European Commission

JRC TECHNICAL REPORTS

Data Evaluation of LUCAS Soil Component Laboratory Data for Soil Organic Carbon

Survey periods 2009, 2012 and 2015

Roland Hiederer

2018

Joint Research Centre

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information
Name: Roland Hiederer
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 440, 21027 Ispra (VA), Italy
Email: roland.hiederer@ec.europa.eu
Tel.: +39 0332 78 57 98

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

JRC112711

Luxembourg: Publications Office of the European Union, 2018

How to cite this report: R. Hiederer, *Data Evaluation of LUCAS Soil Component Laboratory 2009 to 2015 Data on Soil Organic Carbon*, Publications Office of the European Union, Luxembourg, 2018, 63pp. JRC112711

Table of contents

Acknowledgements

The author would like to thank Oihane Fernández-Ugalde (JRC), Marc van Liedekerke (JRC), Luca Montanarella (JRC), Beatrice Eiselt (ESTAT), Alessandra Palmieri (ESTAT) and Simone Scarpa (GFT Italia) for their support in finding the originally submitted data and clarifications on some of the more intricate aspects of the LUCAS Soil component handling of samples and data transmission. Many thanks to Bernd Gawlik (JRC) for his explanations of laboratory methods and limits associated with the analysis of soil carbon.

Abstract

With results of the analysis of the soil samples collected under the 2015 LUCAS Soil component becoming available it is possible to evaluate changes in soil properties to previous soil component surveys of 2009 and 2012 on plot locations with repeated visits.

Changes in organic carbon content could be compared for 14,237 repeated samples collected in 2009 and 2015 and for 1,197 repeated samples from 2012 and 2015. For the pool of all samples from repeatedly visited locations no change in organic carbon content could substantiated. However, when introducing land use as a factor some changes in organic carbon content could be supported by the data. Using information from the main LUCAS survey in the analysis of soil conditions would appear to improve the significance of the data for estimating changes in organic carbon. It would also seem necessary to separate samples from mineral from organic soils in the analysis of changes in organic carbon content.

The evaluation found some inconsistencies in the data within and between surveys concerning the representation of conditions when the results of the measurements were below the detection limit of the instrument or method used. Using only the results of the analysis of the soil samples it would appear impracticable to estimate changes in soil organic carbon content below 10% (cultivated land) or 50% (semi-natural areas). By combining the soil data with land use observations from the main LUCAS survey these levels of detectable changes can be improved.

1 Introduction

Since 2006 the *Land Use and Coverage Area frame Survey* (LUCAS) is carried out at a 3-year interval in the Member States of the European Union to collect land cover and land use data over the whole of the EU's territory[1]. The survey follows a stratified sample design where data are collected by direct observations of about 270,000 points (2012 survey) made by surveyors on the ground information (*in situ*). Over time the focus of the survey shifted from providing information on cropland areas to a cover more generally land cover, land use and landscape. With these changes modifications of the classification specifications and nomenclatures were introduced. Information on the current 2018 LUCAS survey are available from Eurostat at http://ec.europa.eu/eurostat/web/lucas/data/primary-data/2018.

In 2009 a *LUCAS Soil Component* (LUCAS Soil) was added to the survey on land cover/use. The initial motive for the LUCAS Soil Component survey was to collect data on soil organic carbon (SOC) with the aim of providing a quantifiable assessment of SOC content and, based on repeated surveys, an appraisal of the function of soils to act as sources or sinks of atmospheric CO_2. Over time the scope of the LUCAS Soil Component survey increased and additional parameters for data collection and analysis were included[2]. The soil component of the LUCAS survey probably provides the most comprehensive data on soil properties collected under standard conditions at European level (Orgiazzi, et al., 2017).

The 2009 survey in Malta was financed by national funds and includes observations on land cover and use, while the survey for Cyprus was restricted to the soil component. For other countries not included in the 2009 LUCAS soil survey (Bulgaria and Romania) data were collected in 2012. In 2015 the survey was repeated with the aim to cover all EU Member States, this time including all EU28 Member States (Croatia was not included in 2009 a 2012 soil surveys).

With the availability of the 2015 LUCAS Soil data temporal changes in SOC content in the topsoil can be produced for locations where sampling has been taken place in 2009 (or 2012). For consistent and reliable statistics on changes in SOC content data from the two surveys should be comparable. To this purpose a task on evaluating the data on organic carbon (OC) was undertaken as part of a data quality assessment of LUCAS Soil. The methods used, the analysis performed and the results obtained are presented in this report.

Scope of Data Evaluation

A data validation is the first stage in an assessment of data quality. Data are examined and a quality indicator is assigned, but data are not modified or removed. The result is therefore not an assurance of quality.

The objective of the LUCAS Soil data evaluation is the appraisal of the aspects:

- accuracy;

[1] URL: http://ec.europa.eu/eurostat/web/lucas

[2] URL: https://esdac.jrc.ec.europa.eu/projects/lucas

- completeness;
- consistency.

Other aspects of a data quality, such as timeliness, believability or interpretability are not considered relevant to the aim of the task.

In line with the provision of retaining the data in the original state any aspects of data cleaning are limited.

- Missing data: identified, but no substitute is introduced.
- Outliers: identified and highlighted, but not removed.
- Inconsistencies: identified and highlighted, but not removed.

No measures are taken to directly reduce noise in the data. However, data may be excluded from analysis based on conditions that impede a reasonable comparison of data. Such restrictions may be applied when evaluating aspects of consistency, such as the use of a common value for the limit of detection or limiting the comparison of repeated samples to a specific distance between locations.

2 Data Evaluation

2.1 Method

The method for data evaluation follows the steps developed and applied as part of the quality checks of the soil data from other pan-European soil data sets, such as Forest Focus and the BioSoil Demonstration Project (Hiederer, et al., 2008).

Validation is performed in three steps:

- **Compliance**:
  concerns the data format, i.e. the adherence of the data to the data format specifications given in technical documents for data recording and submission.
- **Conformity**:
  concerns the data content, i.e. the values submitted for the parameter under validation for a given survey.
- **Uniformity**:
  concerns the comparability of data between surveys and data from other sources.

When a check highlights a particular condition the message linked to the check is attached to the data concerned. The checks and any parameters used are documented in the database for reference.

2.2 Compliance Checks

Data format specifications for Compliance Checks for the 2009 LUCAS Soil component are taken from tender and meeting documents (JRC, 2009). During the Kick-off Meeting on 08.01.2010, Ispra, some additional specifications were set for 2009 (SGS, 2010). Relevant to OC is that for samples for organic soils ("*very high organic matter*") a texture analysis was not needed. This should be indicated in the comment field (SGS, 2010). In the *Preliminary Test Plan* (SGS, 2010) the limit of quantification for $CaCO_3$ was given as 0.1%, no limit was stated for OC (indirect method used, measured is Total carbon from which CaCO3 concentration is subtracted to obtain OC content).

For the results of the laboratory analysis of the 2012 soil surveys (Bulgaria and Romania) no further details on data formats specifications could be found. The same criteria as for the 2009 data were applied.

For the analysis of the 2015 samples details of requirements are given in the Technical Specifications of the tender (JRC, 2016). Specifications for the data delivery format are given in Section 1.3 Delivery of results. The need to link the [Soil-ID] code to the LUCAS database is explicitly stated, as is the file format and the order of parameters to be reported. These are not principally different from the 2009 format specifications. A change to the order of parameters was presented during the Kick-off Meeting on 07.12.2016. The

central laboratory introduced a leading parameter that defines a [Lab ID] for each sample (SGS, 2016). As a result, the positions of the parameters reported is shifted by one position from the tender specifications.

The specifications taken from the documents for defining checks of data compliance are:

- Annex I to Service Contract 385355 specifies in Chapter 4.1 that the results of the "… *conventional parameters have to be delivered in excel (100% excel-compatible) workbook* …" format. A comparable specification is given for the 2015 data (JRC, 2016).
- The parameters to be included and the order of the parameters along columns of the workbook are specified for both, mandatory and supplementary properties.
- For organic carbon (OC) and calcium carbonates ($CaCO_3$) the results should be reported in g kg^{-1} (specification: "g/kg").
- OC should be reported with 1 decimal, $CaCO_3$ with 0 decimals.
- For the link to the sample plots, which are reported in the LUCAS main survey, the identifiers should be numeric. This is not specified in the 2009 Technical Annex to the contract, but in Eurostat description of the LUCAS data fields.

Not included in the specifications are any details related to the coding of missing data or data formats[3]. Missing data occur frequently for data on the particle size distribution for organic soils. Some checks concerning the prospect of linking the soil analysis to the main LUCAS survey are defined, but were not specified for the deliverables of the soil analysis results.

2.3 Conformity Checks

The Conformity Checks evaluate actual values reported in the data delivered. They are intended to highlight conditions where values are not defined or are outside possible ranges. Codes or identifiers that are not defined generally lead to a warning, while data out of range are considered errors. The range limits are thus set with the aim to exclude impossible values, not to define a range of credible values.

The documents do not specify how missing data are to be represented in the data. From the data it may be assumed that when data for whole samples are missing the sample is absent from the data. When data for a parameter are missing the field contains no entry, which is translated into a NULL entry in the database.

[3] The file *Contents-LUCAS-primary-data-2009-20140618-0.xls* specifies the data format for the field [SOIL_LABEL] as "Number". Yet, not all identifiers of samples for results from the analysis of soil samples comply with a numeric format in column "sample ID". http://ec.europa.eu/eurostat/documents/205002/208938/Contents-LUCAS-primary-data-2009-20140618-0.xls/27fe0910-4150-4299-9e89-8cc0adf73d83

The documents do not specify particular codes for use, such as country codes or data outside the quantification limits of the instrument used. The checks identify any codes used, but in the absence of a reference these are not further evaluated.

2.4 Uniformity Checks

Checks of data Uniformity cover aspects of relationships between parameters for a given survey year and temporal consistency of specifications and values over several surveys. For temporal consistency of values the location of the soil sample plots and data from the LUCA LUC survey are used.

The checks include comparing data for a single parameter over time and the consistency of the relationship between several parameters for a survey year and over time.

2.5 Summary of Checks

A list of the checks for all three evaluation steps is given in Table 1

Table 1: List of checks for evaluation of results from soil sample analysis

Check ID	Check Message	Object Name	Range Limit	Severity
	Compliance			
1	File not in 100% Excel-compatible format.	File	XLS	Warning
2	Soil sample identifier not included as 1st parameter.	Table	1	Warning
3	OC results not included as 8th parameter.*	Table	8	Warning
4	OC results not included as 10th parameter.	Table	10	Warning
5	CaCO3 results not included as 9th parameter.*	CACO3	9	Warning
6	CaCO3 results not included as 11th parameter.	CACO3	11	Warning
7	OC not reported with 1 decimal.	OC	1	Warning
8	CaCO3 not reported with 0 decimal.	CACO3	0	Warning
9	OC not reported in g kg-1.	OC	g/kg	Error
10	CaCO3 not reported in g kg-1.	CACO3	g/kg	Error
11	*Invalid field type for sample data.*	*SOIL_ID*	*ALPHA*	*Error*
12	Character in numeric field.	SOIL_ID	NUM	Error
	Conformity			
13	LUCAS LUC: Invalid code for the observation point identifier.	PLOT_ID	NUM	Error
14	LUCAS LUC: Invalid x-coordinate of plot.	GPS_LAT		Error
15	LUCAS LUC: Invalid y-coordinate of plot.	GPS_LONG		Error
16	LUCAS LUC: Invalid code for E/W.	GPS_EW	CODE	Error
17	LUCAS LUC: Plot not in country.	NUTS0		Error
18	LUCAS LUC: Duplicate identifier for point.	POINT_ID	COUNT>1	Error
19	LUCAS LUC: Duplicate identifier for soil sample.	SOIL_ID	COUNT>1	Error
20	SOIL: Duplicate identifier for point in soil sample data.	SOIL_ID	COUNT>1	Error
21	A soil sample identifier is duplicated.	SOIL_ID	COUNT>1	Error
22	A record is duplicated.		COUNT>1	Error
23	A value is missing for soil sample identifier.	SOIL_ID	BLANK	Error
24	A value is missing for parameter OC.	OC	BLANK	Warning

Check ID	Check Message	Object Name	Range Limit	Severity
25	A value is missing for parameter CaCO3.	OC	BLANK	Warning
26	Invalid code for the Sample Identifier.	SOIL_ID	NUM	Error
27	Invalid code for missing data of OC results.	OC	CODE	Error
28	Invalid code for quantification limit of OC results.	OC	< 2.0	Warning
29	Invalid code for missing data of CaCO3 results.	CACO3	CODE	Warning
30	Invalid code for quantification limit of CaCO3 results.	CACO3	< 1	Warning
31	Code is not in list of defined valid codes.	OC	CODE	Warning
32	Minimum OC value below LOD.	OC	2	Error
33	Maximum OC value not within range.	OC	700	Error
34	Minimum CaCO3 below LOD.	CACO3	1	Error
35	Maximum CaCO3 not within range.	CACO3	1000	Error
	Uniformity			
36	Difference between GPS co-ordinates and OBS_DIST exceeds limit.	MIN_DIST	100	Warning
37	OC value below LOD of CACO3.	OC	1	Error
38	Unusually high CACO3 content for organic OC content (> 200 g kg-1).	CACO3	50	Warning
39	Change in detection limit for OC from 2009 to 2012.	OC	LOD	Warning
40	Change in detection limit for OC from 2009 to 2015.	OC	LOD	Warning
41	Change in detection limit for OC from 2012 to 2015.	OC	LOD	Warning
42	Change in detection limit for CACO3 from 2009 to 2012.	CACO3	LOD	Warning
43	Change in detection limit for CACO3 from 2009 to 2015.	CACO3	LOD	Warning
44	Change in detection limit for CACO3 from 2012 to 2015.	CACO3	LOD	Warning
45	Change of OC not within acceptable variation (+/- %).	OC	15	Warning
46	Change of CACO3 not within acceptable variation (+/- %).	CACO3	15	Warning
47	Change of Total Carbon not within acceptable variation (+/- %)	TOTC	15	Warning
48	Change in OC content without change in land cover/use.	OC	15	Warning
49	No change in OC content after change in land cover/use.	OC	15	Warning
50	Change from organic to mineral soil type.	OC	120, 200	Warning
51	Change from mineral to organic soil type.	OC	120, 200	Warning

* 2009: The order of parameters specified under Section 4.1 of the contract was later modified to allow for the parameter "coarse fragments" to be included, which is reported in column 2.
2015: The order of the parameters specified was altered during the Kick-off Meeting of 07.12.2016 to include the parameter LAB_ID.

Some of the checks concern conditions in the LUCAS main survey data and are marked in the table. These checks concern information on the location of the observation points, which are not recorded in the soil data.

2.6 Evaluation Environment

For the evaluation checks to be applied consistently and results reported in a uniform structure the data has to be stored in common structure and processing arrangement. The common structure is a relational database (RDB), the evaluation was carried out using a relational database management system

(RDBMS), in preference to a spreadsheet application. It represents a processing database and is specifically created to process the LUCAS Soil data. The database can be used as a data repository or for data dissemination, but is not expressly designed for these purposes.

The database model follows the design of other databases for storing physical and chemical properties of soils from sample surveys, such as SPADE/M (Hiederer, *et al*., 2006) or the BioSoil Demonstration Project. The general design of the soil property databases follows the principles of data normalisation. Distinguished as data objects are the plot where a sample is taken, the depth layer and the measured or observed properties. LUCAS Soil only collects soil properties from a single sample at a plot location and for a single layer with a fixed depth. This very much simplifies the database model.

Where possible the data were evaluated in the files provided by Eurostat or the central laboratory. However, to provide qualifiers for individual values the data were transferred to tables of a simple database model. To map the results of the soil sample analysis for a given survey and to compare parameters over time to assess change the soil data has to be linked to the plot data of the main LUCAS survey. There are thus two databases with different sets of attributes and different evaluation tasks. Both tasks would provide separate indicators of the evaluation of the land cover/use and soil data. As a consequence, the results of the soil analysis are linked to the LUCAS observations by two intermediate tables. The tables contain the identifiers that are considered to point to valid parameters in either survey. For the soil data the links are specific by parameter. For the evaluation of the analysis of the soil data the links only cover the parameters organic carbon and $CaCO_3$ laboratory data.

To facilitate more advanced processing of the results of the analysis of the soil samples the data from all submissions were combined and normalised. All data are stored using alpha-numeric format. Where numeric computations are required the data are converted.

3 Data Evaluation Results

3.1 Source Data

LUCAS Soil data are observations of soil properties related to a point in the landscape. As regards the main LUCAS survey the Soil survey is a sub-sample of the main LUCAS survey points. As a consequence, the locations where soil samples were taken are not recorded in the data files containing the soil properties, but are taken to be the nominal locations of the main survey. The geographic locations of the main LUCAS survey and their identifiers are a subset of the 2 x 2 km LUCAS grid, which is publicly available[4].

The actual observation of the main survey, but also the soil survey, may be some distance from the nominal position. For the main survey distance and direction to the nominal grid position are recorded and given in the data files in the field [OBS_DIST][5]. For the locations of soil samples taken outside the LUCAS survey (Cyprus and Malta in 2009) this information is not available.

In the standard configuration data on soil properties are provided by a central laboratory together with a unique identifier for the laboratory sample. The information on the observations made as part of the main LUCAS survey at the sample sites (point) is provided by Eurostat for an identifier taken from the LUCAS Grid. For the 2009 and 2015 LUCAS Soil surveys Eurostat provided the JRC with a table that links identifier of the soil laboratory sample analysed with the identifier of the point of the LUCAS Grid where a soil sample has been taken. However, there are notable exceptions from the standard configuration for data.

LUCAS Soil data originates from several surveys:

- 2009 LUCAS Soil Component for EU23, 20,000 samples;
- 2009 LUCAS Soil Component for EU23, 897 additional samples;
- 2009 exclusive soil and land cover/use survey for Malta;
- 2009 LUCAS Soil for Cyprus;
- 2012 LUCAS Soil Component for Bulgaria and Romania;
- 2015 LUCAS Soil Component for EU28.

None of the data available for these surveys fully agrees with any of the other soil survey data in either format, content or structure. In addition, numerous derivate versions have been generated, using different file formats and diverse reasoning for added processing. An overview of the data available is given in the Annex. No claims for completeness of the data sources are made.

[4] URL: http://ec.europa.eu/eurostat/web/lucas/data/lucas-grid

[5] [OBS_DIST]: Distance between theoretical and reached point in meters.

As a consequence of the particularities of the various surveys and data submitted the transfer of the data to the processing database is presented for each survey.

3.1.1 2009 LUCAS Soil for EU23

The main data of the LUCAS Soil 2009 survey is the file *5.6.LUCAS_Results of 20,000 soil samples.xls*. This would appear to be the original data submitted by the central laboratory. There are derivate versions of the data, such as the data available from the *European Soil Data Centre* (ESDAC) (*LUCAS_TOPSOIL_v1.xls*)[6]. The data available from ESDAC does not represent the original data. One deviation of the data from the original is the substitution of entries that indicate a value below the measurement or quantification limit of the instrument used. Instead, it appears that these non-numeric codes were substituted by numeric values using two methods:

For parameters with quantification limits, including OC and $CACO_3$:

a) Malta and Cyprus: the mean between zero ("0") and the quantification limit is used;

b) EU23 survey data: zero ("0").

Substituting the indicator for a value is a possible approach to treating the data, but a modification of the original data. It can introduce artefacts into the data, if the instrument limit changes over time or from re-analysing the data by a different laboratory the value changes.

According to the contract stipulations the initial results from the laboratory analysis were provided on a monthly basis by the central laboratory. In the final submission the data from 20,000 samples were collected and reported in several files:

- 5.6.LUCAS_Results of 20,000 soil samples.doc
- *5.6.LUCAS_Results of 20,000 soil samples.xls*
- 5.10.LUCAS_Final Report_15June2011.doc
- 5.10.LUCAS_Final Report_15June2011.pdf

The file *5.6.LUCAS_Results of 20,000 soil samples.xls* was used to evaluate the data from the laboratory results. The file format is "XLS" (Microsoft® Excel 97 -2003 Workbook). The data are arranged as a flat table by country, where a record (row) consists of the soil sample ID and each parameter is stored as a field (column). The individual country data are added and stored in a single table. The resulting table cannot be used as a flat data table, because the data headers are repeated for each country. As a consequence, no meaningful data structure for records or fields can be established. A flat data table was generated from the workbook by exporting the data to an ASCII format and importing the data into the database format with all values defined as alpha-

[6] URL: https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data

numeric data format. The non-data entries were then removed by a simple query and the field names were adjusted to the standard names for the parameters reported.

Results from the analysis of additional samples are recorded in the file *LUCAS_RECORD_2011-09-27_Results_897_extra samples.xls*.

The indicator for data under the detection limit of the instrument used is "< 2.0" in the final data ("< 2" in the preliminary data)[7].

The results of the Compliance Checks for the 2009 LUCAS Soil component survey for EU23 are given in Table 2.

Table 2: Results of Compliance Check for 2009 LUCAS Soil Component for EU23

Check ID	Result	Check Label	Check Message
1	OK	EXCEL_COMPATIBLE	File not in 100% Excel-compatible format.
2	OK	POSITION_SAMPLE	Sample identifier not included as 1st parameter.
3	OK	POSITION_OC_2009	OC results not included as 8th parameter
5	OK	POSITION_CACO3_2009	$CaCO_3$ results not included as 9th parameter
7	OK	DECIMAL_OC	OC not reported with 1 decimal.
8	OK	DECIMAL_CACO3	$CaCO_3$ not reported with 0 decimal.
9	OK	UNIT_OC	OC not reported in g kg^{-1}.
10	OK	UNIT_CACO3	$CaCO_3$ not reported in g kg^{-1}.
12	Warning 2.1%	CHAR_IN_NUM_FIELD	Character in numeric field.

Note:
Check No. 11 (VALID_FIELD_TYPE_SAMPLE) is not applicable for the type of file format (XLS).

The file is compliant with the specifications in almost all aspects evaluated. Since the file format does not impose a format type for a parameter characters can be stored in parameters that should be numeric. The only field for which a numeric format is required is the sample identifier. A character in the sample identifiers was found for 425 records. The samples with inconsistent labels can be assigned to countries, because the data file *5.6.LUCAS_Results of 20,000 soil samples.xls* contains an additional column with reference to the country. Sample identifiers with characters were found for samples for 9 countries. A summary is given in Table 3.

[7] One difference noted between monthly data tables and the final data is the indicator given for the limit for measuring Phosphor (P): the monthly data files indicate "< 5,0" while the final data contains the indicator "< 10.0". The use of a "." (point) or a "," (comma) as decimal separator is consistent in the data, but not all indications of measurement limits. This is not considered of relevance to this data evaluation.

Table 3: Character in sample identifier of 2009 LUCAS Soil Component results for EU23

Country	2009 Soil Samples		
	Total	*with character*	*Relative (%)*
Belgium	115	38	33.0
Czech Republic	420	8	1.9
Germany	1,878	3	0.2
Estonia	227	2	0.9
Finland	1,735	54	3.1
Hungary	470	1	0.2
Netherlands	205	2	1.0
Poland	1,699	4	0.2
United Kingdom	1,018	313	30.7
Total	**7,767**	**425**	**2.1%***

* For total of 20,000 samples.

Characters are included to the sample code using distinct methods:

- Append "A" and "B" to one sample label: 68 cases
- Two codes are combined by "-": 57 case, 52 in Finland
- Append "UK" to single sample label: 311 cases, all in United Kingdom

Some combination of adding characters also exist. In addition to samples assigned to countries the result table indicates that a barcode was missing for 7 samples. In total, 2.1% of the samples cannot be assigned to a numeric format, as specified by Eurostat for the link field. The use of a numeric format was not defined in the specifications for the deliverable by the central laboratory and should not be considered as non-compliant data.

In addition, for two samples the length of the label is 8 characters. It would appear that the point identifier was used instead of the sample label. Both samples were described as having missing barcodes.

The reasons for using characters in the sample identifier are varied and explained in the documents provided by the central laboratory. Some data may be recovered and assigned to a LUCAS point, but this requires manual modifications of the data. Such modifications are purposely excluded from the data evaluation task.

3.1.2 2009 LUCAS Soil Component for EU23, 897 Additional Samples

An additional 897soil samples were analysed beyond the 20,000 of the EU23 survey. The results of the analysis was reported in a separate file named *LUCAS_RECORD_2011-09-27_Results_897_extra samples.xls*.

The parameter code does not comply completely with the codes used in the file containing the final results of the main submission, but corresponds to the order of parameters. The results of the compliance evaluation checks are given in Table 4.

Table 4: Results of Compliance Check for 2009 LUCAS Soil Component for EU23, 897 additional samples

Check ID	Result	Check Label	Check Message
1	OK	EXCEL_COMPATIBLE	File not in 100% Excel-compatible format.
2	OK	POSITION_SAMPLE	Sample identifier not included as 1st parameter.
3	OK	POSITION_OC_2009	OC results not included as 8th parameter
5	OK	POSITION_CACO3_2009	$CaCO_3$ results not included as 9th parameter
7	OK	DECIMAL_OC	OC not reported with 1 decimal.
8	OK	DECIMAL_CACO3	$CaCO_3$ not reported with 0 decimal.
9	OK	UNIT_OC	OC not reported in g kg^{-1}.
10	OK	UNIT_CACO3	$CaCO_3$ not reported in g kg^{-1}.
12	Warning 8.0%	CHAR_IN_NUM_FIELD	Character in numeric field.

Note:
Check No. 11 (VALID_FIELD_TYPE_SAMPLE) is not applicable for the type of file format (XLS).

The file contains the results of 897 soils samples plus one sample from the *BacsT/09* analysis. For 72 cases (8.0%) the sample identifier includes a character.

- Append "UK" to single sample label: 22 cases
- Two codes are combined by "-": 5 cases
- Append "B" to one sample label: 3 cases (no "A" appended)
- Prefix "L" to one sample label: 41 cases

The table does not contain a column where samples are assigned to a country. None of the samples with characters can be linked to the link field of the Eurostat table, which leaves 821 samples with a potential link to a LUCAS plot.

3.1.3 2009 Exclusive Soil and Land Cover/Use Survey for Malta

The soil survey of Malta was organised outside the LUCAS main survey and the LUCAS Soil survey of 2009. It therefore covered the soil survey and a reduced the land cover/use surveys. No information about the survey details or any specifications concerning the sample analysis and data formats were available to the evaluation. It was assumed that the specifications for the 2009 analysis were applicable.

It was further not possible to retrieve all files submitted by the survey contractor or the laboratory. Data from the analysis of the 19 soil samples were taken from the ESDAC LUCAS Soil 2009 data file (*LUCAS_TOPSOIL_v1.xls*). The ESDAC data do not contain a country code for the samples from Malta and the country samples were identified by the soil sample ID and the site coordinates. The ESDAC data uses a numeric value when the property is below the detection limit of the analysis used. In these cases the data deviate from the submitted data and care should be taken in a temporal change analysis. For Malta no values for OC below the detection limit of the instrument used. It may be noted that the quantification limit for extractable phosphorous content indicated in the data file does not correspond to the value given for the "LOD or measurement range" in the document "*Final testing plan LUCAS.pdf*", page 2 (5.0 mg/kg), but to the value given in "*5.10.LUCAS_Final Report_15June2011.PDF*", page 15.

Since the ESDAC data inverts the order of the results for pH($CaCl_2$) with pH(H_2O) the order of the parameters was adjusted to the specifications for the 2009 EU23 survey in the processing file for Malta. The results for the compliance evaluations checks for the 2009 survey in Malta are presented in Table 5.

Table 5: Results of Compliance Check for 2009 Soil Survey in Malta

Check ID	Result	Check Label	Check Message
1	N/A	EXCEL_COMPATIBLE	File not in 100% Excel-compatible format.
2	N/A	POSITION_SAMPLE	Sample identifier not included as 1st parameter.
3	N/A	POSITION_OC_2009	OC results not included as 8th parameter
5	N/A	POSITION_CACO3_2009	$CaCO_3$ results not included as 9th parameter
7	OK	DECIMAL_OC	OC not reported with 1 decimal.
8	OK	DECIMAL_CACO3	$CaCO_3$ not reported with 0 decimal.
9	OK	UNIT_OC	OC not reported in g kg^{-1}.
10	OK	UNIT_CACO3	$CaCO_3$ not reported in g kg^{-1}.
12	OK	CHAR_IN_NUM_FIELD	Character in numeric field.

Note:
No original data could be located, data used from ESDAC;
Check No. 11 (VALID_FIELD_TYPE_SAMPLE) is not applicable for the type of file format (XLS).

The original data for Malta submitted by the central laboratory could not be retrieved. It was therefore not possible to evaluate the formal aspects of the data. From the data in the ESDAC table it would appear that no particular issues concerning data conformity exist for the laboratory results of the 2009 Malta soil data.

3.1.4 2009 Exclusive Soil and Survey for Cyprus

As for Malta the 2009 soil survey of Cyprus was organised outside the LUCAS main survey and the LUCAS Soil EU23 survey of that year. However, in contrast to data from Malta, the results from the analysis of the 90 soil samples from Cyprus were included in the data provided by the central laboratory.

Specific data land cover/use observed at the sites of the soil samples could not be discovered.

The soil property data for Cyprus was taken from the file *5.6.LUCAS_Results of 20,000 soil samples.xls*.

Table 6: Results of Compliance Check for 2009 Soil Survey in Cyprus

Check ID	Result	Check Label	Check Message
1	OK	EXCEL_COMPATIBLE	File not in 100% Excel-compatible format.
2	OK	POSITION_SAMPLE	Sample identifier not included as 1st parameter.
3	OK	POSITION_OC_2009	OC results not included as 8th parameter
5	OK	POSITION_CACO3_2009	$CaCO_3$ results not included as 9th parameter
7	OK	DECIMAL_OC	OC not reported with 1 decimal.
8	OK	DECIMAL_CACO3	$CaCO_3$ not reported with 0 decimal.
9	OK	UNIT_OC	OC not reported in g kg^{-1}.
10	OK	UNIT_CACO3	$CaCO_3$ not reported in g kg^{-1}.
12	OK	CHAR_IN_NUM_FIELD	Character in numeric field.

Note:
Check No. 11 (VALID_FIELD_TYPE_SAMPLE) is not applicable for the type of file format (XLS).

No anomalies in data compliance were found for the 2009 soil analysis data for Cyprus.

3.1.5 2012 LUCAS Soil for Bulgaria and Romania

For Bulgaria and Romania a soil component survey was organised as part of the main LUCAS survey in 2012. The results of the soil analysis were submitted as separate files for each country. The specifications for the analysis and data submission could not be retrieved. Therefore, and because the same central laboratory performed the analysis, it was assumed that they would correspond to those of the 2009 analysis.

The data files used for the evaluation were *Bulgaria.csv* and *Romania.csv*. Files with an XLSX suffix also exist with an earlier date. However, these files seem

to have been generated from other data for the surveys and are not suitable for evaluation.

It could not be established whether the file format for the data should have been as specified for 2009. The file structure differs to some degree from the 2009 data: The first parameter is the LUCAS point identifier, second and third parameters are the projected point coordinates. The sample identifier is given as 17th parameter, followed by geographic coordinates, which in the absence of further information were taken to relate to the observation point. The arrangement of the analysis results follows the 2009 data structure. The header text does not completely agree between the two files, but the differences are restricted to parameters that were not specified to be included in 2009. Due to the extensive similarities the two files were merged and evaluated as a single submission[8].

A summary of the results of the data compliance evaluation of the data for Bulgaria are given in Table 7.

Table 7: Results of Compliance Check for 2012 LUCAS Soil Survey in Bulgaria and Romania

Check ID	Result	Check Label	Check Message
1	N/A	EXCEL_COMPATIBLE	File not in 100% Excel-compatible format.
2	N/A	POSITION_SAMPLE	Sample identifier not included as 1st parameter.
3	N/A	POSITION_OC_2009	OC results not included as 8th parameter
5	N/A	POSITION_CACO3_2009	$CaCO_3$ results not included as 9th parameter
7	OK	DECIMAL_OC	OC not reported with 1 decimal.
8	OK	DECIMAL_CACO3	$CaCO_3$ not reported with 0 decimal.
9	OK	UNIT_OC	OC not reported in g kg^{-1}.
10	OK	UNIT_CACO3	$CaCO_3$ not reported in g kg^{-1}.
12	OK	CHAR_IN_NUM_FIELD	Character in numeric field.

Note:
Check No. 11 (VALID_FIELD_TYPE_SAMPLE) is not applicable for the type of file format (CSV, no delimiter).

The results for OC and $CaCO_3$ are reported with 1 and 0 decimals, as specified for the 2009 data. No indicators of the reporting units are provided, but the range of values indicates that they correspond to the 2009 units (g kg^{-1}).

[8] The country code was added

3.1.6 2015 LUCAS Soil for EU28

The results of the analysis of the 2015 LUCAS Soil component survey available to this evaluation were the files submitted by the central laboratory on a monthly basis from April, 2017 until April, 2018. This covers all results from the groups of organic soils (Group 1), samples from repeated visits to the same site (Group 2) and samples collected at new sites (Group 3).

The files are in XLS format (JRC, 2016), but with a structure that differs from the 2009 data and between submissions. The main changes are:

- An internal sample identifier has been added as first parameter. The sample identifier from the previous survey is recorded in the second column as parameter [Client ID] (SGS, 2016).
- The parameter "Electrical conductivity" has been added to the data for all samples.
- For sites that were visited during a previous soil survey no data on coarse fragments and particle distribution are included.
- No comments or notes are included in the data submitted.

In addition to the analysis results submitted were also some corrections or adjustments. These corrections or adjustments introduce some complications to the evaluation. The corrections and modifications should reasonably change the submitted data. For data submitted until September, 2017 a consolidated file was submitted (Interim_progress_report_September_2017.xlsx). Data were also assembled by Group (1, 2 and 3) at the JRC. However, the data underwent some processing and are therefore not suitable for evaluating the submitted data. Because the amount of adjustments applied to the data could not be established with any detail the data submitted monthly by the central laboratory were analysed. To simplify the analysis all files were merged into a single table. The modifications reported in the files LUCAS_Status_Report_August_2017_corrected_56187.xls, *LUCAS_Status_Report_CEC_March_2018.XLS* and *PSD_results_Group_1_corrected.xls* were included.

The results of the checks applied are summarised in Table 8.

Table 8: Results of Compliance Check for 2015 LUCAS Soil Component Survey

Check ID	Result	Check Label	Check Message
1	OK	EXCEL_COMPATIBLE	File not in 100% Excel-compatible format.
2	Warning	POSITION_SAMPLE	Sample identifier not included as 1st parameter.
4	Warning	POSITION_OC_2015	OC results not included as 10th parameter
6	Warning	POSITION_CACO3_2015	$CaCO_3$ results not included as 11th parameter
7	OK	DECIMAL_OC	OC not reported with 1 decimal.
8	OK	DECIMAL_CACO3	$CaCO_3$ not reported with 0 decimal.
9	OK	UNIT_OC	OC not reported in g kg^{-1}.
10	OK	UNIT_CACO3	$CaCO_3$ not reported in g kg^{-1}.
12	Error 0.0%	CHAR_IN_NUM_FIELD	Character in numeric field.

Note:
Check No. 11 (VALID_FIELD_TYPE_SAMPLE) is not applicable for the type of file format (XLS).

In general, the files submitted do not correspond in all aspects to the specifications, neither those given in the tender documents, not those presented by the central laboratory during the Kick-off meeting. Common conditions of non-conformity in the data are:

- The files submitted used different field names for the laboratory identifier ([Sample ident]) and the soil sample identifier ([Client ID]) than specified in the documents ([Lab ID] and[SOIL-ID]).
- A field containing the country code was added as 3rd parameter.
- Positions of [SAND] and [SILT] are interchanged.
- Positions of [P] and [N] are interchanged.
- For samples that were already analysed during a previous survey (repeated samples) the fields for coarse fragments and texture fractions are missing.
- A field for comments is missing.

All this may not cause any particular problem to processing the data, but are aspects of non-conformity in the data delivered for 2015. The switch in the position of [SAND] with [SILT] and [N] with [P] from the previous version and the specifications of the contract is potentially hazardous. It may be too easily overlooked in the analysis.

The Error in the numeric field concerns 8 occurrences of a character in the sample identifier code. In 6 cases the character was "ö". The remaining cases concern the sample "30133", which exists as "30133A" and "30133B". It would

appear that the character "ö" should have been an "8", since the corresponding identifiers are missing. This indicates an unfavourable condition when reading the labels as a possible source.

3.2 Conformity Checks

The documents do not specify how missing data are to be represented in the data. From the data it may be assumed that when data for whole samples are missing the sample is absent from the data. When data for a parameter are missing the filed contains no entry, which is translated into a NULL entry in the database.

The documents do not specify particular codes for use, such as country codes or data outside the quantification limits of the instrument used. The checks identify any codes used, but in the absence of a reference these are not further evaluated.

3.2.1 2009 LUCAS Soil Component for EU23, 20,000 plus 897 extra Samples

For the results of the analysis of the samples of the 2009 LUCAS Soil component the data from the 897 extra samples were merged with the file containing the results from the analysis of the 20,000 samples. This approach appears reasonable, because the extra data contain results from the same survey and were analysed under the same conditions as the main data. The outcome of the Conformity Checks for the combined data for 2009 (except for Malta) are presented in Table 9.

Table 9: Results for Conformity Checks for 2009 LUCAS Soil Component plus 897 extra samples

Check ID	Result	Check Label	Check Message
21	OK	DUPLICATE_SOIL_ID	A sample identifier is duplicated.
22	OK	DUPLICATE_RECORD	A value is missing for soil sample identifier.
23	OK	MISSING_DATA_SOIL_ID	A value is missing for sample identifier.
24	OK	MISSING_DATA_OC	A value is missing for parameter OC.
25	OK	MISSING_DATA_CACO3	A value is missing for parameter $CaCO_3$.
26	Error 7.8%	VALID_CODE_SOIL_ID	Invalid code for the sample identifier.
27	N/A	VALID_CODE_MISS_OC	Invalid code for missing data of OC results.*
28	OK (0.0%)	VALID_CODE_LOD_OC	Invalid code for quantification limit of OC results.**
29	N/A	VALID_CODE_MISS_CACO3	Invalid code for missing data of $CaCO_3$ results.*
30	OK	VALID_CODE_LOD_CACO3	Invalid code for quantification limit of $CaCO_3$ results.
31	N/A	VALID_CODE_DEFINITION	Code is not in list of defined valid codes.
32	OK	RANGE_OC_MIN	Minimum OC value below LOD.
33	OK	RANGE_OC_MAX	Maximum OC value not within range.
34	OK	RANGE_CACO3_MIN	Minimum $CaCO_3$ below LOD.
35	OK	RANGE_CACO3_MAX	Maximum $CaCO_3$ not within range.

Note:
* no valid code specified.
**code differs between results submitted for main sample data and extra samples.

No duplicate identifiers or records were found in the combined data. Also, no cases of missing entries for either, OC and $CaCO_3$ were apparent. To check for valid codes in the soil data the reference codes for the sample identifier can be taken as the identifiers recorded in the Eurostat data for the LUCAS main survey. The file available to serve as reference is *Ext_Results_2009(26July2011).csv*. Of the 20,896 [9] soil samples in the combined soil file 19,156 have an equivalent code in the LUCAS main survey data. To provide a value for number of valid links the number of soil samples for Malta (19 samples) and Cyprus (90 samples) should be deducted from the total number, because the LUCAS survey was not carried out in 2009. This results in 20,787 soil samples with a potential link to the plot data, of which 1,631 (7.8%) cannot be linked to the LUCAS survey. The use of the code "9" (1 case) as a sample identifier constitutes and error in reporting the [SOIL_ID].

[9] Analysis results from one sample are for reference sample BacsT /09.

According to the documentation for the LUCAS data (Eurostat, 2011) this code is reserved to indicate points in the sample that were replaced.

The check for a valid codes used to report analysis results of a parameter compares the codes found in the data and the specifications. These do not explicitly define the codes for conditions when measurements result in values below the detection limit or contain any reference to code missing data.

For indicating the condition of a measurement below the detection limit for OC data the entry "< 2.0" is used. The limit was specified in the final test plan (SGS, 2010), but was not defined in the preliminary test plan (SGS, 2010). The qualifier was used for 43 samples (0.2%). The data for the 897 extra samples uses the code "< 2,0" for such conditions. This code was used in 3 cases.

To indicate an analysis result for $CaCO_3$ that was below the detection limit of the instrument used the entry "< 1" was used. This was the only code and used for 7518 samples (37.6%).

No values were found below the minimum limit for OC (2.0 g kg^{-1}), neither for the maximum value (586.8 g kg^{-1}). The same applies for Carbonates, where the minimum value is 1 g kg^{-1}, and the maximum value is 944 g kg^{-1}.

3.2.2 2009 Soil and Land cover/use survey for Malta

For the exclusive soil and land cover/use survey in Malta for 2009 a summary of the results of the Conformity checks are presented in Table 10.

Table 10: Results for Conformity Checks for 2009 Soil and Land cover/use survey for Malta

Check ID	Result	Check Label	Check Message
21	OK	DUPLICATE_SOIL_ID	A sample identifier is duplicated.
22	OK	DUPLICATE_RECORD	A value is missing for soil sample identifier.
23	OK	MISSING_DATA_SOIL_ID	A value is missing for sample identifier.
24	OK	MISSING_DATA_OC	A value is missing for parameter OC.
25	OK	MISSING_DATA_CACO3	A value is missing for parameter $CaCO_3$.
26	OK	VALID_CODE_SOIL_ID	Invalid code for the sample identifier.
27	N/A	VALID_CODE_MISS_OC	Invalid code for missing data of OC results.*
28	N/A	VALID_CODE_LOD_OC	Invalid code for quantification limit of OC results.**
29	N/A	VALID_CODE_MISS_CACO3	Invalid code for missing data of $CaCO_3$ results.*
30	N/A	VALID_CODE_LOD_CACO3	Invalid code for quantification limit of $CaCO_3$ results.**
31	N/A	VALID_CODE_DEFINITION	Code is not in list of defined valid codes.
32	OK	RANGE_OC_MIN	Minimum OC value below LOD.
33	OK	RANGE_OC_MAX	Maximum OC value not within range.
34	OK	RANGE_CACO3_MIN	Minimum $CaCO_3$ below LOD.
35	OK	RANGE_CACO3_MAX	Maximum $CaCO_3$ not within range.

* no valid code specified.
** case does not occur in data.

Due to the nature of the 2009 Malta survey the identifiers cannot be linked to the Eurostat reference data. None of the sample identifiers of the soil survey is present in the Eurostat table, which allows merging the data from Malta with the other 2009 soil data.

The data does not contain any codes to indicate results of the analysis that are below the detection limit of the instrument used for OC and $CaCO_3$. The values are further within the ranges for OC (minimum: 2.8 g kg^{-1}; maximum: 28 g kg^{-1}) and for CaCO3 (minimum: 173 g kg^{-1}; maximum: 905 g kg^{-1}).

3.2.3 2009 LUCAS Soil Survey for Cyprus

The results of the analysis for the 90 samples collected in Cyprus were included in the 2009 EU23 data. As a consequence of the nature of the survey, none of the sample identifiers is referenced in the Eurostat reference table. The data contains two samples with OC values below the detection limit and 8 samples

for $CaCO_3$. The value ranges were within the set limits for OC and $CaCO_3$, which were checked in the main 2009 data file.

3.2.4 2012 LUCAS Soil Component for Bulgaria and Romania

The results of the analysis of the soil samples collected in 2102 in Bulgaria and Romania were merged into a single file and then subjected to the Conformity Checks. The data header information differs in some minor detail between the files and was adjusted to the standard naming convention for field names. The total number of samples in the merged file is 2034, (661 from BG and 1373 from RO).

To link the soil data to the LUCAS samples of the main survey the original files contain the identifier of the LUCAS survey. However, no specific file containing the results from the main LUCAS survey was discovered.

The results of are summarised in Table 11.

Table 11: Results for Conformity Checks for 2012 LUCAS Soil Component for Bulgaria and Romania

Check ID	Result	Check Label	Check Message
21	Error 0.3%	DUPLICATE_SOIL_ID	A sample identifier is duplicated.
22	OK	DUPLICATE_RECORD	A value is missing for soil sample identifier.
23	Warning 0.2%	MISSING_DATA_SOIL_ID	A value is missing for sample identifier.
24	OK	MISSING_DATA_OC	A value is missing for parameter OC.
25	OK	MISSING_DATA_CACO3	A value is missing for parameter $CaCO_3$.
26	OK	VALID_CODE_SOIL_ID	Invalid code for the sample identifier.
27	Error 0.7%	VALID_CODE_MISS_OC	Invalid code for missing data of OC results.*
28	Warning 11.4%	VALID_CODE_LOD_OC	Invalid code for quantification limit of OC results.
29	Error 0.4%	VALID_CODE_MISS_CACO3	Invalid code for missing data of $CaCO_3$ results.*
30	Warning 33.4%	VALID_CODE_LOD_CACO3	Invalid code for quantification limit of $CaCO_3$ results.
31	N/A	VALID_CODE_DEFINITION	Code is not in list of defined valid codes.
32	Error 0.7%	RANGE_OC_MIN	Minimum OC value below LOD.
33	OK	RANGE_OC_MAX	Maximum OC value not within range.
34	Error 0.4%	RANGE_CACO3_MIN	Minimum $CaCO_3$ below LOD.
35	OK	RANGE_CACO3_MAX	Maximum $CaCO_3$ not within range.

* no valid code specified.

Of the samples for Bulgaria three are duplicated in the file processed. This brings the number of samples with useable soil data for Bulgaria to 655. Data are missing for four instances of a [POINT_ID] in the file. All instances are for data from Bulgaria.

For Bulgaria the results from 661 samples are recorded in the file. Entries exist for all samples for OC and $CaCO_3$. For the 1,373 samples from Romania data and point identifiers exist, but the sample identifier is missing for 4 samples. The cases concerned are given in Table 12:

Table 12: Missing Sample Identifier in data of 2012 LUCAS Soil Survey in Bulgaria and Romania

Country	POINT_ID
Romania	52582808
Romania	52602800
Romania	54762632
Romania	54762638

This situation is also present in the "XLSX" files. In the absence of the submitted data and a report from the central laboratory a cause for the missing identifiers could not be identified.

In a deviation from other data it appears that a qualifier of "-999" has been used, probably to mark cases were results from the analysis are absent. This entry is given for 15 samples for data from Romania for the OC field. In the absence of the proper documents and under the assumption that the specifications for data coding were no different for 2012 than for 2009 this data coding is treated as an error.

The entry used to mark results of the analysis of OC that were below the detection limit of the instrument used is "<6". This code was used for 231 samples (11.4%; 47 for results from Bulgaria, 184 for results from Romania). This condition leads to a warning, since it is not documented.

A code of "<0.5" was used to mark a result that was below the detection limit of the instrument used for $CaCO_3$. The code occurs as an entry in the field for 679 samples (33.4%; 180 for results from Bulgaria, 499 for results from Romania).

A code of "-999" was used in 8 cases (0.4%) in the field containing the results of the $CaCO_3$ analysis. In one case a result from Bulgaria was given this code, in 7 cases the results were from the analysis of data from Romania.

The use of "-999" is not defined and the use would lead to errors for all instances in the data under the corresponding Conformity Check. This more general check has already been processed specifically for OC and CaCO3 and was net applied.

The check for the valid minimum value for OC is affected by the 15 cases of the undefined entry of "-999". Other than this no values are recorded below 6 g kg^{-1} and the maximum value is 129.1 g kg^{-1}. Correspondingly affected is the check on the minimum $CaCO_3$ value. The minimum value is 1 g kg^{-1} and the maximum 623 g kg^{-1}.

The entry "<0.5", which is used to mark analysis results for CaCO3 below the detection limit of the instrument, is unusual. It is not documented and requires 1 decimal although the data in the field is given with 0 decimals. Consequently, there are no values reported < 1 g kg^{-1}.

3.2.5 2015 LUCAS Soil for EU28

For the evaluation of data Conformity the analysis results from all submitted files were merged into a single file[10]. The XLS format was converted to a CSV and all data were imported as in alpha-numeric format into database tables. One may note that the additional information on the country, where the soil samples were taken, use various spellings, such as "*Luxembourg*" and "*Luxemburg*" or "*U. Kingdom*" and "*UK*". All country names were converted to the standard country codes in the database tables.

The files submitted monthly by the central laboratory were combined to a single table. The following files containing submitted corrections were added to the monthly data:

- *Correction_July_2017.xlsx*
- *LUCAS_Status_Report_August_2017_corrected_56187.xls*
- *PSD_results_Group_1_corrected.xls*
- *LUCAS_Status_Report_CEC_March_2018.XLS*

The structure of the files containing the corrections do not in all cases match the structure of the files containing the submitted analysis results. At times, only the data that changed were included in the file. Introducing the corrected data into the data table amounts to a partial up-date of a record in the data table, which then originally submitted and corrected data.

To store all originally submitted and corrected data in a single table the data were normalised for the parameters. A field was added to denote which data to use in the evaluation.

The results of the Conformity Checks applied to the 2015 soil data are presented in Table 13.

[10] The files *LUCAS_Results_Group_1.xls*, *LUCAS_Results_Group_2.xls* and *LUCAS_Results_Group_3.xls* could not be used for the evaluation. These files do not conform to the specified data structure and omit some of the corrections submitted. It is suggested to clearly indication the nature of the data they contain.

Table 13: Results for Conformity Checks for 2015 LUCAS Soil Component survey

Check ID	Result	Check Label	Check Message
21	Error 0.1%	DUPLICATE_SOIL_ID	A sample identifier is duplicated.
22	OK	DUPLICATE_RECORD	A value is missing for soil sample identifier.
23	OK	MISSING_DATA_SOIL_ID	A value is missing for sample identifier.
24	OK	MISSING_DATA_OC	A value is missing for parameter OC.
25	OK	MISSING_DATA_CACO3	A value is missing for parameter $CaCO_3$.
26	Error 0.9%	VALID_CODE_SOIL_ID	Invalid code for the sample identifier.
27	Warning	VALID_CODE_MISS_OC	Invalid code for missing data of OC results.*
28	Warning	VALID_CODE_LOD_OC	Invalid code for quantification limit of OC results.
29	Comment	VALID_CODE_MISS_CACO3	Invalid code for missing data of $CaCO_3$ results.*
30	Error 49.0%	VALID_CODE_LOD_CACO3	Invalid code for quantification limit of $CaCO_3$ results.
31	Error 0.0%	VALID_CODE_DEFINITION	Code is not in list of defined valid codes.
32	Error 0.3%	RANGE_OC_MIN	Minimum OC value below LOD.
33	OK	RANGE_OC_MAX	Maximum OC value not within range.
34	Error 49.0%	RANGE_CACO3_MIN	Minimum $CaCO_3$ below LOD.
35	OK	RANGE_CACO3_MAX	Maximum $CaCO_3$ not within range.

* no valid code specified.

A duplicate entry for the sample identifier was found for 12 cases. These samples had different laboratory identifiers and data attached to them. There is no obvious pattern to the condition, neither for country nor for the group of data or corrections of analysis results.

No duplicate records were found and all records have an identifier for the samples. There are no cases where either OC or $CaCO_3$ have blank entries.

Invalid sample identifiers are the 8 cases with a character in the code, but also those identifiers that cannot be linked to the LUCAS main survey data. The file used to find valid identifiers was *LUCAS_DMT_2015.accdb*. The file does not contain data sampled outside EU28. As a consequence, any identifiers from results of the analysis of soil samples outside EU28 will be highlighted as missing. In total, a missing link was found for 1,474 soil sample identifiers. For EU28 a total of 203 soil analysis results could not be found in the Eurostat file. The distribution of the countries concerned is presented in Table 14.

Table 14: Cases of identifiers for 2015 Soil Component samples not in Eurostat reference table (LUCAS_DMT_2015.accdb)

Country	Samples without link	
	Number	*%*
Austria	25	4.4
Bulgaria	65	10.9
Estonia	1	0.5
Spain	6	0.2
Finland	1	0.1
Croatia	46	22.7
Hungary	3	0.7
Italy	1	0.1
Latvia	3	1.0
Netherlands	4	2.3
Poland	42	3.1
Romania	2	0.2
Sweden	4	0.2

The highest rates of soil samples without reference in the LUCAS main survey were found for Croatia (22.7%) and Bulgaria (10.9%). The overall rate for EU28 comes to 0.9% of all samples.

As in other years with soil surveys, for the 2015 soil data no codes to be assigned to cases where no data could be reported were specified. For OC in one case the code "NVL" was used, presumably to signify the absence of a result from the analysis (SOIL_ID: 56187; Lithuania). No such code or similar was found for [CACO3].

No code was specified for 2015 data to indicate an analysis result below the detection limit of the instrument used. For OC a code "<0.0" was found in two cases ([SOIL_ID]: 44375 and 44395; Serbia).

For the analysis of CaCO3 no particular code for results below the detection limit of the instrument used were noticed. Instead, a value of zero ("0") was used in 11,288 cases. This amounts to 49.0% of all reported results. The condition has been considered to be erroneous, since the condition should have been coded clearly, such as using the specified detection limit ("< 1.0"), as has been applied in the data of other survey years. The severity of the evaluation outcome is justified by the potential problems that are introduced by using a value zero where such a measurement was not made. This can affect the data for OC, which is not measured directly, but through $CaCO_3$.

The use of the codes for the detection limit and missing data that are not specified ("NVL", "<0.0") generates an error condition in the checks for Conformity. However, since the methods to be applied to indicated these conditions are not specified this cannot be considered non-conform.

The check on the minimum OC content reported is impeded by the lack of the LOD for OC. It would be reasonable to assume a value of 2 g kg-1, as given for

previous years using the same method (Bernd Gawlik, JRC, personal communication). Yet, for 70 samples a value below this limit is reported, with a minimum of 0.1 g kg^{-1}. This results in an error condition for 0.3% of the reported analysis results. The maximum value for OC is 560.2 g kg^{-1}, which is within the range considered feasible.

The use of a value zero ("0") for $CaCO_3$ leads to all cases being classified below the minimum limit for the parameter (49.0%). The maximum value reported for $CaCO_3$ is 976 g kg^{-1}, which is within the limits of the check.

3.3 Uniformity Checks

Checks of data Uniformity cover aspects of relationships between parameters for a given survey year and temporal consistency of specifications and values over several surveys. For temporal consistency of values the location of the soil sample plots and data from the LUCA LUC survey are used. The checks for Uniformity were applied to data that were combined by survey period.

3.3.1 Single Year Cross-Parameter Checks

The values of the OC analysis results were compared to the reported results of the $CaCO_3$ analysis. The results for all survey years are presented in Table 15.

Table 15: Results of Uniformity Checks for Cross-Parameters

Check ID	Result	Survey Period	Check Message
37	OK OK Warning 0.1%	2009 2012 2015	OC value below detection limit of CACO3.
38	Warning 0.1% OK OK	2009 2012 2015	Unusually high CACO3 content for organic OC content (> 200 g kg-1).

For the check of LOD consistency between $CaCO_3$ and OC values for the 2009 and 2012 data from the sample analysis did not reveal any obvious conflicts. For 2015 the data does not contains a value of zero ("0") to signify analysis results below the detection limit of the instrument or method used. According to the documentation the LOD for CaCO3 was 0.1% (SGS, 2016) (1 g kg^{-1}).

In total there were 25 cases (0.1%) where an OC value as the result of the analysis was reported that was below the LOD for $CaCO_3$. An additional 45 (0.2%) cases reported OC values below the LOD for OC from 2009.

For 2009 data 12 cases were reported where for OC content indicating organic soils (> 200 g kg^{-1}) the $CaCO_3$ content exceeds 50 g kg^{-1}. No such cases were reported for 2012 and 2015 data.

In the course of this evaluation no other cross-parameter checks were performed, such as consistency of particle distribution of the C/N ratio. This would have required evaluating these parameters, which was outside the scope of this work.

3.3.2 Multi-Year Code Consistency Checks

The temporal consistency of the codes used to indicated results of the analysis below the detection limit of the instrument used and any missing data was evaluated based on the parameters given in Table 16.

Table 16: Summary of codes used for LOD and instances of missing data for OC and $CaCO_3$

Survey	Parameter	Comment
2009 EU23		Code for missing OC data.
		Code for missing $CaCO_3$ data.
	< 2.0	Code for quantification limit for OC analysis.
	< 1	Code for quantification limit for $CaCO_3$ analysis.
2009 897 Extra		Code for missing OC data.
		Code for missing $CaCO_3$ data.
	< 2,0	Code for quantification limit for OC analysis.
	< 1	Code for quantification limit for $CaCO_3$ analysis.
2009 Malta		Code for missing OC data.
		Code for missing $CaCO_3$ data.
		Code for quantification limit for OC analysis.
		Code for quantification limit for $CaCO_3$ analysis.
2012 BG and RO	-999	Code for missing OC data.
	-999	Code for missing $CaCO_3$ data.
	<6	Code for quantification limit for OC analysis.
	<0.5	Code for quantification limit for $CaCO_3$ analysis.
2015	NVL	Code for missing OC data.
		Code for missing $CaCO_3$ data.
	<0.0	Code for quantification limit for OC analysis.
	0	Code for quantification limit for $CaCO_3$ analysis.

These parameters were retrieved from the contract documentation or additional documents submitted and were used for the valuation of consistency in codes. The results of the evaluation are presented in Table 17.

Table 17: Results of Uniformity Checks for consistent Multi-Year Coding

Check ID	Result	Survey Period	Check Message
39	Warning	2009 and 2012	Change in detection limit for OC from 2009 to 2012.
40	Warning	2009 and 2015	Change in detection limit for OC from 2009 to 2015.
41	Warning	2012 and 2015	Change in detection limit for OC from 2012 to 2015.
42	Warning	2009 and 2012	Change in detection limit for $CaCO_3$ from 2009 to 2012.
43	Warning	2009 and 2015	Change in detection limit for $CaCO_3$ from 2009 to 2015.
45	Warning	2012 and 2015	Change in detection limit for $CaCO_3$ from 2012 to 2015.

The value for the OC LOD changed from 2009 (< 2.0 g kg^{-1}) to 2012 (< 6.0 g kg^{-1}). The limits were reflected in the data reported for OC for both years.

No specific value for the LOD for OC measurements was indicated in the 2015 data (2 cases of "<0.0"). The 2015 data contains 70 (0.3%) cases were the value reported was below the 2009 LOD. Applying the LOD of 2012 to 2015 OC data results in 987 cases in the 2015 where data are below the detection limit indicated. However, it does not seem reasonable to use this comparison as an evaluation criterion.

For $CaCO_3$ analysis results the LOD changed from < 1 g kg^{-1} in 2009 to < 0.5 g kg^{-1} in 2012. Due to the specifications of the number of decimals for the parameter (0) the change in LOD is of no effect. The lack of an identifier for analysis results below the LOD for $CaCO_3$ in 2015 is an inconsistency in data reporting. In spite of this, no values below the LOD of 2009 are reported for 2015 data. Thus, any practical consequences depend on the method used to treat the 2015 data. This evaluation equally applies when comparing the LOD data from the 2012 to the 2015 data.

3.4 Link of Soil Data to LUCAS Main Survey Point Data

For the 2009 surveys, except for Cyprus and Malta, The results from the analysis of the soil samples can be related to geographic locations and the observations of the main LUCAS survey through a table provided by Eurostat. The table contains the soil identifier and the plot identifier as part of a site record. For the 2009 soil data from Cyprus and Malta and the 2012 data from Bulgaria and Romania separate tables are provided that link the soil sample to the LUCAS plot, but that do not contain the observations of the main LUCAS survey for those years. For these surveys the soil data can be related to LUCAS plot observations by the LUCAS micro data files.

3.4.1 2009 Soil Survey in LUCAS Survey Plot Data

The link of the soil sample data with the geographic position of the LUCAS plot and observations is recorded in the file *EXT_results_2009_final.csv* (last modified: 17.01.2011), which was provided by Eurostat. After the discovery of some inconsistent entries by the JRC (Francesca Bampa, Delphine de Brogniez) the data were amended and Eurostat provided the JRC with a new version (*Ext_Results_2009(26July2011).csv*). More inconsistencies were found in the data, which led to two more files. From the information available it is not apparent, if these modifications were made by Eurostat or the JRC[11]. Therefore, the file from 26.07.2011 is used as the reference for linking the soil analysis data to the sample site.

The file in CSV format was imported into a database table with the structure of the link table. The table contains 234,709 unique entries for the field [POINT_ID] and 19,686 entries in the field [SOIL_LABEL]. The soil sample code used to indicate a LUCAS point without a soil sample is zero ("0"). The information whether a soil sample was taken at the LUCAS plot is coded in the field [SOIL_SURVEY]. The field data are defined in the file *Contents_microdata09.xls* (Eurostat, 2011) as:

- 1 = soil sample taken;
- 2 = soil sample not taken;
- 3 = point not in soil sample.

The file contains some residual inconsistencies in the definition of the relationship between soil samples and survey plots that prevent an unambiguous link of the soil sample data to the LUCAS point data:

- 873 records contain code "0" in the field [SOIL_SURVEY]
  code "0" is not defined;
- 447 records contain code "1" in field [SOIL_SURVEY], but "9" as [SOIL_LABEL]
  [SOIL_LABEL] = 9: point in soil sample but replaced;
- 429 records contain code "1" in field [SOIL_SURVELABELSURVEY];
- 491 entries in the field [SOIL_LABEL] are linked to more than one entry in the field [POINT_ID] where [SOIL_LABEL] is not "0".

The LUCAS data file contains 20,377 unique combinations of [POINT_ID] and [SOIL_LABEL], where the value for the [SOIL_LABEL] is not "0" or "9". For 9 cases a value for the [SOIL_LABEL] is combined with a value of "3" for the [SOIL_SURVEY] (point not in soil sample). However, 7 of these samples are present in the soil data (IDs: 118, 12271, 17204, 21311, 21538, 2629 and 9905).

[11] Several files in CVS, XLS and DBF format were generated that link the results of the soil analysis to LUCAS sample points. In most cases the documentation of the files is rudimentary or non-existing and the processing applied cannot be convincingly substantiated or transparently presented.

A link between the soil data and the LUCAS main survey, and thus the plot coordinates, can be established for 19,152 samples. However, in 451 cases the link contains duplicates for the soil identifier in the LUCAS file. This leaves 18,705 unique and unambiguous links of 2009 soil data with the main LUCAS data, which includes the 7 cases where the [SOIL_SURVEY] contains the entry "3". Of these, 31 have entries of "88.88888" as co-ordinates, which reduces the count to **18,674** samples that can be linked to a geographic position of a LUCAS plot remain.

The geographic co-ordinates for **99** soil samples from 2009 Malta and Cyprus were recovered from the data delivered as part of the soil survey.

3.4.2 LUCAS 2009 Malta and Cyprus Links

The locations of the sample points of the 2009 Malta survey are given in the file *Malta_Lucas.xlsx* from 30.01.2012. To provide an appraisal of the position of the survey sites to LUCAS points the soil survey locations were assigned to the nearest LUCAS Grid position. The location of the Malta 2009 sample sites with respect the nominal LUCAS Grid positions is presented in Figure 1.

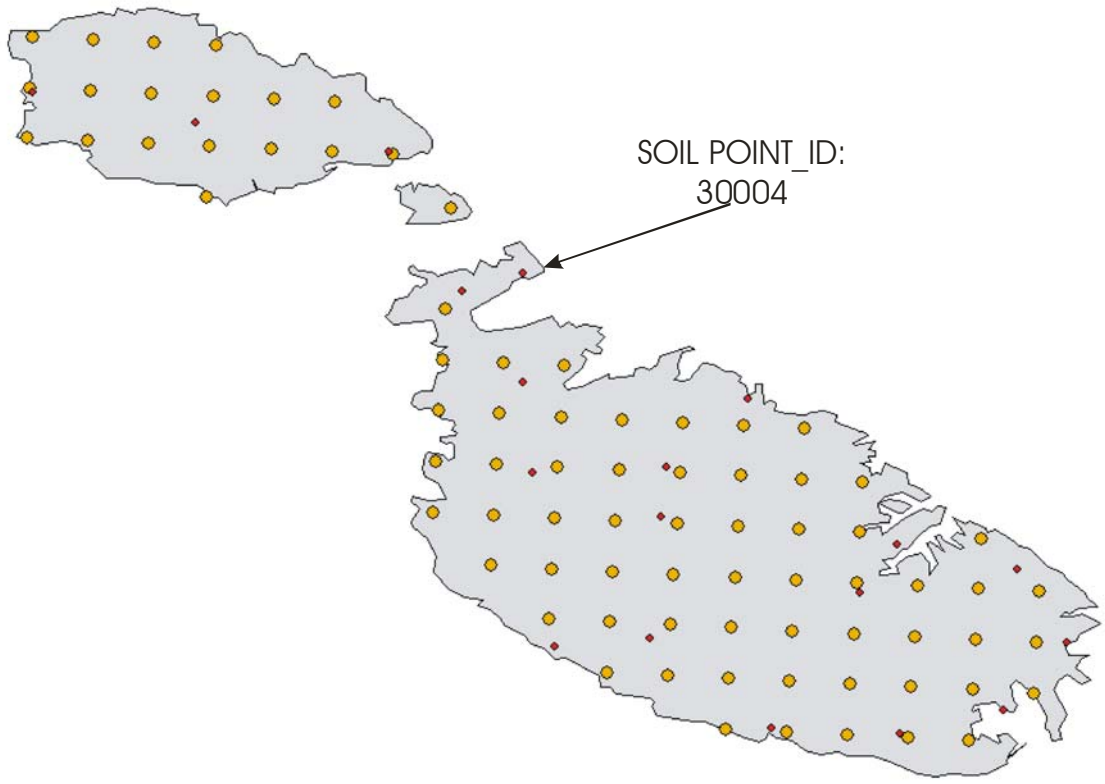


Figure 1: 2009 Malta soil and land cover/use survey sample sites and 2 x 2km LUCAS Grid positions

This procedure was successful for 18 soil sample sites. One sample site (code 30004) could not be assigned to a LUCAS Grid position. The nearest grid

position is more than 2km away and that a LUCAS Grid position is absent (point would be locates in the Sea; Eurostat, personal communication). The point identifier for the soil sample was therefore generated from the geographic location for that position. For Malta using a geographic rather than a database link can help identifying samples of the 2015 survey that were taken at or close to the 2009 sample locations.

A similar approach can be used for data from Cyprus to identify sites for comparing 2009 to 2015 soil data. However, for Cyprus no observations for land cover/use were reported in the data provided. Therefore, changes in land cover/use cannot be used to support the comparison of the results of the analysis of the soil samples.

For both data sets the approach of using geographic links was not followed in the evaluation, since these samples are not linked to the 2009 LUCAS data.

3.4.3 2012 LUCAS Soil Component Survey Geographic Co-Ordinates

For the 2012 LUCAS Soil Component survey samples for Bulgaria and Romania data of the analysis are reported for 2,034 samples. The co-ordinates of the LUCAS points are included in the soil data. For one sample (SOIL_ID: 36360) the coordinates are outside the country border. For the sample, a value of "8" is given for all GPS-related measurements and several observations in the data available to the evaluation and the Eurostat LUCAS 2012 Micro Data (Eurostat, 2011).

Duplicate entries for the soil sample identifier are present for 3 cases of samples from Bulgaria (32066, 32324 and 32437). As a result, **2,023** samples can be unambiguously linked to valid geographic co-ordinates of the 2012 LUCAS survey.

3.4.4 2015 LUCAS Soil Component Survey in LUCAS Plot Data

The submitted 2015 soil data can be linked to geographic locations and data from the LUCAS main survey in the file *LUCAS_DMT_2015.accdb*, which was provided by Eurostat. The file parameters are specified in the documentation to the LUCAS micro data (Eurostat, 2016).

The file contains 22,143 entries for identifiers of soil samples ([51-Soil Label]). Of these 172 have duplicate entries, all marked as soil taken ([50-Soil sample taken] = 1). The file contains one duplicate entry for the point identifier ([B-Point ID] = 51643122). The identifier is assigned to a point in Bulgaria and Poland, with different co-ordinates and identifiers for soil samples. Both soil samples are marked as taken. In two cases, the field [Point Longitude E/W] contains a value of "8". The corresponding samples (SOIL_ID: 12042, 60656) were excluded from link. The co-ordinates of these identifiers also contained the code for ("88.888888"), which was not present for any other linked soil sample identifier.

This brings the number of soil samples with links to 2015 LUCAS data to **21,342**.

3.4.5 Number of Repeated Soil Samples

The 2015 soil survey should take a large portion of samples at locations of the 2009 and 2012 surveys. For presumed samples of mineral soils the number of samples with repeated surveys sent for analysis by the central laboratory was reported to be 15,999. A not specifically identified number of the 1,334 samples of presumed organic soils would also come from sites previously visited.

To identify the samples from repeatedly visited sites the soil sample identifier from the analysis of the samples is linked to the corresponding field in the Eurostat file provided for 2009 and 2015. For data from 2012 the LUCAS point identifier was part of the data delivered. Additional information on LUCAS land cover/use is retrieved from the Eurostat micro data file for 2012 (Eurostat, 2011).

For the soil surveys of 2009 and 2015 a total of 14,765 samples can be linked. For the soil surveys of 2012 and 2015 the number of links is 1,421 unique links. This brings the number of analysed soil samples from repeatedly visited sites to 16,186.

To provide an appraisal of the geographic co-location of the sites between surveys the distance between the GPS co-ordinates was calculated as the only available information on location of the sample point. The calculated distance from the GPS co-ordinates was compared to the distance of the LUCAS point to the observations reported. From the two differences the minimum distance was used to provide a measure of the temporal variation in the position of a point. The findings are summarized for different ranges in Table 18.

Table 18: Minimum Distance of GPS Position to Observation Distance from Point

Minimum Difference GPS to OBS_DIST	2009 to 2015	Accum. Share	2012 to 2015	Accum. Share
	Points	*%*	*Points*	*%*
< 1 m	7485	50.6	628	43.5
1m to < 10 m	6796	96.5	802	99.1
10 m to < 25 m	338	98.8	8	99.7
25 m to < 50 m	93	99.5	3	99.9
50 m to < 100 m	46	99.8	2	100.0
>= 100m	34	100.0	0	100.0

The analysis indicates that approx. 99% of all repeated samples were taken to within 25 m of the LUCAS observations (2009 to 2015: 98.8%; 2012 to 2015: 99.7%) over 99.5% to within 50m (2009 to 2015: 99.5%; 2012 to 2015: 99.9%). Notable is a distance of over 15,000 m for 9 points, which form an isolated cluster in the data, one for a point in the United Kingdom and 8 for points in Portugal.

Treating the points with a difference >= 100m between GPS co-ordinates and observation distance from points as coming from different locations the number of samples for comparing the results from the laboratory analysis between soil surveys comes to a total of **16,152** samples (14,731 for 2009 and 2015; 1,421 for 2012 and 2015).

3.5 Uniformity Checks

The link of soil samples to the LUCAS allows an evaluation of the relation of soil properties with site conditions, such as land use, and an appraisal of temporal changes of the soil properties.

3.5.1 Soil Properties by Soil Survey

The data that can be subjected to Uniformity Checks is restricted to results from the analysis that can be compared. The restrictions applied for the evaluation were:

- restricting processing to values that are within the specified range;
- results from samples that can be linked to the LUCAS plot data;
- limiting the distance of observation from LUCAS plots.

Summary statistics for OC and all samples with the restriction applied are presented in Table 19.

Table 19: Statistical summary indicators for 2009, 2012 and 2015 LUCAS Soil Component Surveys

Organic Carbon	Statistical Indicator		Survey Year 2009	2012	2009 + 2012	2015
Within range	Count	*No.*	20871	1785	22656	22940
	Mean	*g kg^{-1}*	49.7	19.7	47.3	43.2
	Median	*g kg^{-1}*	20.7	17.5	20.1	20.7
	Std.Dev.		90.7	10.4	87.5	75.7
Link to Plot	Count	*No.*	18664	1779	20443	21282
	Mean	*g kg^{-1}*	49.3	19.8	46.7	43.7
	Median	*g kg^{-1}*	20.4	17.5	19.9	20.5
	Std.Dev.		90.8	10.4	87.2	77.4
Distance Limit	Count	*No.*	18110	1716	19826	21102
	Mean	*g kg^{-1}*	49.7	19.8	47.1	43.9
	Median	*g kg^{-1}*	20.5	17.6	19.9	20.5
	Std.Dev.		91.5	10.5	87.9	77.7

The number of values to process decreases with more stringent conditions. Of the 20,917 results reported of the OC analysis for 2009 and 2012 19,826 (94.8%) remain as data for further evaluation. For the 2015 survey the number of OC data decreases from 23,010 results to 21,102 (91.7%).

The difference in number of the results reported for analysis and the number used to provide the summary statistics is necessarily lower since results below the LOD are excluded, but need to be reported. Applying restrictions of value range, links to the LUCAS survey or distance of observation to the data have a very limited effect on the mean OC content or the standard deviation.

The relative frequency distribution of OC values for the combined 2009/2012 and the 2015 soil data is presented in Figure 2.

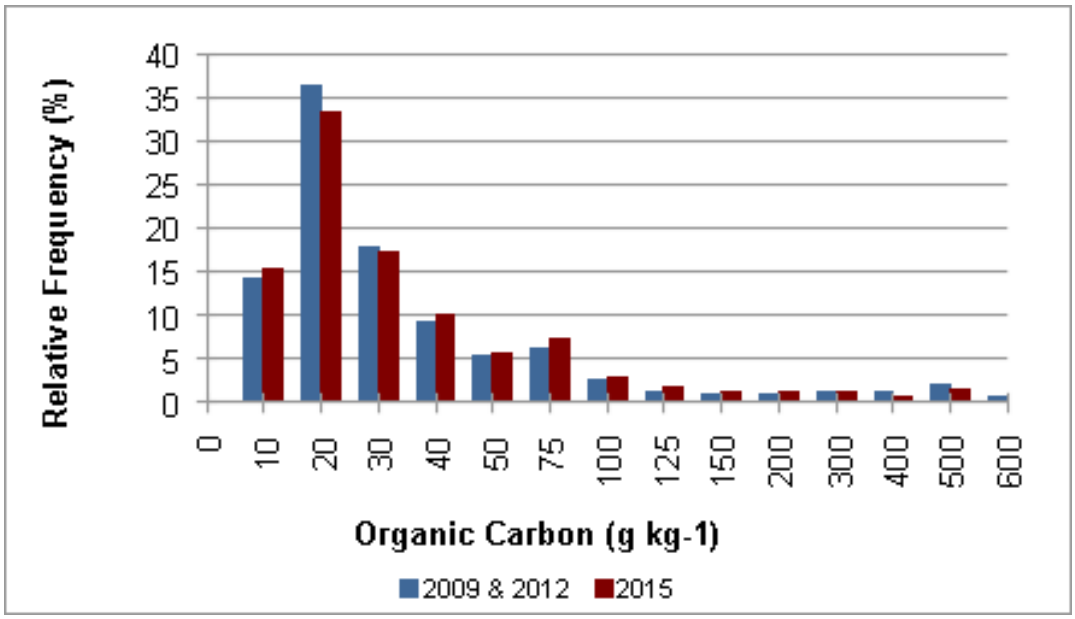


Figure 2: Frequency distribution of 2009 & 2012 and 2015 results for the analysis of Organic Carbon

The OC content results of the analysis of the soil samples has a skewed distribution. Most samples had OC contents of 20 g kg^{-1} or less, 50.5% for the 2009/2012 data and 51.3% of the 2015 soil data.

For OC one of the major factors influencing the soil property over a relatively short period is land use (LU). To provide an indication of the OC content related to LU the LUCAS land cover information was aggregated into the main land use categories of the *Land Use, Land Use Change and Forestry* (LULUCF) sector for reporting of and accounting for greenhouse gas (GHG) emissions (IPCC, 2006). A summary of OC content by LU category is presented in Table 20.

Table 20: Distribution of OC by major Land Use Category

LU*	2009		2012		2015	
	Count	Mean	Count	Mean	Count	Mean
	No.	*g kg^{-1}*	*No.*	*g kg^{-1}*	*No.*	*g kg^{-1}*
1	4040	41.3	560	22.4	4602	40.3
2	6958	18.7	797	16.7	7466	17.8
3	21	17.4			11	16.2
4	955	15.8	51	19.5	1093	16.7
6	5685	96.7	292	23.6	7240	77.5
7	75	378.5	2	28.2	46	356.7
8	40	34.0			48	31.9
9	336	24.2	14	15.5	596	17.1
Total	18110	49.7	1716	19.8	21102	43.9

* 1: Permanent Grassland; 2: Long-term cultivated; 3: Rice (paddy); 4: Permanent crops; 5: Set-aside; 6: Natural vegetation; 7: Wetlands; 8: Settlements; 9: Other areas

The summary indicators show that on average the OC content is lowest on arable land and permanent crops, and considerably higher for other land uses. The LU category "Other areas" has low OC content because it includes areas with no or low vegetation, such as bare areas and sand dunes.

The mean OC content for the 2009 and 2012 surveys samples is notably higher than the mean OC for the 2015 samples. However, it should be noted that the surveys take samples from different populations. Therefore, the difference in the mean is not an indication of change of OC content, but due to the dissimilar areas covered by samples. With the characteristics of the various surveys for deciding on sample locations and the distribution of OC content values it would not appear appropriate to perform a standard *t*-test on the mean values.

3.5.2 Changes in Soil OC content between Soil Surveys

To compare the results from the 2009/2012 survey with those from the 2015 and evaluate temporal changes the samples considered should relate to the same conditions. With the changes in the area covered between surveys, but also the changes to the selection of sample sites, only data from samples collected at the same plot location are included (repeated samples).

The frequency distribution of the differences are presented in Figure 3.

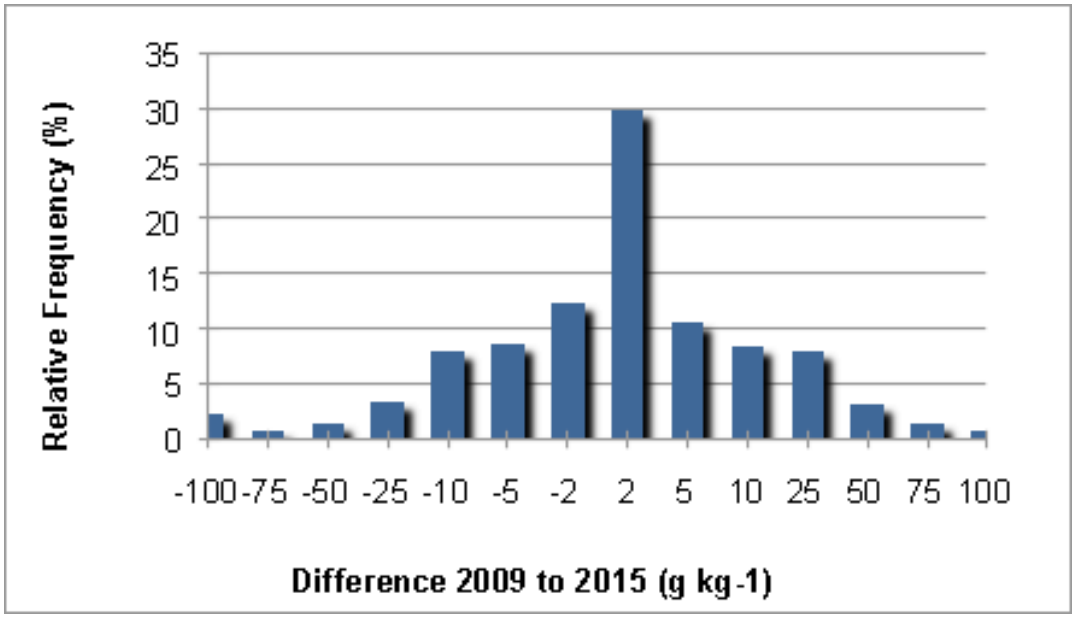


Figure 3: Frequency distribution of differences in 2009 and 2015 Organic Carbon

The distribution of the differences indicates that most values for OC of the repeated samples (29.8%) are within -2 to 2 g kg^{-1} of the value of the previous survey.

When looking at the frequency distribution of the relative changes from 2009/1012 to 2015 the situation depicted in Figure 4 emerged.

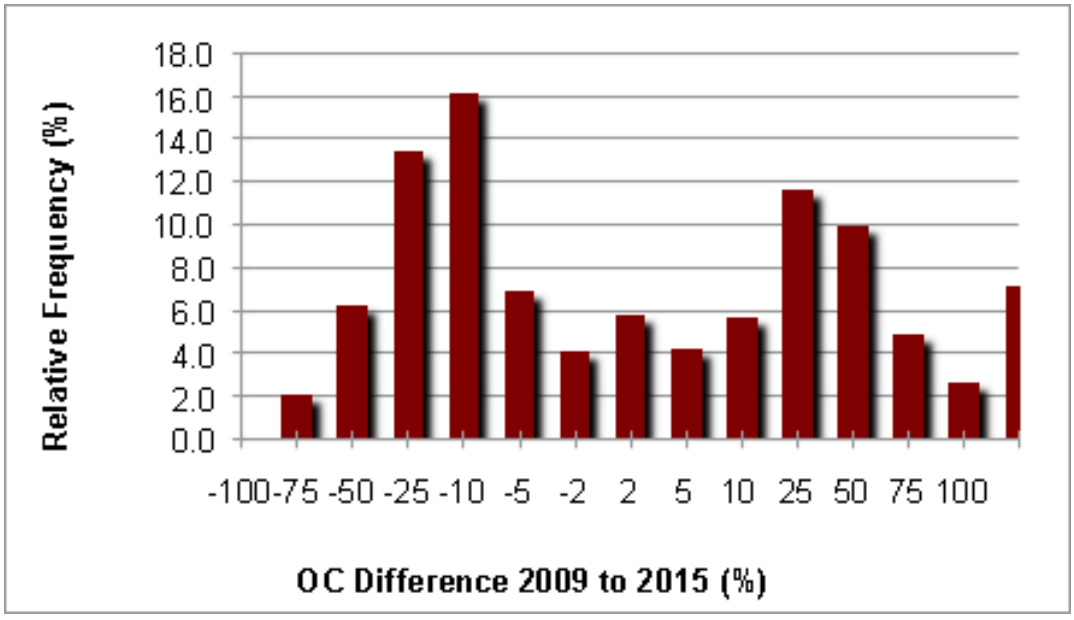


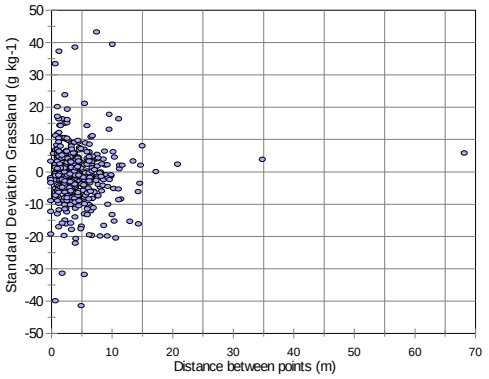
Figure 4: Frequency distribution of relative differences in 2009 and 2015 Organic Carbon

The frequency indicates that small relative difference are not reported for OC. As a rough guide and for the ensemble of repeated soil samples the relative difference needs to exceed 10% of OC content before it is registered in the results of the analysis. This is in line with the mean OC content of the samples and the LOD of the method. It also corresponds to the acceptable variation of measurements of total Carbon for ISO 10694:1995 (SGS, 2011). This is approx. the variation one may expect when analysing the same sample by different laboratories for forest soils (Cools & De Vos, 2010).

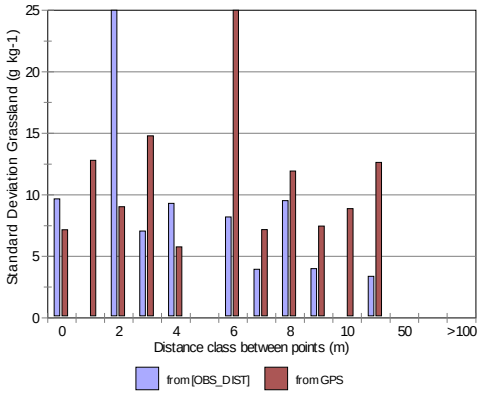
The distance between the locations at which soil samples were taken may be expected to influence the difference between OC content reported for a site.

To provide an appraisal of the effect of the distance on the difference in OC information from the plot data ([OBS_DIST]) and the GPS co-ordinates were used[12]. The data between the 2012 and 2015 surveys (Bulgaria and Romania) and for LU categories that did not change between years were used for the comparison, because the real changes in OC over two years for areas without LU change should be very small. Any variations in the OC content between years may be attributed to differences in the samples rather than changes in OC content at a sampling site. The results are presented in Figure 5.

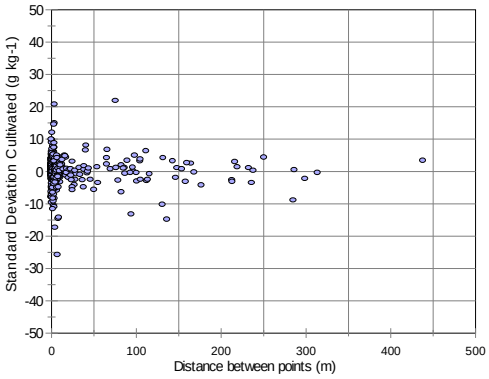
[12] [OBS_DIST]: Distance between theoretical and reached point in meters.
To estimate the distance between the geographic locations of repeated samples the difference between the distances for two years was used. For GPS locations the distance in ETRS_LAEA projected locations was used.



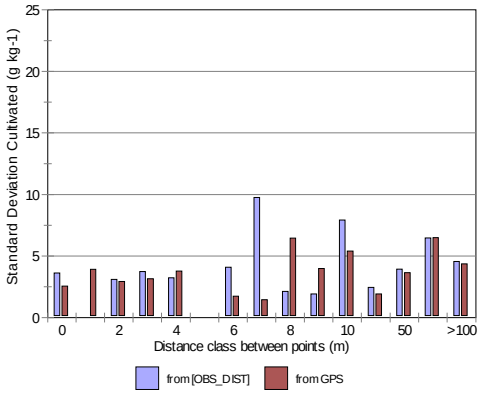
a) OC difference Grassland



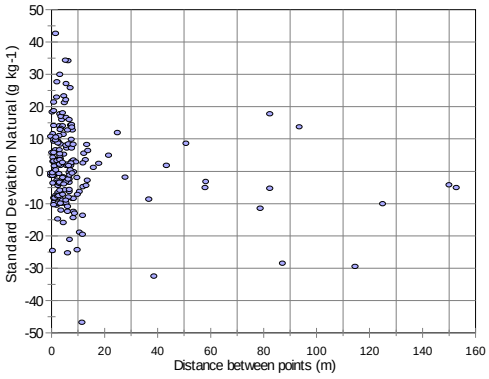
b) Std. Dev. Frequency Grassland



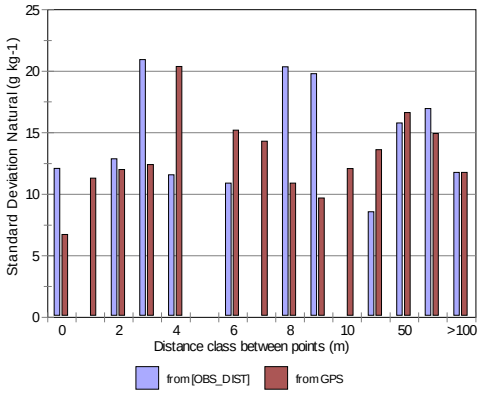
c) OC difference Cultivated



d) Std. Dev. Frequency Cultivated



e) OC difference Natural



f) Std. Dev. Frequency Natural

Figure 5: Standard deviation for OC difference and frequency distribution by distance class for selected Land Use Categories (2012 and 2015)

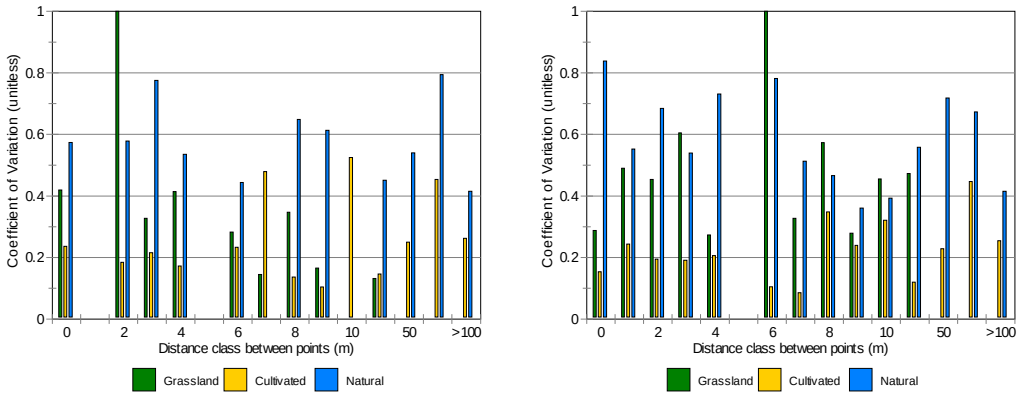
The graphs indicate that there is no increase in the standard deviation of the difference between OC content results from the analysed samples on plots without LU change. There is very notable variability in the OC content difference for grassland and semi-natural areas and, by comparison, less for samples from

cultivated land. This may indicate more homogenous conditions of soils on cultivated land, but the standard deviation is affected by the mean of the sample and this varies notably between LU categories.

A method of standardising s_{diff} between samples with different mean values is the Coefficient of Variation (CV). The CV_{diff} used for the purpose is defined as:

$$CV_{diff} = \frac{s_{diff}}{\bar{x}}$$

The CV_{diff} for OC_{diff} and distances taken from the [OBS_DIST] field and the GPS co-ordinates are presented in Figure 6.



a) [OBS_DIST] difference *b) GPS difference*

Figure 6: Coefficient of Variation for difference in OC content from plot distances derived from the field (a) [OBS_DIST] and calculated from (b) GPS co-ordinates for selected Land Use categories (2012 and 2015)

The CV_{diff} indicates that variability in the difference in OC content are smallest for samples taken from cultivated land and largest for samples from semi-natural land, with samples from grassland situated in between those categories. The data do not show an effect of distance on the variability of the difference in OC content, regardless of the LU category.

For repeated samples for 2009, 2012 and 2015 within 100m of observation the summary statistics given in Table 21 were obtained:

Table 21: Statistical summary indicators for repeated samples of 2009, 2012 and 2015 LUCAS Soil Surveys

Organic Carbon	Statistical Indicator		Survey Year 2009	2015	2012	2015
Repeated samples	Count	*No.*	14237		1197	
	Mean	*g kg^{-1}*	45.8	45.6	20.1	20.3
	Median	*g kg^{-1}*	19.7	19.5	17.7	17.5
	Std.Dev.		84.7	85.5	10.1	14.3

For OC of samples taken repeatedly the difference in the mean between surveys is 0.2 g kg^{-1} for the 2009 to 2015 and the 2012 to 2015 results. The situation of the 2012 results for OC marked with "<6" or "-999" is peculiar when compares to the OC results for 2015 for the same plots. For 169 such instances in the 2012 data the OC content was > 6 g kg^{-1} for 157 cases and < 6 g kg^{-1} for 12 cases in 2015. This amounts to 12.2% of all soil samples collected in 2012 and 2015 (samples were not included in OC change for repeated samples in Table 21 and subsequent results).

One option of appraising whether the difference in the mean between the surveys could indicate a change in OC or is within the variation of all sample means is to treat the data from the surveys as coming from two independent samples. The null-hypothesis is that the samples come from the same population and that the difference in the mean is within the range of all sample mean differences. Given the large size of the samples the *t*-value is thus:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma_{diff}}$$

Under the assumption that the standard deviation of the population difference σ_{diff} can be approximated from the variance s^2 of the two samples *t* becomes 0.19. The probability of getting a *t*-value as large as this or larger for the degrees of freedom (14,236) of the OC content variability is 0.85. Therefore, the probability of finding a difference of the means of 0.2 is rather low and the null-hypothesis cannot be rejected.

With samples taken at the same location one may argue that the soil samples taken during the 2015 survey are not independent form the samples of the previous surveys and come from paired (correlated) samples. The probability of the *t*-value of 0.46 for the difference would then be 0.65.

That the results of the repeated sample surveys are actually paired is not so obvious. While the overall change may be small, the changes at the repeated sample sites display a degree of variability for the results reported between surveys. The OC content reposted for repeated samples can be visualised by a scatter plot, as presented in Figure 7.

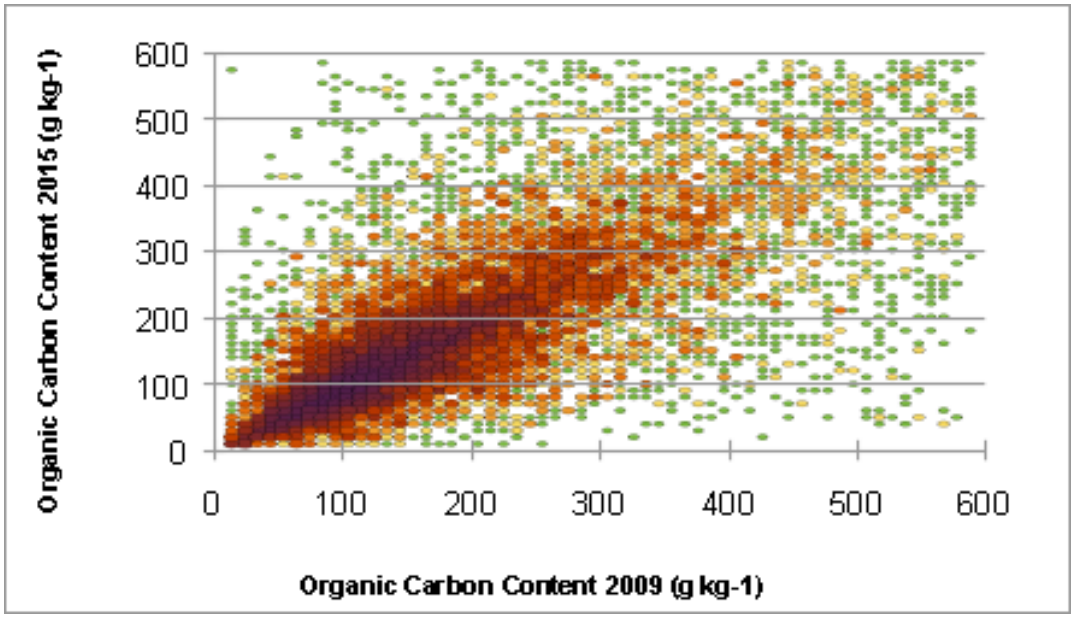


Figure 7: Scatterplot of Organic Carbon Content for 2009 vs. 2015 LUCAS Soil Component data for repeated samples

The graph shows the distribution of the value pairs with the frequency of the occurrence of a pair indicated by an increasing tone of red. The data pairs cluster along the diagonal, although there are various deviations from the diagonal.

3.5.3 Changes in Soil OC Content by Land Use Category

Whilst the overall summary did not provide evidence of change in OC there may be compensating effects from changes in land use categories.

The results of the distribution of OC across the land use categories from those samples where the category remained unchanged are presented in Table 22.

Table 22: Statistical indicators of OC by major Land Use Category for soil samples remaining in the same Land Use Category (2009 to 2015 and 2012 to 2015)

LU*	Count	2009 Mean	2015 Mean	Count	2012 Mean	2015 Mean
	No.	*g kg^{-1}*	*g kg^{-1}*	*No.*	*g kg^{-1}*	*g kg^{-1}*
1	2230	42.0	43.8	362	23.6	23.4
2	5018	17.9	17.3	476	16.6	16.8
3	5	22.8	19.2			
4	704	15.6	16.4	27	18.2	18.1
6	4167	91.7	90.4	165	23.7	25.5
7	23	432.6	458.5			
8	4	22.8	10.3			
9	80	9.1	10.3			
Total	12231	48.0	47.8	1030	20.3	20.5

* 1: Permanent Grassland; 2: Long-term cultivated; 3: Rice (paddy); 4: Permanent crops; 5: Set-aside; 6: Natural vegetation; 7: Wetlands; 8: Settlements; 9: Other areas

For repeated samples for 2009 to 2015 that remain in the same LU Category there is no significant change in the mean OC for all categories and individual categories. Some conditions shown by the statistical indicators were observed:

- Settlements
  There is a notable change in the mean of OC for the category "Settlements". This category covers all artificial areas, not only built-up areas. However, there are only 4 sites where the category did not change between surveys. Given the number of samples in the soil surveys (2009: 40; 2015: 48) this is remarkable. It would appear that either samples in areas of "Settlements" were removed from repeated visits or there was a peculiar change in the classification.
- Wetlands
  Similarly, repeated samples from "Wetlands" (23) are comparatively restricted when compared to the number of samples in the soil surveys (2009: 75; 2015: 46).

Such areas have low levels of temporal fluctuations and the presence in the sample should only change when a selection procedure is applied to the sample design.

Table 23: Testing Hypothesis for difference in mean of OC by major Land Use Category for soil samples remaining in the same Land Use Category (2009 to 2015)

LU	Diff. Observed	Std. Dev. Difference σ_{Diff}	*t*-value	P 2-tailed	Conf. Int. (95)
1:1	1.86	1.73	-1.07	0.285	± 3.44
2:2	-0.59	0.40	1.49	0.137	± 0.79
3:3	-3.64	5.43	0.67	0.522	± 12.5
4:4	0.86	0.63	-1.38	0.169	± 1.24
6:6	-1.22	2.79	0.44	0.663	± 5.56
7:7*	25.9	41.24	-0.63	0.533	± 83.1
		< 120 g kg^{-1} in surveys:			
1:1	1.28	0.59	-2.16	0.031	± 1.18
2:2	-0.44	0.21	2.11	0.035	± 0.41
4:4	1.05	0.60	-1.74	0.082	± 1.19
6:6	-0.70	0.63	1.11	0.269	± 1.26

* Contains mineral soil types

The deviations between years also concern changes from and to organic soils. These changes were evaluated using a threshold of >120 g kg^{-1} to specify mineral soils and 200 g kg^{-1} to specify organic soils. The changes found between repeated samples for 2009 and 2015 are:

- 2009 organic -> 2015 mineral: 148 (21.1%)
- 2009 mineral -> 2015 organic: 122 (18.1%)

The changes from and to organic soils concern about 20% of all samples with organic soils. This is rather unusual, because such drastic changes in OC content over a period of 6 years are highly unlikely to occur in the soil stratum. They may occur in the sample as a result of different treatment of an organic layer over the soil layer or simply by taking samples at a site that differs from the previous sample location. Combined with the low consistency of sample locations assigned to wetlands it may be hypothesised that the sampling locations varied between surveys.

The OC results of the analysis were evaluated for changes in Land Use Categories a direction of change can be assumed. The change from long-term cultivated areas to grassland or natural vegetation may be assumed to result in an increase in OC over time. Similarly, a change in land use in the opposite direction should result in a decrease in OC content. The results of the evaluation are presented in Table 24.

Table 24: Mean OC for sites with changes from and to Long-term Cultivated and Permanent Grassland / Natural areas

LU*	Count	2009 Mean	2015 Mean	Count	2012 Mean	2015 Mean
	No.	*g kg^{-1}*	*g kg^{-1}*	*No.*	*g kg^{-1}*	*g kg^{-1}*
2 -> 1, 6	416	21.2	22.5	52	17.7	17.8
1, 6 -> 2	457	27.0	25.4	50	18.4	17.4

* 1: Permanent Grassland; 2: Long-term cultivated; 6: Natural vegetation

The changes in the mean OC for the samples agree with the expected direction, but are too small to substantiate a tangible change. If one assumes a relative change in soil OC from after land use change from grassland or natural vegetation to long-term cultivated land use of 25%[13] a soil with an OC content of 27.0 g kg^{-1} would stabilise after 20 years at about 20.3 g kg^{-1}. Extrapolating the observed trend over 20 years the average soil OC would amount to 21.7 g kg^{-1}.

For the changes in OC content reported for the soil samples of the survey from 2012 to the same location in 2015 the summary statistics are provided in Table 22.

For 86% of the sites of 2012 with valid repeated samples for OC content no changes in LU category were observed. Notable is the absence of samples from areas of settlements in 2012. All sites with other land uses in 2012 have changes LU category in 2015. Changes in OC content are small for all samples.

The results of the analysis of the probability of rejecting the hypothesis that there is no change in OC content between the surveys is presented in Table 25.

[13] The relative change of 25% is estimated from "Table 5.5: Relative stock change factors (Flu, Fmg, and Fi) (over 20 years) for different management activities on Cropland" (IPCC, 2006), as the mean for Temperate dry and Temperate moist regimes.

Table 25: Testing Hypothesis for difference in Mean of OC by major Land Use Category for soil samples remaining in the same Land Use Category (2012 to 2015)

LU	Diff. Observed	Std. Dev. Difference σ_{Diff}	*t*-value	P 2-tailed	Conf. Int. (95)
1:1	-0.29	0.72	0.23	0.82	± 2.56
2:2	0.20	0.80	-0.67	0.50	± 0.60
3:3*					±
4:4	-0.06	0.13	0.03	0.98	± 4.40
6:6	1.81	13.7	-1.12	0.26	± 3.21
7:7*					
8:8*					
8:9*					

* no data for analysis

There is no evidence that the changes observed in OC content sufficiently substantial to reject the null-hypothesis that there was no change on the sites where the LU category remained the same in 2012 and 2015.

To deduct from these results that the repeated soil survey does not contain sufficient evidence of a change in soil OC is only applicable when looking at the ensemble of the repeated samples. The changes from all samples contain compensations for losses and gains in OC that may occur at a thematically or spatially delimited level. However, such investigations are outside the scope of this work.

3.5.4 Changes in $CaCO_3$ Content

Changes in $CaCO_3$ influence the results reported for the analysis of OC content, because of laboratory method used to establish OC. Of the total number of 14,770 samples that can be linked for the 2009 and 2015 surveys for $CaCO_3$ 5.536 (37.5%) are assigned analysis results below the LOD (< 1 g kg^{-1}) in the 2009 data and 7,911 (53.6%) in the corresponding 2015 samples. For 5,215 links the value in 2009 was the code used to identify a value below the LOD and a value of "0" or "1" in the 2015 data. For 8.8% these cases the $CaCO_3$ reported for 2009 was > 10 g kg^{-1}.

There is some variability in the results of the sample analysis, but this is very much restricted to values close to the LOD for $CaCO_3$.

Table 26: Statistical indicators of $CaCO_3$ by major Land Use Category for soil samples remaining in the same Land Use Category (2009 to 2015 and 2012 to 2015)

LU*	Count	2009 Mean	2015 Mean	Count	2012 Mean	2015 Mean
	No.	*g kg^{-1}*	*g kg^{-1}*	*No.*	*g kg^{-1}*	*g kg^{-1}*
1	473	116.5	107.6	108	57.5	66.1
2	2196	171.4	167.9	154	44.1	50.5
3						
4	562	262.1	254.5	8	97.4	63.9
6	313	158.7	137.3	20	50.8	80.7
7	1	189.0	108.0			
8						
9	59	300.0	283.9			
Total	4343	179.4	172.8	290	50.0	58.7

* 1: Permanent Grassland; 2: Long-term cultivated; 3: Rice (paddy); 4: Permanent crops; 5: Set-aside; 6: Natural vegetation; 7: Wetlands; 8: Settlements; 9: Other areas

Overall there has been a reduction in $CaCO_3$ from 2009 to 2015 and an increase for the repeated samples from 2012 to 2015. None of the changes are sufficiently conclusive to reject the hypothesis that there was no change in $CaCO_3$ from previous surveys to 2015.

There is no reliable relationship between land use and changes in $CaCO_3$ and the evaluation uses the entirety of $CaCO_3$ data reported for the surveys. An evaluation of results of the laboratory analysis reported for $CaCO_3$ by changes to and from samples collected on cultivated land was not performed.

The total soil carbon content was estimated from the sum of the OC and $CaCO_3$ analysis results. A summary of the distribution of total soil carbon by major LU category for 2009/2012 and 2015 repeated samples is given in Table 27.

Table 27: Statistical indicators of total soil carbon by major Land Use Category for soil samples remaining in the same Land Use Category (2009 to 2015)

LU*	Count	2009 Mean	2015 Mean	t-value	P 2-tailed	Conf. Int. (95)
	No.	*g kg^{-1}*	*g kg^{-1}*			*g kg^{-1}*
1	2258	68.9	66.2	0.91	0.36	± 6.02
2	5396	91.9	89.8	0.76	0.45	± 5.63
3	8**	20.1	20.6			±
4	723	219.4	214.4	0.47	0.64	± 21.0
6	4206	104.7	102.9	0.58	0.56	± 6.02
7	26	433.8	429.1	0.12	0.42	± 77.9
8	4**	23	22.75			
9	82	225.1	231.3	0.33	0.74	± 70.5
Total						

* 1: Permanent Grassland; 2: Long-term cultivated; 3: Rice (paddy); 4: Permanent crops; 5: Set-aside; 6: Natural vegetation; 7: Wetlands; 8: Settlements; 9: Other areas
** Sample size too low.

To compute the statistics in the table a value of zero (0) was used for analysis results with an OC content or $CaCO_3$ concentration below the LOD. Relative to the variability of the results of the analysis the changes in total soil carbon between surveys would not provide sufficient support to reject the hypothesis of no change between surveys.

Also for total soil carbon the changes following a change in LU category from and to cultivated was assessed. The results are presented in Table 28.

Table 28: Mean total carbon for sites with changes from and to Long-term Cultivated and Permanent Grassland / Natural areas (2009 to 2015 and 2012 to 2015)

LU*	Count	2009 Mean	2015 Mean	Count	2012** Mean	2015** Mean
	No.	*g kg^{-1}*	*g kg^{-1}*	*No.*	*g kg^{-1}*	*g kg^{-1}*
2 -> 1, 6	428	80.5	77.3	53	27.6	27.9
1, 6 -> 2	471	72.8	70.1	53	27.7	27.6

* 1: Permanent Grassland; 2: Long-term cultivated; 6: Natural vegetation
** for total C > 2 g kg^{-1}.

The expected trend for OC content following a change from cultivated land to grassland or semi-natural areas is an increase in OC content and should, since $CaCO_3$ should not be significantly affected by the change in LU category, also lead to an increase in total soil carbon. This general development was not found in the data. Rather, total soil carbon content decreased after a change from cultivated to grassland or semi-natural land use. Given that the OC content

increased after such a change in LU category this peculiarity may be attributed to a decrease in $CaCO_3$. The probability that there is no change in total soil carbon is 0.73 and the difference may be attributed to be within the variability of the means of differences.

Table 29: Summary of results of Uniformity Checks for multi-year single-parameter parameters

Check ID	Result	Survey Change	Check Message
45	62.6%	2009 and 2015	Change of OC not within acceptable variation (>15 +/- %).*
46	24.6%	2009 and 2015	Change of $CaCO_3$ not within acceptable variation (>15 +/- %).*
47	61.6%	2009 and 2015	Change of Total Carbon not within acceptable variation (>15 +/- %).*
48	-0.2 g kg^{-1}	2009 and 2015	Change in OC content without change in land cover/use.
49	38.1%	2009 and 2015	No change in OC content after change in land cover/use (<15 +/- %).*
45	51.2%	2012 and 2015	Change of OC not within acceptable variation (15+/- %).*
46	68.3%	2012 and 2015	Change of $CaCO_3$ not within acceptable variation (15 +/- %).*
47	59.9%	2012 and 2015	Change of Total Carbon not within acceptable variation (+/- %)
48	-0.2 g kg^{-1}	2012 and 2015	Change in OC content without change in land cover/use.
49	55.1%	2012 and 2015	No change in OC content after change in land cover/use.
50	21.1%	2009 to 2015	Change from organic to mineral soil type.
	N/A	2012 to 2015	Change from organic to mineral soil type.
51	18.1%	2009 to 2015	Change from mineral to organic soil type.
	N/A	2012 to 2015	Change from mineral to organic soil type.

* Excluded are cases with a value zero (0) in either year.

4 Summary and Conclusions

The evaluation if the data from the LUCAS soil component surveys from 2009, 2012 and 2015 concentrated on OC. As a consequence of the laboratory method used to report on the OC content $CaCO_3$ concentrations were included in the evaluation. With the availability of data from the 2015 survey changes in soil properties over the samples collected in 2009/ 2012 could be calculated.

- **2009 Soil Component Survey**

  For 2009 results of the laboratory analysis were submitted for a total of 20,916 soil samples. Of these, 19 were from separate soil surveys carried out in Malta and 90 in Cyprus. For the OC content of 46 samples (0.2%) a value below the limit of detection of the method used was indicated.

  A link to the data of the LUCAS main survey could be established for 18,674 soil samples (89.7%, excluding Malta and Cyprus). There number has been affected by the use of alpha-numeric codes for the identifier in the soil data, which do not match the data provided by Eurostat.

- **2012 Soil Component Survey**

  For LUCAS Soil Component surveys of 2012 surveys in Bulgaria and Romania results of the analysis of 2,027 samples were received, of which 2,023 had unique identifiers. The detection limit for OC content seems to have been 6 g kg^{-1} for the analysis instead of the previously 2 g kg^{-1}. The LOD for $CaCO_3$ was 0.5 g kg^{-1} instead of 1 g kg^{-1}. In addition, in a deviation from other data a code of "-999" was used, which is presumed to indicate missing data from the analysis. With 245 cases the proportion of samples with OC below LOD or missing data is comparatively high (12.1%).

- **2015 Soil Component Survey**

  The results of the analysis of the soil 2015 Soil Component survey were submitted on a monthly basis with some corrections. Preliminary results of the laboratory analysis for submitted for a total of 23,010 cases[14]. The structure of the parameters differs to some degree from data submitted for previous periods. In a deviation from previous surveys conditions of measurements below the LOD were reported with a value zero (0) in the data for $CaCO_3$. The submitted files also contain values below the LOD used for previous surveys. A link to previous surveys could be established for 14,768 samples for 2009 and for 1,421 samples for 2012.

- **Changes between Surveys**

  For comparing data between surveys and assessing the degree of change of soil property the distance between the locations were samples were collected could be of influence. The distance between sample locations of repeated points is not immediately available. The information related to distance in the LUCAS data concerns the distance between the location reached and the theoretical position of the LUCAS grid point. However, the variability in the

[14] Cases correspond to records in the data, not samples.

difference of OC content between surveys is largest when the difference in locations in both surveys approaches zero. That the distance between repeated soil sampling locations may be an issue is supported by the variability of the difference in OC content between surveys. There appears to be a background variability in the data which varies from 10% to 50%, depending on land use. These levels of "noise" in the data are estimated when removing inexplicable conversions between distinctly mineral and organic soils. While the number of samples with OC content > 200 g kg^{-1} decreased by 4% from the repeated samples in 2009 to 2015, 21.1% changed from organic to mineral (< 120 g kg^{-1}) and 18.1% from mineral to organic.

The overall change in OC content from repeated samples from 2009 and 2012 to 2015 was 0.2% (2009 to 2015: 45.8 to 45.6 g kg^{-1}; 2012 to 2015: 20.1 to 20.3 g kg^{-1}). In both cases the data did not provide sufficient support to reject the hypothesis that the samples were taken form a population without change in OC content.

An evaluation based on major land use categories indicated that there may have been a tangible change in OC content on soil samples collected on cultivated land and grassland from 2009 to 2015. The results suggest that although there is not sufficient evidence of an overall change in OC between surveys there may well be tangible changes within and between land use categories.

Conclusions:

- The use of numeric codes to indicate specific conditions in fields containing numeric measurements should be avoided. The use of codes for values below the detection limit of the method or instrument used in the laboratory should be more clearly specified and documented. The coding applied should also be more consistent between surveys. Using a value zero (0) to indicate data below the LOD is not a viable method.
- The identifiers for the soil samples in the results of the laboratory analysis reference to alpha-numeric codes should be numeric and provide unambiguous links to the LUCAS main survey observations.
- Conditions when the results of the analysis are below the limit of detection should be coded clearly and consistently and not indicated by a value of zero (0).
- The format of the data submitted should be changed to avoid the need for an order for the parameters reported and a code for missing data.
- In the analysis results from samples of mineral soil should be treated separately from samples of organic soils.
- For an estimation of changes in OC content between surveys information on land use should be included. In the analysis a separation of areas remaining in a land use category and those changing land use may be advantageous. Particular care seems to be required when areas are broadly classified as settlements or wetlands.

Annex

LUCAS Data Sources

ID	Survey Type	Survey Data	Data Type	Date	Comment
	LUCAS Land cover/use	Eurostat Web site			Eurostat LUCAS Web site According to the text the following data should be available from the site: • micro-data • photos • statistical tables and soil data The site "Primary Data" lists **soil data** as available from the site. The site provides a link to ESDAC for the 2009 LUCAS Soil data.
1		Reference data 2009	point	31.03.2011	EU-23, 234707 records, no fields related to soil (LUCA_2009_SH.zip) No reference for other years on the Web-site.
2		LUCAS micro data 2009	CSV	25.11.2016	EU-23 cover (EU23_2009_20161125.csv)
3		LUCAS micro data 2012	CSV	19.10.2016	EU-28 cover (EU-2012-20161019.csv)
4		LUCAS micro data 2015	CSV	28.10.2016	EU-28 cover (EU28_2015_20161028.csv)
5		GISCO.LUCA_PT_2009	point		234709 records, includes LUCA_SOIL_LABL
6		GISCO.LUCA_PT_2012	point		270276 records, includes LUCA_SOIL_LABL
7		GISCO.LUCA_PT_2015	point		348536 records, includes LUCA_SOIL_LABL
	LUCAS Soil	ESDAC Web site			Information and download of "LUCAS 2009 TOPSOIL data".
8		Soil data 2009	XLS	05.092013	(LUCAS_TOPSOIL_v1.xls)
		JRC internal			Server or shared
9		Survey: LUCAS Soil 2009	CSV	14.01.2011	Data for ESTAT (Folder) This is preliminary data with several stages of modifications (see readme - LUCAS2009_estat.docx)
10		MOSES: 2009 survey results	XLS	14.06.2011	Final results of the contract no. 385355 to SGS to analyze the LUCAS soil samples. Report, (5.6.LUCAS_Results of 20.000 soil samples.xls)
11		MOSES: 2009 Wageningen	XLS	14.06.2011	Analysis of lab samples (5.5.LUCAS_results of Wageningen tests in 2010-2011.xls)

ID	Survey Type	Survey Data	Data Type	Date	Comment
12		MOSES: 2009 extra samples	XLS	05.10.2011	Extra samples (LUCAS_RECORD_2011-09-27_Results_897_extra samples.xls)
13		Shared: 2009/12 re-analysis	XLS	18.03.2018	Re-analysis of 2009/12 lab data (Re_analysis_LUCAS2009_2012.xlsx)
14		Local: Malta	DBF, XLS	08.11.2011	2009 survey, only plot data (malta_lucas.*)
15		?: Cyprus	?	?	No original survey data found.
16		Shared: 2009 survey	XLS	11.05.2017	All 2009 data (LUCAS_2009.xlsx)
17		MOSES: 2009/12 survey results	SHP	17.06.2016	Point data (LUCAS_2009_plusBulgariaRomania), includes BG, MT and RO
18		MOSES: 2009/12 survey results	SHP	15.07.2016	Point data (LUCAS_LAEA_ALL), includes BG, CY, MT and RO
19		LUCAS Topsoil survey 2012	XLS, CSV	23.07.2015	Bulgaria, separate data file (Bulgaria.xlsx), CSV generated later
20			XLS, CSV	23.07.2015	Romania, separate data file (Romania.xlsx), CSV generated later
21			XLS	12.03.2018	Bulgaria, separate data file (LUCAS_2012_BG.xlsx)
22			XLS	12.03.2018	Romania, separate data file (LUCAS_2012_RO.xlsx)
		LUCAS 2015			
23		JRC internal	XLS	12.03.2018	Repeated sample sites (LUCAS2015_2009_repeated_points_mineral_soils.xlsx)
24			XLS	12.03.2018	Repeated sample sites (LUCAS2015_2009_repeated_points_organic_soils.xlsx)
25			XLS	12.03.2018	Repeated sample sites for BG and RO (LUCAS2015_2012_repeated_points_BG_RO.xlsx)
26		MOSES: 2009/12 + HM data	XLS	14.03.2018	Consolidated data with heavy metal data (All_Reference_Working.xlsx)
27		MOSES: 2009/12 point	SHP	14.03.2018	Points of consolidated data with heavy metal data (ConsolidatedFile.shp)

1 http://ec.europa.eu/eurostat/cache/GISCO/geodatafiles/LUCA_2009_SH.zip
2 http://ec.europa.eu/eurostat/documents/205002/208938/EU23_2009_20161125.csv
3 http://ec.europa.eu/eurostat/cache/lucas/EU-2012-20161019.csv
4 http://ec.europa.eu/eurostat/documents/205002/6786255/EU28_2015_20161028.csv
5 GISCO SDE
6 GISCO SDE

7 GISCO SDE
8 file:////ies.jrc.it/H05/SOIL/ESDAC/MOSES%20data/LUCAS-2009-data-as-distributed-by-ESDAC/LUCAS_TOPSOIL_v1.xlsx
9 file:////ies.jrc.it/H05/lucas2009_estat/readme%20-%20LUCAS2009_estat.docx
10 file:////ies.jrc.it/H05/SOIL/ESDAC/MOSES%20data/LUCAS_project/LUCAS_SOIL_2009/DATA/Final_Report/5.6.LUCAS_Results%20of%2020,000%20soil%20samples.xls
11 file:////ies.jrc.it/H05/SOIL/ESDAC/MOSES%20data/LUCAS_project/LUCAS_SOIL_2009/DATA/Final_Report/5.5.LUCAS_results%20of%20Wageningen%20tests%20in%202010-2011.xls
12 file:////ies.jrc.it/H05/SOIL/ESDAC/MOSES%20data/LUCAS_project/LUCAS_SOIL_2009/DATA/LUCAS_RECORD_2011-09-27_Results_897_extra%20samples.xls
13 \\ies.jrc.it\H05\SOIL\share\OihaneFernandezUgalde\LUCAS_Topsoil_Survey\Quality_control_lab_analyses\Re_analysis_LUCAS2009_2012.xlsx
14 N/A
15 N/A
16 file:////ies.jrc.it/H05/SOIL/share/OihaneFernandezUgalde/LUCAS_Topsoil_Survey/LUCAS_2009/LUCAS_2009.xlsx
17 file:////ies.jrc.it/H05/SOIL/ESDAC/MOSES%20data/LUCAS_2009_plusBulgariaRomania-shapefile/
18 file:////ies.jrc.it/H05/SOIL/ESDAC/MOSES%20data/LUCAS_2009_plusBulgariaRomaniaCyprus-shapefile/
19 file:////ies.jrc.it/H05/SOIL/ESDAC/MOSES%20data/LUCAS_Romania_Bulgaria/Bulgaria.xlsx
20 file:////ies.jrc.it/H05/SOIL/ESDAC/MOSES%20data/LUCAS_Romania_Bulgaria/Romania.xlsx
21 file:////ies.jrc.it/H05/SOIL/share/OihaneFernandezUgalde/LUCAS_Topsoil_Survey/LUCAS_2012/LUCAS_2012_BG.xlsx
22 file:////ies.jrc.it/H05/SOIL/share/OihaneFernandezUgalde/LUCAS_Topsoil_Survey/LUCAS_2012/LUCAS_2012_RO.xlsx
23 file:////ies.jrc.it/H05/SOIL/share/OihaneFernandezUgalde/LUCAS_Topsoil_Survey/LUCAS_2015_repeated_points/LUCAS2015_2009_repeated_points_mineral_soils.xlsx
24 file:////ies.jrc.it/H05/SOIL/share/OihaneFernandezUgalde/LUCAS_Topsoil_Survey/LUCAS_2015_repeated_points/LUCAS2015_2009_repeated_points_organic_soils.xlsx
25 file:////ies.jrc.it/H05/SOIL/share/OihaneFernandezUgalde/LUCAS_Topsoil_Survey/LUCAS_2015_repeated_points/LUCAS2015_2012_repeated_points_BG_RO.xlsx
26 file:////ies-ud01.jrc.it/soil/ESDAC/MOSES%20data/LUCAS-2009-2012-consolidated+ESTAT+HM/All_Reference_Working.xlsx
27 file:////ies-ud01.jrc.it/soil/ESDAC/MOSES%20data/LUCAS-2009-2012-consolidated+ESTAT+HM/

References

Cools, N. & De Vos, B., 2010. *6th FSCC Interlaboratory Comparison,* Brussels: Instituut voor Natuur- en Bosonderzoek (INBO), Brussel.

Eurostat, 2011. *LUCAS Micro Data 2009.* [En ligne] Available at: http://ec.europa.eu/eurostat/documents/205002/208938/Contents-LUCAS-primary-data-2009-20140618-0.xls/27fe0910-4150-4299-9e89-8cc0adf73d83
[Accès le 2 May 2018].

Eurostat, 2011. *LUCAS Micro Data 2012.* [En ligne] Available at: http://ec.europa.eu/eurostat/documents/205002/208012/Contents-LUCAS-primary-data-12-20140618-.xls/811dfa11-89f7-428c-9459-ff56baa07a9f
[Accès le 02 May 2018].

Eurostat, 2016. *LUCAS SURVEY 2015 WEB CSV Record Descriptor,* Luxembourg: European Commission.

Hiederer, R. et al., 2008. *Forest Focus Monitoring Database System – Validation Methodology,* Luxembourg: Office for Official Publications of the European Communities.

IFN-INRA, 2009. *List of the checks performed for Biosoil data validation,* s.l.: Inventaire forestier national (IFN) and Institut national de la recherche agronomique (INRA).

IPCC, 2006. *2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 4, Agriculture, forestry and other Land Use.* Hayama: IEGS.

JRC, 2009. *Service Provision for technical support in the LUCAS study, provision of central laboratory services for soil analysis.* Ispra(Varese): European Commission Joint ResearchCentre.

JRC, 2016. *Tender Specifications: Laboratory analysis of soil samples collected in the context of the LUCAS 2015 survey.* Ispra(Varese): European Commission Joint Research Centre.

Orgiazzi, A. et al., 2017. LUCAS Soil, the largest expandable soil dataset for Europe: a review. *European Journal of Soil Science*, 11 November, 69(1), pp. 140-153.

SGS, 2010. *Kick-off Meeting Minutes LUCAS project: No 2009/S 122-176960,* Ispra, Italy: JRC.

SGS, 2010. *LUCAS Project: Final Test Plan,* Budapest, Hungary: SGS Hungary, Member of SGS Group (Société Générale de Surveillance.

SGS, 2010. *LUCAS Project: Preliminary Test Plan,* Ispra, Italy: SGS Hungary, Member of Société Générale de Surveillance.

SGS, 2011. *Final technical report and executive summary Lucas soil study,* Budapest, hungary: SGS Hungary Ltd..

SGS, 2016. *LUCAS Project: Preliminary Test Plan 2016,* Nyiregyháza: SGS Hungary LTD. Laboratory.

SGS, 2016. *Template Specimen,* Nyiregyháza: SGS Hungary LTD. Laboratory.

List of Acronyms

Acronym	Label
ASCII	American Standard Code for Information Interchange
$CaCO_3$	Calcium carbonate
CSV	Comma separated values file
EU23	European Union, 23 states participating in survey
EU28	European Union of 28 Member States
GIS	Geographic Information System
LOD	Limit of determination
IPCC	Intergovernmental Panel on Climate Change
LU	Land use
LULUCF	Land use, land use change and forestry sector
NUTS	Nomenclature des Unités territoriales statistiques
OC	Organic carbon
RDB	Relational data base
RDBMS	Relational data base management system
SOC	Soil organic carbon
TOTC	Total carbon
XLS	Microsoft® Excel 97 -2003 Workbook
XLSX	Microsoft® Office Open XML SpreadsheetML File Format

List of Figures

List of Tables

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