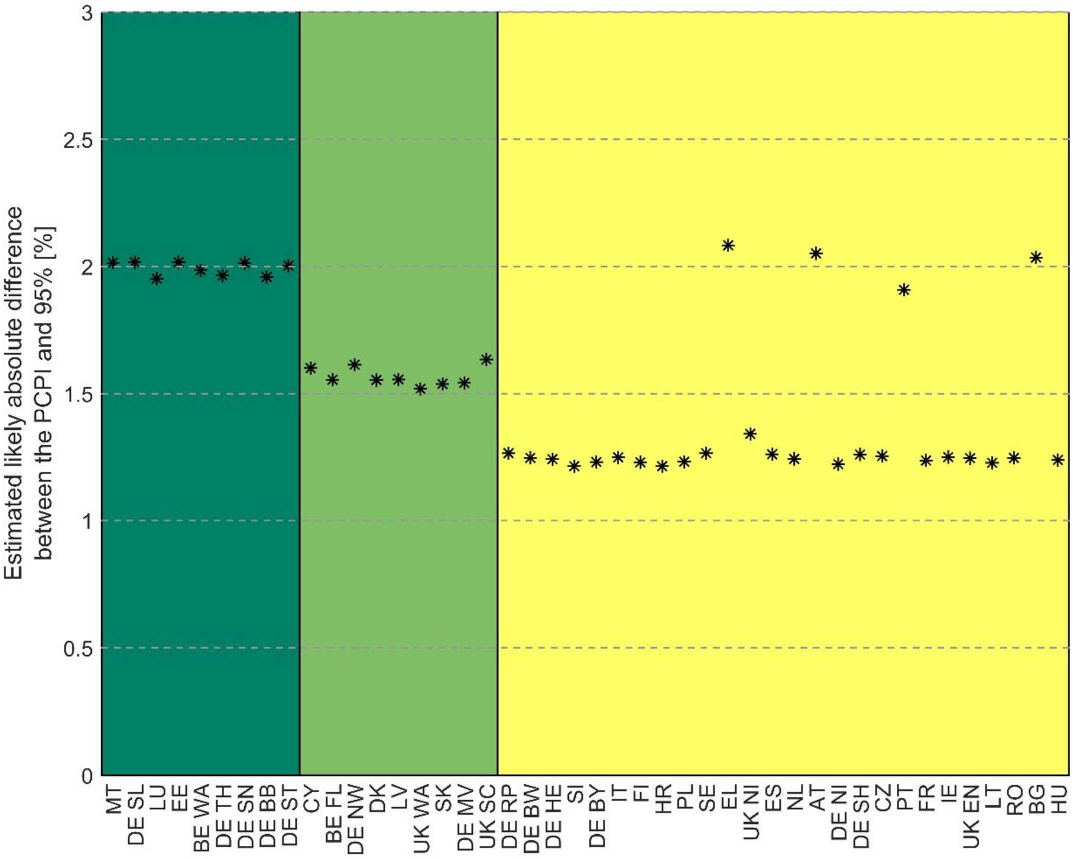


Figure 11. Estimated likely absolute difference between the PCPI and 95% for a fully random sample (i.e. on the whole territory). The LPIS systems are sorted according to their LPIS QA sample size (dark green for 500, light green for 800 and yellow for 1250) and their standard deviation (decreasing order)



2.4.5 Comment on the different indexes

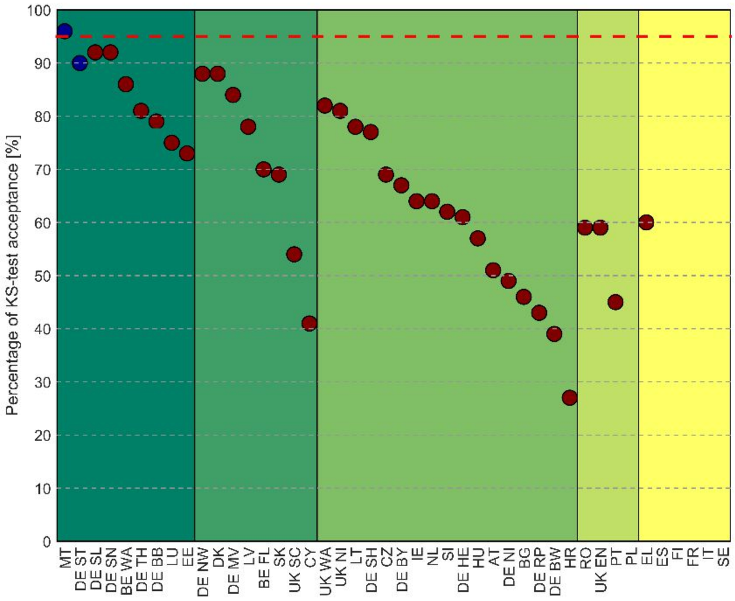
Similarly to the Z-test, the PCPI is sensitive to any shift between both distributions. However, while the Z-test focuses on the bias (i.e. shift between the means), the PCPI index is only focused on the accurate description of the extremes of the distributions (i.e. the tails). The mean of the sample can be different while unchanging the limits of the PCPI.

Similarly to the Chi-square test, the PCPI is sensitive to discrepancies between the population and the sample variances. A sample with a smaller variance is likely to produce a narrow PCPI curve and thus the percentage of the population belonging to this interval will be too small. Reversely, a sample with a larger variance is likely to produce a wide PCPI and thus the percentage of the population belonging to this interval will be too large.

Similarly to the KS statistic and test, the PCPI is thus able to detect discrepancies in terms of the mean and the variance combined. However, while the KS test will reject any discrepancy between the population CDF and the ECDF on the whole domain of these functions, the PCPI is only focusing on the accurate description of the extremes and the central part. The two curves might be different but will pass the test as long as both extremes and the central parts are correctly defined. This is important because, in the conditions of the LPIS QA with the typical sample size, the KS test flags as *not equal* two distributions differing from more than 5% vertically. This condition is difficult to meet for any LPIS QA sample without an unnecessarily large coverage of the LPIS territory. Unnecessary, as a test was ran on the smallest systems only with the technical limit of image acquisition (i.e. the image provider only delivers images with at least 100 km² coverage) showed that in most of the cases, the area coverage that would be needed in order to approach the similar conditions as for a full random selection would be at least 5 times the current area coverage and thus costs (applied for the 2015 and 2016 LPIS QA exercises; see Figure 12).

These partial results indicate that to pass the KS test, the increases of costs would be more than fivefold, without even being sufficient to satisfy the expected 95% acceptance rate. This shows again that the KS test is too challenging and should not be considered as the main driving index for establishing the representativeness of a sample.

Figure 12. Expected minimum increase of image coverage (colors of the circles) and corresponding KS test acceptance rate. Results are presented for the smallest of the 44 LPIS systems. The dashed red line represents the percentage of KS test acceptance for a fully random selection over the whole LPIS territory. Background colors represent the number of images for each LPIS system as implemented in 2015 and 2016



2.4.6 Conclusion: PCPI selected for representativeness assessment

After several testing and simulations and with the concern of keeping the costs under control, the comparisons of the PCPI values were kept as the driving index for measuring the representativeness of the sample because the first 3 indexes were considered respectively too restrictive, not adequate to skewed distributions or unnecessarily sensitive to small discrepancies between the sample and the population. The PCPI values are seen as a balanced solution because they ensure that the whole distribution is covered by the sample while avoiding that the large (in KS terms) sample size of LPIS QA would almost systematically reject the equality of the distributions/means/variances.

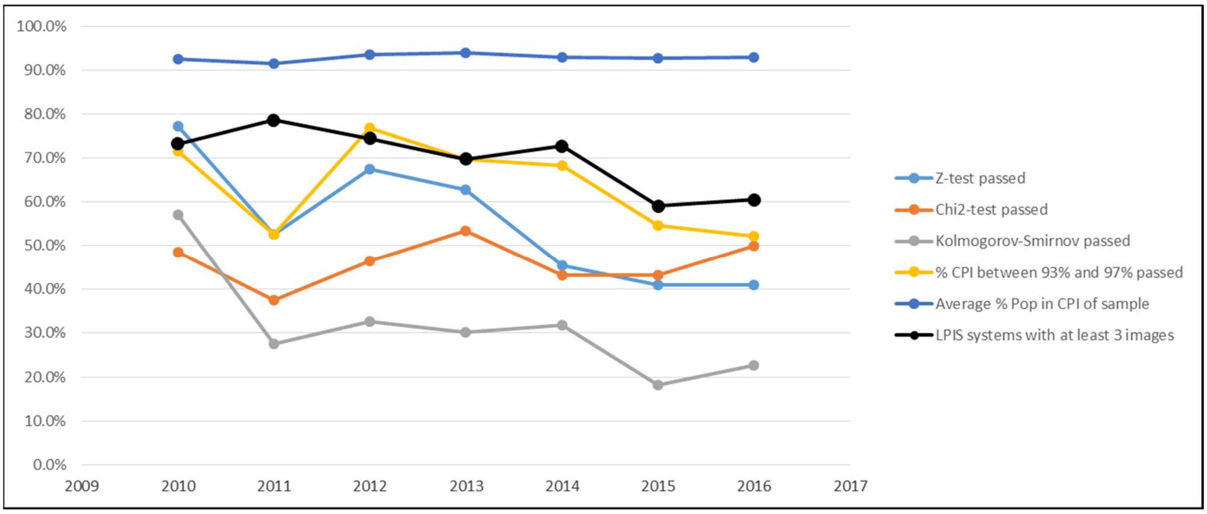
3 Results

3.1 Ex-post representativeness evaluation of the samples over 2010-2016

The different representativeness indexes were computed on the samples generated for the previous LPIS QA campaigns (2010-2016). The two periods corresponding to the two types of image acquisition procedure (i.e. 2010-2014 and 2015-2016) must be considered separately.

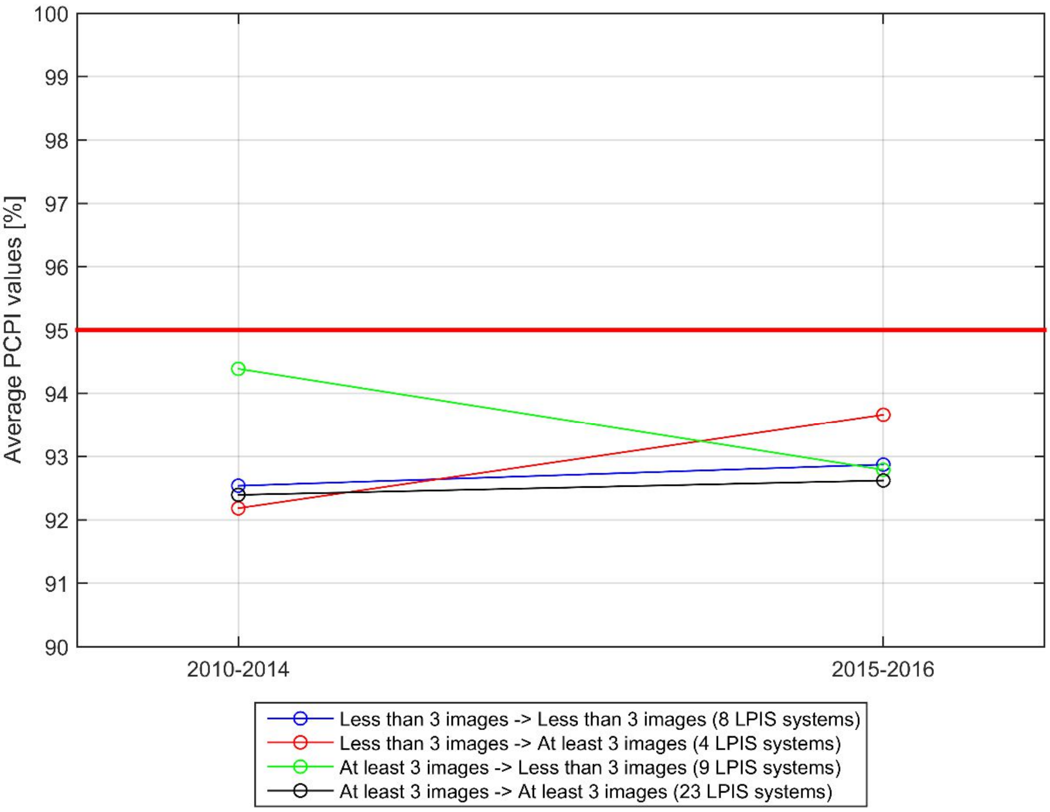
Figure 13 shows the values of the indexes averaged over the MSs for each year between 2010 and 2016. As one can see in this graph, the acceptance rate of the test based on the different indexes are somehow volatile. However, the average PCPI of the sample is relatively stable and close to the expected 93%-97% interval (i.e. 95% with the 2% margin). For the Z-test and the KS-test, the values of the indexes seem to have slightly decreased in 2015-2016. This decreased could be partially explained by the evolution of the percentage of LPIS systems for which the LPIS QA exercise was applied on at least 3 images (between 70 and 75% for the 2010-2014 period and 60% for the 2015-2016 period). This shows that the historical recommendation of minimum 3 images for the purpose of the LPIS QA was indeed relevant. For the remaining indexes, no particular trend can be observed.

Figure 13. Observed acceptance rates for the different indexes over the 2010-2016 period. The number of tested LPIS systems were 36 in 2010, 40 in 2011, 39 in 2012 and 2013 and 44 in 2014-2016



The results presented in Figure 13 must be put in parallel with the drastic decrease of image coverage between both periods 2010-2014 and 2015-2016 (see Figure 1). Obviously, one should recall that the imagery for LPIS QA 2010-2014 was taken from the CwRS images. However, in that light, one can conclude that over-shooting the total imagery coverage like it was the case in 2010-2014 is ineffective since its impacts on the representativeness values are almost completely negligible. Additionally, Figure 14 shows the evolutions of the average PCPI values over both periods according to the "minimum of 3 images per LPIS systems" rule before and after the new LPIS QA image acquisition procedure. One can see that for the LPIS systems where this rule ceased to be applied (9 systems; green curve), the average PCPI decreased. The opposite trend is seen for the LPIS systems where the rule is now applied (4 systems; red curve). For the systems where the rule remained either applied (23 systems; black curve) or not applied (8 systems; blue curve), the averaged PCPI remained almost stable with a slight increase in both cases.

Figure 14. Evolution of the PCPI average values between the periods 2010-2014 and 2015-2016 as a function of the "minimum of 3 images" rule before and after the new LPIS QA image acquisition procedure



3.2 Evaluation of the 2015-2016 sample procedure

The proposed methodology (see sections 2.2 to 2.4 for the details) was applied on the LPIS population data for the 44 different LPIS systems (25 MS, 2 regions in Belgium, 4 countries in United Kingdom and 13 Bundesländer in Germany). The simulations showed that an additional scene was necessary for 7 LPIS systems (namely BE-WA, CY, DE-BB, DE-TH, EE, HR and LU) and two additional scenes for PL only. For the remaining 36 systems, the current image attributions were judged sufficient (see Table 3). With only 8 LPIS systems potentially affected by necessary actions out of 44, one can conclude that the level of representativeness was already satisfying for the majority of systems.

3.3 New density masks for 2017

For the 2015-2016 period, the density of RP was estimated using a 10 km regular grid over the whole Europe. For each cell, several statistics were collected from the LPIS population data (e.g. number of RP, total agriculture area, minimum/maximum/mean parcel area…). This regular grid proved to be useful for working at pan-European level. However, some limitations became evident, namely:

- The 10 km resolution of the grid is workable at pan-European level but is can be coarse when focusing on a particular MS or LPIS system;
- The cross-boundary grid cells were intersected by the administrative boundaries so that the cells straddle on 2 or more LPIS systems, making it difficult to establish the correct density of RP on either side;

— The grid covered unnecessarily large areas of sea/ocean.

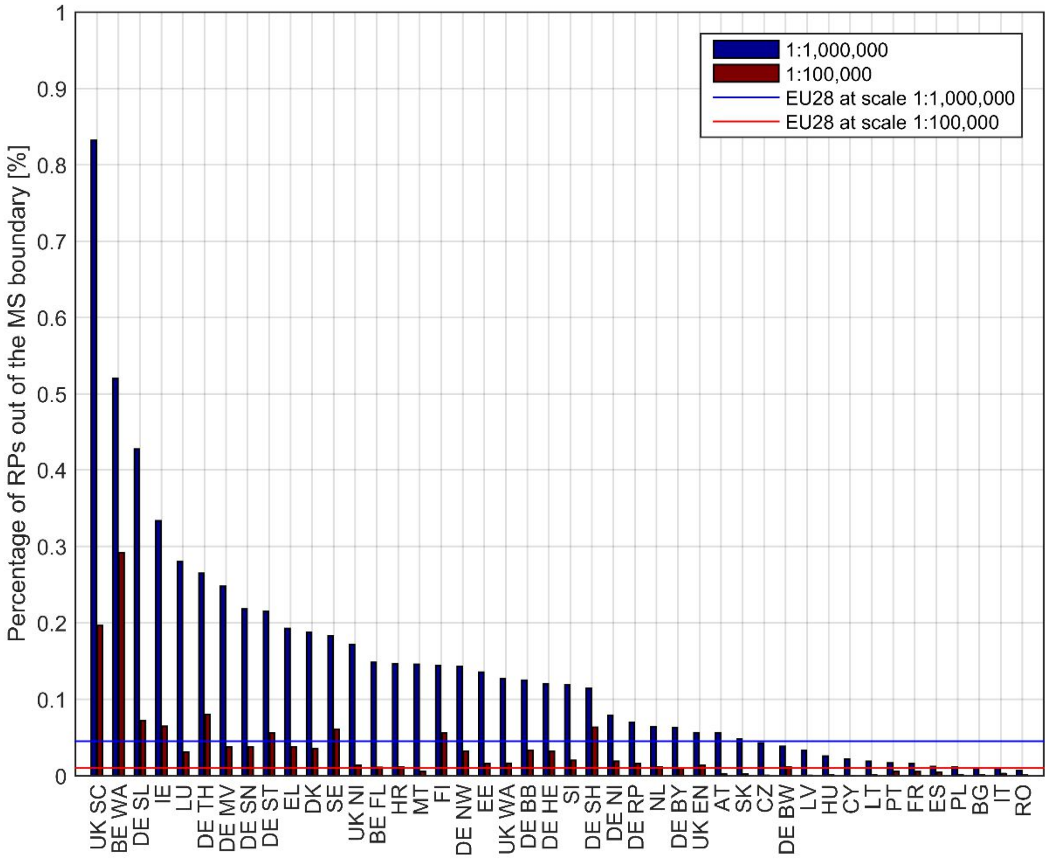
For those reasons, it was decided to produce specific regular grids (i.e. one per LPIS system) and to increase the resolution to 2 km.

As the minimum LPIS zone coverage is 12 km by 12 km (due to technical limitation imposed by the one sensor characteristics), the density is computed using a moving windows of the same size. Consequently, for each point on the regular grid, all the RPs located within 5-7 km the point are counted. Relying on smaller moving windows for the establishment of the density masks (instead of the 15km by 15km coverage of the images that are actually acquired for the LPIS QA exercise) is expected to under estimate the number of the potentially covered RPs. In this way, the chances to sufficiently cover the target sample size is maximized.

As for the previous 2015-2016 period, a limit of 2 RP per km² was set to determine the low and high density of RPS.

Thanks to the finer resolution of the administrative boundaries (1:100,000 instead of 1:1,000,000), the number of RPs not covered by the zonings for image acquisition significantly dropped for each of the 44 LPIS systems (see Figure 15). In terms of percentage per LPIS system, the maximum reached in UK-SC with 0,83% is now reached in BE-WA with 0,29%. The remaining "outside" RPs will still not be sampled during the LPIS QA exercise. However, they represent less than 0,01% of the RPs over the 44 LPIS systems (compared to 0,045% previously) and can easily be targeted during specific audit visits to the corresponding LPIS system.

Figure 15. Comparison of the percentage of reference parcels outside of the administrative boundary when using the 1:1,000,000 scale (blue) and the 1:100,000 scale (red). The percentage at EU28 are also shown for both scales

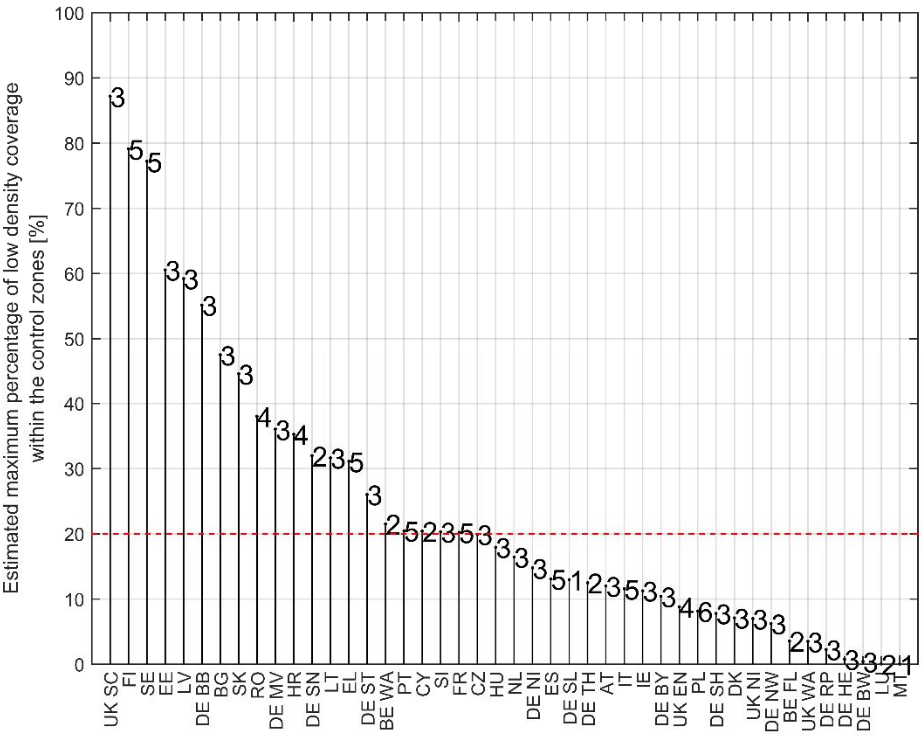


3.4 Assessing the impact of low density mask

Low density masks (i.e. zones where the density of RPs is smaller than 2 RPs per km²) are not equally affecting representativeness for each LPIS system. Some systems are extensively covered by its low density mask (e.g. BG, EE, FR, FI, RO, SE) and some are almost not covered at all because their RPs spatial repartition is almost homogeneous or simply because of the RP design of the system (e.g. cadastral RP type generally have high number of small RPs).

In order to assess the impact of the low density mask, control zones were simulated for each LPIS system. The number of control zones per system was updated accordingly with the results of the simulations (i.e. the final 2017 image provision submission). For each set of simulated control zones, the proportion of the low density mask that was covered was computed (i.e. area of low density on total control zone area). This simulation procedure was repeated 100 times for each LPIS system. Figure 16 shows the result of these simulations. If the percentage of low density area covered had less than 95% chances to be larger than 20%, then the corresponding LPIS system was considered sufficiently homogeneous and both density mask were merged (excluding thus only the parts of the territory without any RP). LPIS total size were also taken into account for merging small LPIS systems (BE-WA, CY and SI) or systems with close to 20% coverage (i.e. CZ, PT). For the remaining 15 LPIS systems (i.e. BG, DE-BB, DE-MV, DE-SN, EE, EL, FI, FR, HR, LT, LV, RO, SE, SK and UK-SC), the two level RP density masks were kept separated; the low density mask was then integrated in the pan-European low density mask.

Figure 16. Estimated percentage of low density coverage within control zones. Segments represent the 95% smallest coverage for each LPIS system while the numbers are the number of control zones. Member states are sorted in decreasing order of this estimated percentage



3.5 Final propositions for image provision for 2017-2018

The results of the simulations on the 2015-2016 zonings suggested an increase of 10 images compared to the 2015-2016 baseline (see Table 3). Those increases mainly concern small systems with 1 yearly allocated image in the 2015-2016 period plus HR and PL.

Following these first results and recalling the 20 years legacy of CwRS, it was clear that relying on a single image might not be appropriate for inferring the result of the sample on the whole population. The reason is that it could highly depend on where this unique zone is located. In order to circumvent this possibility, it was decided to guarantee a minimum of 2-3 images per LPIS system (except for DE-SL and MT for which the representativeness of the samples is already satisfactory and for which the unique image of 15 km by 15 km already represents a significant percentage of their total territory: 9% for DE-SL and 70% for MT). Consequently, for LPIS systems with less than 2-3 allocated images in 2015-2015 and/or for which the simulations did not trigger a sufficient increase, additional images were proposed in order to meet this minimum criterion.

In addition to the change of sampling procedure to a stratified random sample and to the update of the zonings for image acquisition, three different scenarios were thus compared for the 2017 LPIS QA image acquisition procurement (see Table 3 for details):

- "Status Quo": No change of the number of images because of budget limitation and constraints in the image procurement procedure;
- "Partial increase": Increase of the number of images for the smallest LPIS systems and for PL (for a total of 10 images);
- "Full increase": Increase of the number of images according to the different analyses and simulations performed for each LPIS system (for a total of 21 images).

The baseline is also established for the period 2015-2016. For this baseline, not only the number of images was kept unchanged but also the simple random sampling algorithm (i.e. as opposed to the stratified random sampling that is proposed for 2017).

3.6 Re-evaluation of the representativeness under the upgraded situation

The three proposed scenarios were assessed using the same procedure of randomly simulated images as in section 2.2. However, this time the stratified random sampling (section 2.3) replaces the previous simple random sampling. The proposed indexes were all applied in order to evaluate the expected representativeness for each of the scenarios. Figures 19-21 (in Annex) show the values of the acceptance rates for the 2015-2016 initial situation and for the three scenarios for the Z test, the Chi-square test and the KS test respectively. Table 1 shows the expected number of cases where the situation should improve for these three indexes and the three scenarios. One can see that the situation is expected to improve in a majority of LPIS systems (regardless of the scenario). As can be expected, the improvements are the highest for the "Full increase" scenario.

Table 1. Number of LPIS systems out of 44 with an expected improvement of the different indexes and the three different scenarios.

Indexes	Status Quo	Partial increase	Full increase
Z test	16	20	24
Chi-square test	25	27	27
KS test	24	28	28

As previously mentioned, the evaluation of the representativeness is mainly focused on the PCPI. Figure 17 shows the distance between PCPI and 95% for the 2015-2016 baseline and the three scenarios. These distances generally decreased thanks to the change of sampling method, even without any increase of image number. Exceptions were found for the LPIS systems of PT and SK. For PT, this is explained by the particular spatial distribution of the RPs (many medium-small parcels in the North, many small parcels in the middle and few large parcels in the South) that, associated with the image allocation (2 images on the North, 1 image on the middle and 1 on the South), puts too much weight on the smaller RPs. For SK, it is due to its two dominant landscapes with a few large parcels in lower altitudes and many medium-small parcels in higher altitudes. In these conditions, the risk that both allocated images would not contain the large parcels is high. On a second step, the increase of the number of images leads to important improvement of the representativeness for the affected LPIS systems (e.g. CY, DE-ST, DE-TH, EE, HR, PT and SK). It is worth noting that the situations in CY and LU improved but remain above the desired 2% limit. However, a further increase of the number of images for these two smaller systems was not considered justifiable.

The PCPI values were then clustered in classes of distance to the 95% objective: "Less than 1% difference", "Between 1-2% difference" and "More than 2% difference". The number of LPIS systems in each class and for the different scenarios is presented in Table 2. A score of the representativeness was assigned in order to quantify the benefits of each of the scenarios compared to the 2015-2016 baseline.

Figure 17. Representativeness evaluated using the PCPI for each LPIS system and for (i) the base line of (2015-2016 in blue), (ii) the updated zonings but with no changes of number of images (Status Quo in green), (iii) the new zonings with a partial increase of the images (Partial increase in black) and (iv) the new zonings with a full increase of the images (Full increase in dashed red)

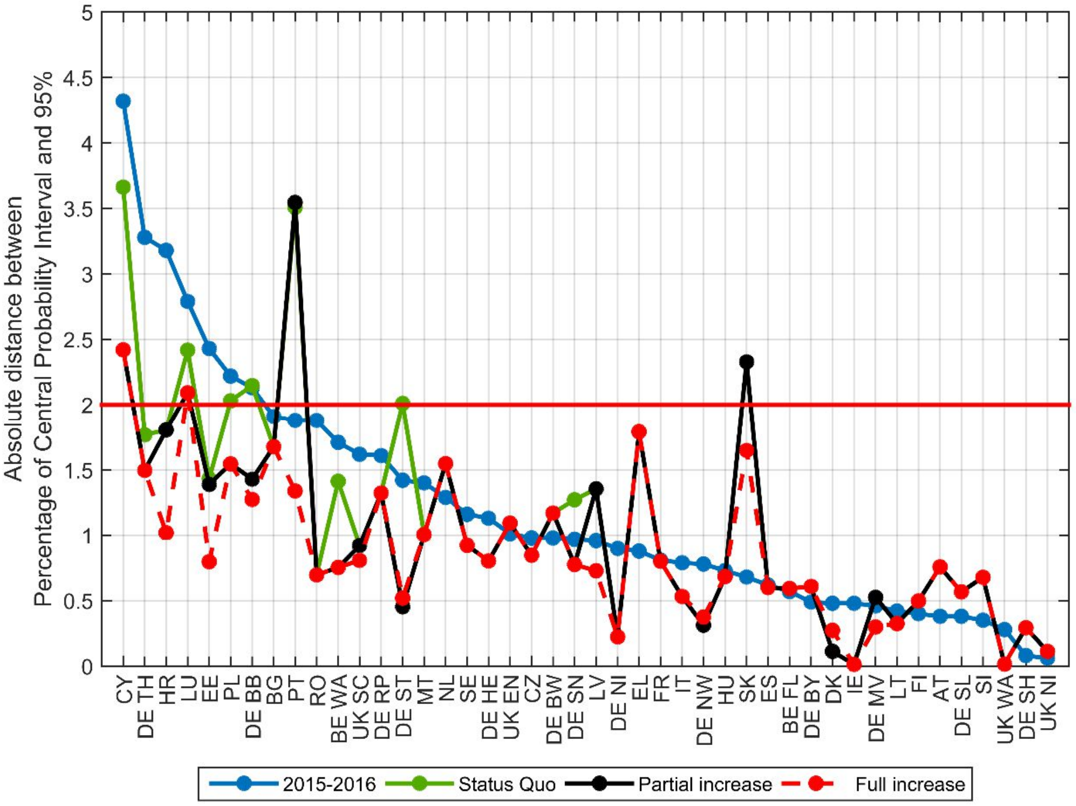


Table 2. Classes of the representativeness based on the PCPI index for the 2015-2016 period and for the different scenarios

Classes of PCPI	2015-2016	Status Quo	Partial increase	Full increase
Good (less than 1% difference)	26	25	28	31
Intermediate (between 1-2% difference)	11	14	12	11
Poor (more than 2% difference)	7	5	4	2
Score (Good=2, Intermediate=1, Poor=0)	63	64 (+2%)	68 (+8%)	73 (+16%)
Total images	123	123 (+0%)	133 (+8%)	144 (+17%)

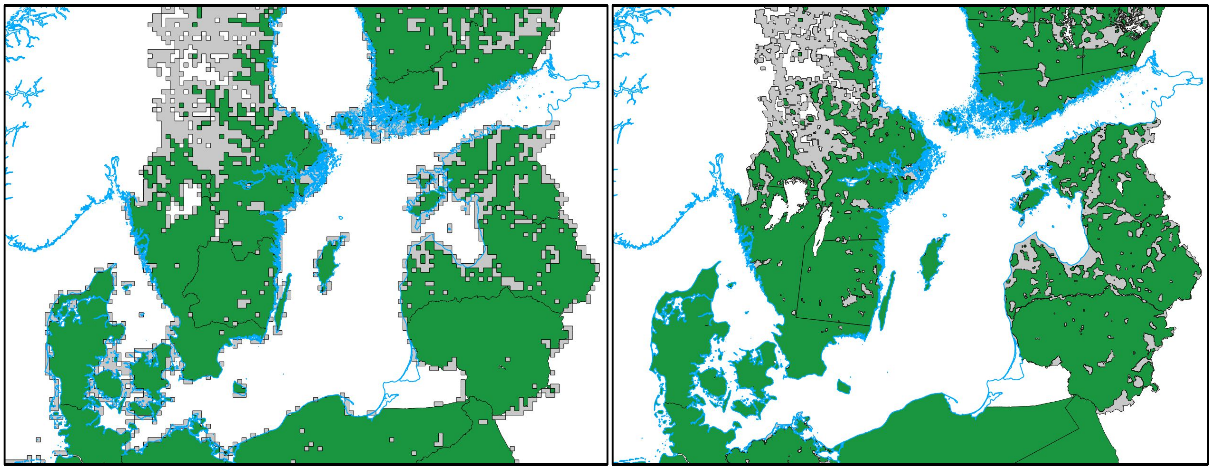
Following these results, it seems opportune to adopt the "Full increase" scenario for 2017.

4 Impact on the 2017 LPIS QA campaign

The findings of this analysis were completely embedded in the 2017 LPIS QA campaign. The resulting changes and the consequences for 2017 are:

1. New NUTS polygons with better scale (from 1:1.000.000 to 1:100.000; EuropeanBoundaryMap© v9.1 from GISCO);
2. Density computed at MS level not full Europe, avoiding "mixed cells" ;
3. Pan EU mask only with those MSs that are "the most affected by abundance of low RP density" (i.e. proportion of low density has 95% chances to exceed 20% of the total images coverage"; see the details in section 3.4);
4. The remaining MSs are considered "sufficiently homogeneous" (i.e. low and high density shapefiles are merged);
5. Specific acquisition zones were created for expanses that have a short available window for image acquisition to ensure that images are equally probable to be acquired in less favoured conditions (i.e. high alpine zones in AT and DE-BY with a long persistent snow cover);
6. This results in 86 final acquisition zones compared to 78 in 2015-2016 (see Figure 18 for an illustration of the improved definitions of the zonings);
7. The samples will now be generated using a stratified random sampling procedure with the different images as strata;
8. An increase of 21 images was adopted in line with the results of the representativeness analyses shown in the previous sections (details of the changes can be seen in Table 3).

Figure 18. Illustration of the improvements in the zoning definition from the 2015-2016 period (left) and the 2017 exercise (right)



As the generation of the acquisition zones is now semi-automatic, yearly updates for 2018 (using the most recent LPIS population data) are foreseen in Autumn 2017. However, this operation could be affected by any delay of some MSs for sending their LPIS population data.

5 Continued monitoring

As previously explained, three member states (Belgium, Germany and United Kingdom) operate and quality assess their LPIS at regional level. Consequently, the corresponding LPIS systems are processed separately during the LPIS QA. At the same, some large member states are also managing their territories through several paying agencies but perform one common LPIS assessment. The underlying assumption is that their LPIS system is following the same recipe (often based on a national cadastre) regardless of the paying agencies. The veracity of this assumption could be further investigated by comparing the QEs according to the paying agencies of these countries. If the assumption is proven untrue (e.g. by unexpected regional variations in the cadastre), the consequence would be that the sampling would be performed at the level of the paying agencies and not at the level of the whole Member State. In such case, both the number of images and the total number of LPIS QA inspections would increase. For instance, in Germany, the total number of images in the "Full increase" scenario is 35 and the number of inspections is 11,600. By comparison, should Italy be split according to its 9 paying agencies, the allocation of images in Italy would increase from 5 to about 22 (an average of 2.5 images per paying agency) and the number of inspections would probably be around 12,500. The reasoning is the same for Spain that, with its 17 paying agencies, the image allocation would increase from 5 to about 42 and a number of inspections that would amount to around 21,250.

Furthermore, some stakeholders observed during the first years of the new image acquisition process that some images of the same LPIS system were quite close to each other (e.g. one image acquired in 2016 is located relatively close to an image acquired in 2015). Although this is not an issue regarding the random selection of the control zones and thus also of the RPs for inspection, it might generate the subjective perception that the LPIS QA exercise is targeting particular locations and not sufficiently covering some parts of LPIS.

6 Conclusions

After its 2016 performance audit on the Land Parcel Identification System (LPIS), the European Court of Auditors recommended that the Commission should carry out a cost-benefit analysis in order to determine whether the representativeness of QA samples could be improved so that a better coverage of the population of parcels in LPIS be achieved.

To address the analysis, an appropriate indicator for representativeness was developed and benchmarked against several know statistical tests. Indeed, the existing statistical tests where either too focused or too demanding to be applicable for the LPIS conditions. The new indicator, "percentage of the population in the 95% central probability interval (PCPI)" was then applied to the actual samples of the past to measure their performance. The results showed that representativeness was acceptable, but could suffer if less than three clusters/zones were applied to some populations.

At the same time, some performance issues from the 10km resolution density maps used for stratification were addressed by upgrading this mask to a 2km resolution. Beyond the resolution of the stratification control layers, JRC also modified the core sampling algorithm from full random to stratified random sampling.

A simulation of this sampling scenario, including the uplift to 3 or more zones and stratified random sampling, demonstrated that the new methodology remediated most territories where sub-optimal conditions could occur. Due to the relatively small size in combination with the parcel distribution, only Malta and Cyprus still do not meet the set indicator threshold, but only just.

These results reconfirmed that the initial approach of spatial stratification does not interfere with the representativeness in se, but it also found that the current implementation could be improved.

All remedial actions combined, led to a 15 percent increase of the number of control zones over Europe, but did not require any procedural modification or other additional inspection workload for the Member States. This revised sampling methodology thus considers the Court's cost-benefit concerns and has been fully implemented for the 2017 LPIS QA campaign.

References

Kenney, J.F. and Keeping, E.S. (1951) *Mathematics of Statistics*. Part Two. 2nd ed. D. Van Nostrand Company, Inc. Princeton: New Jersey.

Kolmogorov, A.N. (1933), Sulla determinazione empirica di una legge di distribuzione, *Giornale dell'Istituto Italiano degli Attuari*, Vol. 4, pp. 83–91.

Sawilowsky, Shlomo S. (2009), New Effect Size Rules of Thumb, *Journal of Modern Applied Statistical Methods*, Vol. 8(2), Article 26.

Smirnov, N. (1948), Table for estimating the goodness of fit of empirical distributions, *Annals of Mathematical Statistics*, Vol. 19, pp. 279–281.

"Student" (1908), The probable error of a mean, *Biometrika*, Vol. 6(1), pp. 1-25.

List of abbreviations and definitions

AP	Agricultural parcel (LPIS RP design type)
CAP	Common Agricultural Policy
CDF	Cumulative density function (statistics)
CP	Cadastral parcel (LPIS RP design type)
CPI	Central probability interval (statistics)
CwRS	Control with remote sensing
DG	Directorate General
EC	European Commission
ECDF	Empirical cumulative density function (statistics)
ECA	European Court of Auditors
EU	European Union
FB	Farmer block (LPIS RP design type)
GIS	Geographic Information System
GISCO	Geographic Information System of the Commission
JRC	Joint Research Centre
KS	Kolmogorov-Smirnov (statistical test)
LPIS	Land parcel identification system (formally: identification system for agricultural parcels)
MS	Member state
NUTS	Nomenclature of territorial units for statistics
OTSC	On-the-spot check
PA	Paying agency
PB	Physical block (LPIS RP design type)
PCPI	Percentage of population in central probability interval (statistical test)
QA	Quality assurance
QE	Quality element
RP	Reference parcel
TB	Topological block (LPIS RP design type)

List of figures

List of tables

Annexes

Annex 1. Supporting figures and tables

Table 3. Summary of the image allocations per LPIS system for the three different scenarios and the estimated values of the PCPI. The colors indicate where the PCPI is smaller than 93% (blue) or between the limits of 93%-97% (green)

	2015-2016		Status Quo		Partial increase		Full increase	
MS	PCPI	Images	PCPI	Image increase	PCPI	Image increase	PCPI	Image increase
AT	94.6%	3	94.2%	-	94.2%	-	94.2%	-
BE-FL	94.4%	2	94.4%	-	94.4%	-	94.4%	-
BE-WA	93.3%	1	93.6%	-	94.2%	+ 1	94.2%	+ 1
BG	93.1%	3	96.7%	-	96.7%	-	96.7%	-
CY	90.7%	1	91.3%	-	92.6%	+ 1	92.6%	+ 1
CZ	94.0%	3	94.2%	-	94.2%	-	94.2%	-
DE-BB	92.9%	1	92.9%	-	93.6%	+ 1	93.7%	+ 2
DE-BW	94.0%	3	93.8%	-	93.8%	-	93.8%	-
DE-BY	94.5%	3	94.4%	-	94.4%	-	94.4%	-
DE-HE	93.9%	3	94.2%	-	94.2%	-	94.2%	-
DE-MV	95.5%	2	94.5%	-	94.5%	-	94.7%	+ 1
DE-NI	94.1%	3	94.8%	-	94.8%	-	94.8%	-
DE-NW	94.2%	2	94.7%	-	94.7%	-	94.6%	+ 1
DE-RP	93.4%	3	93.7%	-	93.7%	-	93.7%	-
DE-SH	94.9%	3	94.7%	-	94.7%	-	94.7%	-
DE-SL	94.6%	1	94.4%	-	94.4%	-	94.4%	-
DE-SN	94.0%	1	93.7%	-	94.2%	+ 1	94.2%	+ 1
DE-ST	93.6%	1	93.0%	-	94.5%	+ 1	94.5%	+ 2
DE-TH	91.7%	1	93.2%	-	93.5%	+ 1	93.5%	+ 1
DK	94.5%	2	94.9%	-	94.9%	-	94.7%	+ 1
EE	92.6%	1	93.6%	-	93.6%	+ 1	94.2%	+ 2
EL	94.1%	5	96.8%	-	96.8%	-	96.8%	-
ES	94.4%	5	94.4%	-	94.4%	-	94.4%	-
FI	94.6%	5	94.5%	-	94.5%	-	94.5%	-
FR	94.2%	5	94.2%	-	94.2%	-	94.2%	-
HR	91.8%	3	93.2%	-	93.2%	-	94.0%	+ 1
HU	94.3%	3	94.3%	-	94.3%	-	94.3%	-
IE	94.5%	3	95.0%	-	95.0%	-	95.0%	-
IT	94.2%	5	94.5%	-	94.5%	-	94.5%	-
LT	94.6%	3	94.7%	-	94.7%	-	94.7%	-
LU	92.2%	1	92.6%	-	92.9%	+ 1	92.9%	+ 1
LV	94.0%	2	93.6%	-	93.6%	-	94.3%	+ 1
MT	93.6%	1	94.0%	-	94.0%	-	94.0%	-
NL	93.7%	3	93.5%	-	93.5%	-	93.5%	-
PL	92.8%	4	93.0%	-	93.5%	+ 2	93.5%	+ 2
PT	93.1%	4	91.5%	-	91.5%	-	93.7%	+ 1

RO	93.1%	4	94.3%	-	94.3%	-	94.3%	-
SE	93.8%	5	94.1%	-	94.1%	-	94.1%	-
SI	94.7%	3	94.3%	-	94.3%	-	94.3%	-
SK	94.3%	2	92.7%	-	92.7%	-	93.4%	+ 1
UK-EN	94.0%	4	93.9%	-	93.9%	-	93.9%	-
UK-NI	95.1%	3	94.9%	-	94.9%	-	94.9%	-
UK-SC	93.4%	2	94.1%	-	94.1%	-	94.2%	+ 1
UK-WA	94.7%	3	95.0%	-	95.0%	-	95.0%	-
AVERAGE	93.8%	121	94.0%	0	94.2%	+10	94.3%	+21

Figure 19. Comparison of the acceptance rate for the Z-test between the baseline of 2015-2016 and the three different scenarios

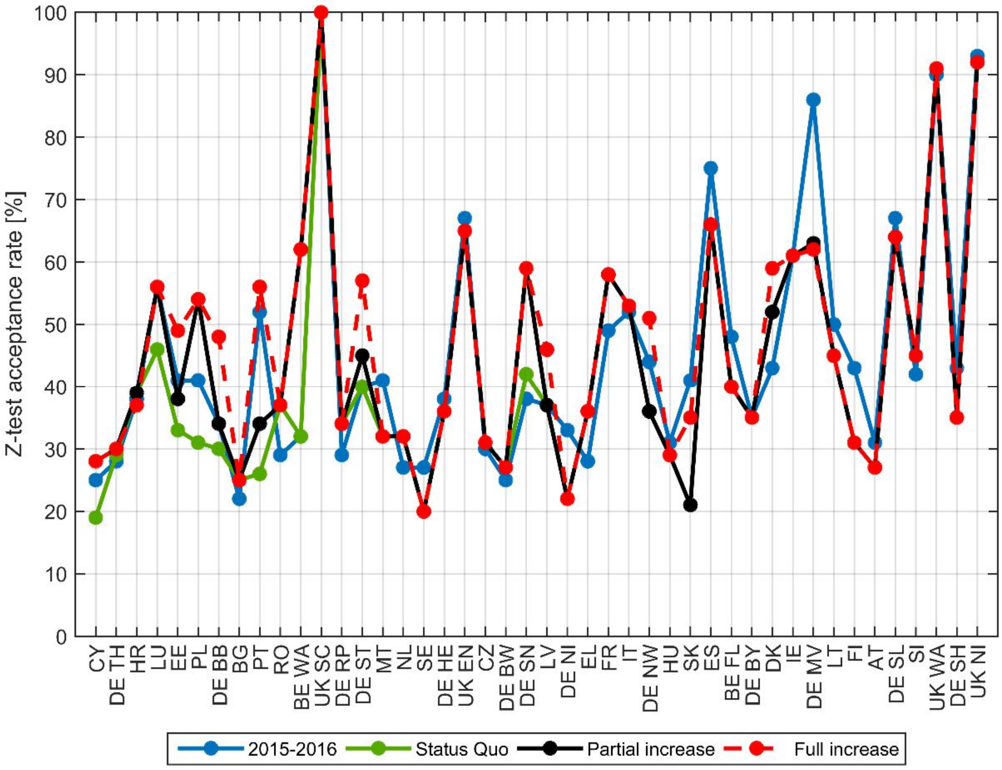


Figure 20. Comparison of the acceptance rate for the Kolmogorov-Smirnov test between the baseline of 2015-2016 and the three different scenarios

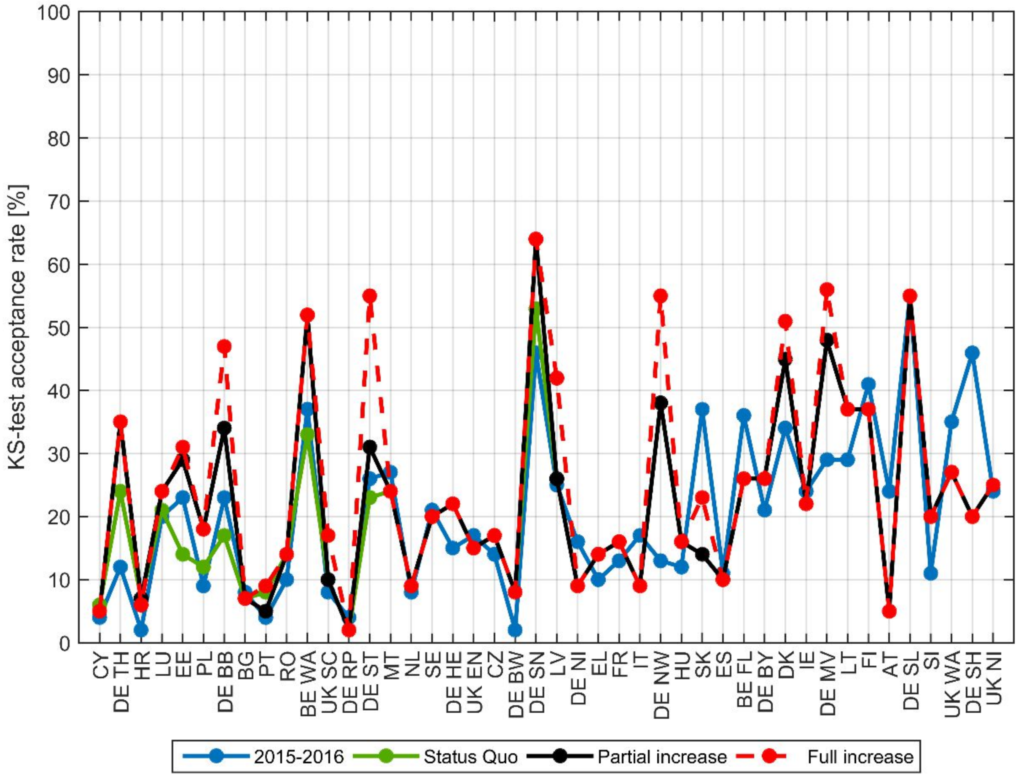
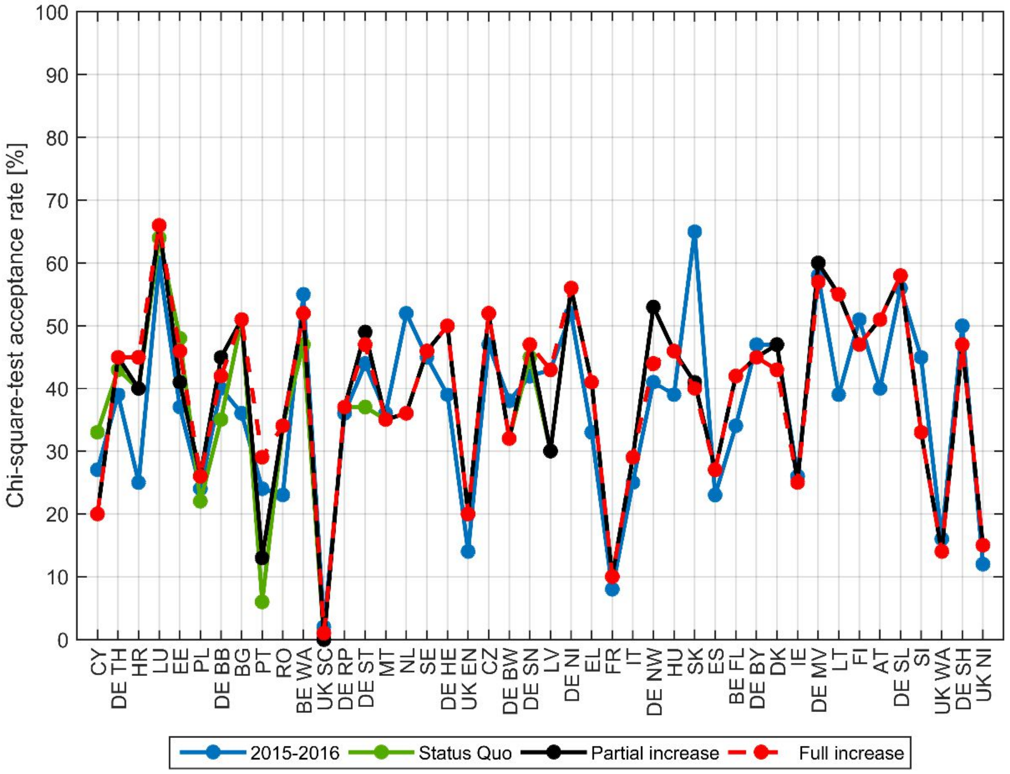


Figure 21. Comparison of the acceptance rate for the Chi-square test between the baseline of 2015-2016 and the three different scenarios



KJ-NA-28755-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/458367

ISBN 978-92-79-73181-5