GEMAS – Extended dataset: Data description

1. Introduction

The "GEMAS - Extended Dataset" is available as a series of Excel workbooks (*.xlsx) or as a series of CSV files. It includes the data of 2213 Ap samples and 2130 Gr samples and thus also the samples collected in the area of eastern Ukraine. It consists of the following files and worksheets:

(a) for the Ap samples

XLSX-File	Worksheet	CSV-File	Content
11 GEMAS Ap - Ident	GEMAS Ap – Ident	11 GEMAS Ap - Ident	Identifier, location reference, information on the sample material, further information on the sample sites [1]
12 GEMAS Ap - AR	Ap AR PDL	12a GEMAS Ap - AR PDL	Analytical data: aqua regia extraction (AR), ICP-AES/ICP-MS, practical detection limit (PDL) [2]
	Ap AR LDL ACME	12b GEMAS Ap - AR LDL ACME	Analytical data: aqua regia extraction (AR), ICP-AES/ICP-MS, “official” detection limit of ACME (LDL) [2]
	DL's	12c GEMAS Ap - AR DL's	Detection limits for aqua regia extraction (AR), ICP-AES/ICP-MS [2]
13 GEMAS Ap - XRF	Ap WD-XRF	13a GEMAS Ap - WD-XRF	Analytical data: total content, WD-XRF [3]
	Ap WD-XRF Oxides	13b GEMAS Ap - WD-XRF Oxides	Analytical data: total content, WD-XRF, oxides of main components (SiO_2, TiO_2, …; L.O.I.) [3]
	Ap ED-XRF	13c GEMAS Ap - ED-XRF	Analytical data: total content, ED-XRF [3]
	DL's WD-XRF	13d GEMAS Ap - WD-XRF DL's	Detection limits of WD-XRF [3]
	DL's ED-XRF	13e GEMAS Ap - ED-XRF DL's	Detection limits of ED-XRF [3]
14 GEMAS Ap - MMI	Ap MMI	14a GEMAS Ap - MMI	Analytical data: MMI® extraction, ICP-MS [4]
	DL's	14b GEMAS Ap - MMI DL's	Detection limits for MMI® extraction, ICP-MS [4]
15 GEMAS Ap - Other data	Ap pH-CEC-C-N-S-χ-CIA	15a GEMAS Ap - pH-CEC-C-N-S-χ-CIA	Supplementary analytical data and indices: pH, CEC, C, N, S, χ, CIA [5]
	Ap Pb isotope ratios	15b GEMAS Ap - Pb isotope ratios	Pb isotope ratios [5]
	Ap PSD	15c GEMAS Ap - PSD	PSD (particle size distribution) [5]
	Ap Kd values	15d GEMAS Ap - Kd values	Kd values for selected elements [5]
	Ap PNEC & RCR	15e GEMAS Ap - PNEC & RCR	PNEC and RCR for selected elements [5]
	DL's	15f GEMAS Ap - Other data DL's	Detection limits for CEC, C, N, S [5]

[1] see chapter 2.2, [2] see chapter 2.3, [3] see chapter 2.4, [4] see chapter 2.5, [5] see chapter 2.6.

(b) for the Gr samples

XLSX-File	Worksheet	CSV-File	Content
21 GEMAS Gr - Ident	GEMAS Gr – Ident	21 GEMAS Gr - Ident	Identifier, location reference, information on the sample material, further information on the sample sites [1]
12 GEMAS Gr - AR	Gr AR PDL	22a GEMAS Gr - AR PDL	Analytical data: aqua regia extraction (AR), ICP-AES/ICP-MS, practical detection limit (PDL) [2]
	Gr AR LDL ACME	22b GEMAS Gr - AR LDL ACME	Analytical data: aqua regia extraction (AR), ICP-AES/ICP-MS, “official” detection limit of ACME (LDL) [2]
	DL's	22c GEMAS Gr - AR DL's	Detection limits for aqua regia extraction (AR), ICP-AES/ICP-MS [2]
23 GEMAS Gr - XRF	Gr WD-XRF	23a GEMAS Gr - WD-XRF	Analytical data: total content, WD-XRF [3]
	Gr WD-XRF Oxides	23b GEMAS Gr - WD-XRF Oxides	Analytical data: total content, WD-XRF, oxides of main components (SiO_2, TiO_2, …; L.O.I.) [3]
	DL's WD-XRF	23c GEMAS Gr - WD-XRF DL's	Detection limits of WD-XRF [3]
25 GEMAS Gr - Other data	Gr pH-CEC-C-N-S-χ-CIA	25a GEMAS Gr - pH-CEC-C-N-S-χ-CIA	Supplementary analytical data and indices: pH, CEC, C, S, CIA [4]
	Gr Pb isotope ratios	25b GEMAS Gr - Pb isotope ratios	Pb isotope ratios [4]
	Gr PSD	25c GEMAS Gr - PSD	PSD (particle size Distribution) [4]
	Gr Kd values	25d GEMAS Gr - Kd values	Kd values for selected elements [4]
	Gr PNEC & RCR	25e GEMAS Gr - PNEC & RCR	PNEC and RCR for selected elements [4]
	DL's	25f GEMAS Gr - Other data DL's	Detection limits for CEC, C, S [4]

[1] see chapter 2.2, [2] see chapter 2.3, [3] see chapter 2.4, [4] see chapter 2.6.

2. Description of the file structures and file contents

2.1. General notes

The structures and contents of the files for the Ap and Gr samples are identical. Therefore, they are described together. Not all parameters that were analysed or determined for the Ap samples were also analysed or determined for the Gr samples. Consequently, some worksheets respectively files or parameters are omitted for the Gr samples.

2.2. GEMAS – Ident

2.2.1. General information

The XLSX files "11 GEMAS Ap - Ident" and "21 GEMAS Gr - Ident" respectively the corresponding CSV files (cf. tables in chapter 1) contain the identifiers of the samples of the GEMAS project, location information on the sampling points (coordinates), information on the sample material and further information on the sampling sites.

2.2.2. Column “ID”

Each sample of the GEMAS project is uniquely identifiable via its ID. It was assigned at randomisation of the samples (see chapter 6.3.1 in Demetriades et al., 2014). The Ap samples have IDs between 3001 and 5585, the Gr samples between 1 and 2464. The ID enables the linkage of the data belonging to a sample from the individual worksheets or files.

2.2.3. Column “Country”

The country code in the column “Country” indicates in which country the sample was collected.

Code in the column „Country”	Country	Code in the column „Country”	Country
AUS	Austria	GER	Germany
BEL	Belgium	HEL	Greece (Hellas)
BOS	Bosnia and Herzegovina	HUN	Hungary
BUL	Bulgaria	IRL	Ireland
CRO	Croatia	ITA	Italy
CYP	Cyprus	LAV	Latvia
CZR	Czech Republic	LIT	Lithuania
DEN	Denmark	LUX	Luxembourg
EST	Estonia	MON	Montenegro
FIN	Finland	NEL	Netherlands, The
FRA	France	NOR	Norway
FOM	F.Y.R.O.M., now North Macedonia	POL	Poland

Code in the column „Country”	Country	Code in the column „Country”	Country
PTG	Portugal	SWE	Sweden
SRB	Serbia	SIL	Switzerland
SKA	Slovakia	UKR	Ukraine
SLO	Slovenia	UNK	United Kingdom
SPA	Spain		

2.2.4. Column “Country-ID”

The column “Country-ID” contains the sequential number of the sample within the country specified in the column “Country”. The numbers were assigned separately for the Ap and Gr samples and are not randomised.

2.2.5. Column “Type”

The column “Type” gives information about the sample material. For agricultural soil samples the column contains the value “Ap”, for grazing land soil samples the value “Gr”. More details on the sample materials can be found in Reimann (2014).

2.2.6. Column “Type2”

The distribution patterns of many of the analysed parameters show that there exist major geochemical differences between the samples from northern and southern Europe. Therefore, in addition to the overall statistics, separate statistics were provided for the samples from northern and southern Europe (Filzmoser et al., 2014). Column “Type2” indicates, in addition to information about the sample material (information from column “Type”), the allocation of the samples to northern (“N”) or southern (“S”) Europe.

2.2.7. Columns “X_Coo” and “Y_Coo”

The columns “X_Coo” and “Y_Coo” indicate the location of the sample site in decimal degrees. The data refer to the coordinate system EPSG:4326 (WGS84 - World Geodetic System 1984). “X_Coo” contains the longitude, “Y_Coo” the latitude.

2.2.8. Columns “X_LAEA” and “Y_LAEA”

For the preparation of the geochemical maps, especially for the grid calculation by kriging, the transformation of the ellipsoidal coordinates (columns “X_Coo” and “Y_Coo”) into Cartesian coordinates was necessary. The target coordinate system of the transformation was EPSG:3035 (ETRS89 - extended/LAEA Europe). This coordinate system is often used in Europe for statistical representations and other purposes where a true-to-area mapping is required. The columns “X_LAEA” (easting) and “Y_LAEA” (northing) provide the transformed coordinates.

2.2.9. Column “Altitude”

The column “Altitude” indicates the elevation of the sample site in metres above sea level.

2.2.10. Column “APS_1960-1990”

The column “APS_1960-1990” (APS: Annual Precipitation Sum) contains the mean annual precipitation in the area of the respective sample site in the years 1960-1990. They were determined based on the WorldClim database version 1.4 (WorldClim.org, 2005). The climate surfaces (grids of spatially interpolated climate data) included in this database have a spatial resolution of 30 arc seconds. This is equivalent to about 0.86 km^2 at the equator and less elsewhere and commonly referred to as ‘1 km resolution’. A detailed description of the methods used to compile and interpolate the climate data can be found in Hijmans et al. (2005). To generate the data in column “APS_1960-1990”, the sample points were intersected with the bioclimatic variable BIO_{12} (Annual Precipitation) of the WorldClim database (UCDAVIS, 2005b) using ESRI ArcGIS (method “Extract values to points” of the Spatial Analyst Tools).

2.2.11. Column “AMT_1960-1990”

The column “AMT_1960-1990” (AMT: Annual Mean Temperature) contains the annual mean temperature in the area of the respective sample site in the years 1960-1990. They have also been determined based on the WorldClim database version 1.4 (WorldClim.org, 2005, see chapter 2.2.9). To generate the data in column “AMT_1960-1990”, the sample points were intersected with the bioclimatic variable BIO_1 (Annual Mean Temperature) of the WorldClim database (UCDAVIS, 2005a) using ESRI ArcGIS (method “Extract values to points” of the Spatial Analyst Tools).

2.2.12. Column "APS_1970-2000"

The column "APS_1970-2000" (APS: Annual Precipitation Sum) contains the mean annual precipitation in the area of the respective sample site in the years 1970-2000. They were determined based on the WorldClim database version 2.1 (WorldClim.org, 2020). As already in WorldClim database version 1.4, the climate surfaces (grids of spatially interpolated climate data) included in the WorldClim database version 2.1 have a spatial resolution of 30 arc seconds. This is equivalent to about 0.86 km^2 at the equator and less elsewhere and commonly referred to as '1 km resolution'. A detailed description of the methods used to compile and interpolate the climate data can be found in Fick & Hijmans (2017). To generate the data in column "APS_1970-2000", the sample points were intersected with the bioclimatic variable BIO_{12} (Annual Precipitation) of the WorldClim database (UCDAVIS, 2020) using ESRI ArcGIS (method "Extract values to points" of the Spatial Analyst Tools).

It should be noted that only an unquantifiable proportion of the changes in the climatic data for the years 1970-2000 compared to those for the years 1960-1990 can be attributed to the effects of global climate change. A more comprehensive database, improved calculation and interpolation methods and integration of satellite-derived data either as covariates or as independent spline variables will probably have a much more significant influence on the calculation results.

2.2.13. Column "AMT_1970-2000"

The column "AMT_1970-2000" (AMT: Annual Mean Temperature) contains the annual mean temperature in the area of the respective sample site in the years 1970-2000. They have also been determined based on the WorldClim database version 2.1 (WorldClim.org, 2020, see chapter 2.2.11). To generate the data in column "AMT_1970-2000", the sample points were intersected with the bioclimatic variable BIO_1 (Annual Mean Temperature) of the WorldClim database (UCDAVIS, 2020) using ESRI ArcGIS (method "Extract values to points" of the Spatial Analyst Tools).

Note the remarks on the changes in climate data for the years 1970-2000 compared to those for the years 1960-1990 in chapter 2.2.11.

2.2.14. Column “Climate”

The column "Climate" contains information about the climate zone in which the sample site is located. The determination of the climate zone was carried out GIS-based by intersecting the sample points with the map of climate zones after Baritz et al. (2005) (see Figure 10.11 in Birke et al., 2014a). Please also note the remarks on GIS-based attribution in chapter 2.2.23.

Code in the column “Climate”	Climate zone
SpBo	subpolar to boreal climate
BoTe	boreal to temperate climate
Temp	temperate climate
Medi	Mediterranean climate

2.2.15. Column “Soiltype”

The feature “Soiltype” was derived from the results of the determination of the particle size distribution (normalised values, see chapter 2.6.1). The classification is based on the soil type classification used for the Austrian soil mapping (BFW).

Code in the column „Soiltype“	Soil type	Sand	Silt	Clay
		%	%	%
S	sand	65 - 100	0 - 30	0 - 10
zS	silty sand	40 - 70	30 - 55	0 - 5
lS	loamy sand	30 - 80	10 - 55	5 - 15
sZ	sandy silt	10 - 45	55 - 75	0 - 15
Z	silt	0 - 25	75 - 100	0 - 25
tS	clayey sand	65 - 90	0 - 10	10 - 25
sL	sandy loam	20 - 75	10 - 55	15 - 25
lZ	loamy silt	0 - 30	55 - 75	15 - 25
sT	sandy clay	50 - 75	0 - 10	25 - 40
L	loam	5 - 65	10 - 55	25 - 40
zL	silty loam	0 - 20	55 - 75	25 - 45
lT	loamy clay	0 - 60	0 - 55	40 - 50
T	clay	0 - 50	0 - 50	50 - 100

2.2.16. Column “Soilclass”

The feature “Soilclass” is also based on the classification used for the Austrian soil mapping (BFW). The soil class was derived from the soil type.

Code in the column „Soilclass“	Soil class	Associated soil types
ll	very light	sand, silty sand (S, zS)
l	light	loamy sand, sandy silt, silt (lS, sZ, Z)
m	medium	clayey sand, sandy loam, loamy silt (tS, sL, lZ)
s	heavy	sandy clay, loam, silty loam (sT, L, zL)
ss	very heavy	loamy clay, clay (lT, T)

2.2.17. Column “CGSG”

The feature “Chemical and Geological Soil Groups” (CGSG) was generated by combining the information from lithological maps and parent material maps with the geochemical signal from the GEMAS project samples. A more detailed description of the approach can be found in Birke et al. (2014a). This feature was only determined for the Ap samples.

Code in the column “CGSG”	Chemical and geological soil group
alk	alkaline rocks
chalk	calcareous rocks
green	greenstone
loess	loess
org	organic soil
other	other
granite	plutonic felsic rocks
prec	Precambrian rocks
quartz	quartz rich soil
schist	schists

2.2.18. Column “Litho”

The column "Litho" contains information about the lithology in the vicinity of the respective sample site. The determination of the lithology was carried out GIS-based by intersecting the sample points with the digital map of the lithology of the continental surfaces after Dürr et al. (2005) (see Figure 10.5 in Birke et al., 2014a). Please also note the remarks on GIS-based attribution in chapter 2.2.18.

Code in the column “Litho”	Lithology
Pa	plutonic felsic rocks
Pb	plutonic mafic-ultramafic rocks
Va	felsic volcanic rocks
Vb	mafic volcanic rocks
Pr	Precambrian rocks
Mt	metamorphic rocks
Cl	complex lithology
Ss	siliciclastic consolidated sedimentary rocks
Sm	mixed consolidated sedimentary rocks
Sc	carbonated consolidated rocks
Su	non- or semi-consolidated sediments
Ad	alluvial deposits
Lo	loess
Ds	dunes and shifting sand
Ep	evaporites
Wb	inland body of water
Ig	polar ice and glaciers

2.2.19. Column “PM_Hart”

Column “PM_Hart” contains information about the parent material associations influencing the formation of soil in the area of the respective sample site. The determination of the parent material associations was carried out GIS-based by intersecting the sample points with the map of the parent material associations after Hartwich et al. (2005) (see Figure 10.6 in Birke et al., 2014a). Please note the remarks on GIS-based attribution in chapter 2.2.23.

Code in the column "PM_Hart"	Parent material association
quart	Quaternary marine deposits
quart(u)	Quaternary marine deposits (undifferentiated), overlying igneous and metamorphic rocks
fluv	fluvial deposits
morain	morainic deposits
glacfluv	glaciofluvial deposits
glaclac	glaciolacustrine deposits
eol	eolian sand
loess	loess
peat	Peat
tertsed	Tertiary sedimentary rocks (undifferentiated)
tertmol	Tertiary molasse
mesotert	Mesozoic, partly Tertiary rocks (undifferentiated)
mesocalc	Mesozoic calcareous sedimentary rocks
flysch	flysch (undifferentiated) with Tertiary and Mesozoic sedimentary rocks
paleo	Paleozoic, partly Mesozoic and Tertiary sedimentary rocks (undifferentiated), often with igneous and metamorphic rocks
volc	volcanic rocks
ignmet	igneous and metamorphic rocks (undifferentiated), partly with sedimentary rocks

2.2.20. Column "PM_Guen"

Column "PM_Guen" contains information about the parent material(s) influencing the formation of soil in the area of the respective sample site. The determination of the parent material(s) was carried out GIS-based by intersecting the sample points with the map of the parent material after Günther et al. (2013) (see Figure 10.7 in Birke et al., 2014a). Please note the remarks on GIS-based attribution in chapter 2.2.23.

Code in the column "PM_Guen"	Parent material
all/coll	alluvium/colluvium
glacioflu	glaciofluvial materials
calc	calcareous rocks
marl	marl
clayed	clayey materials
sandy	sandy materials
san/fly/mo	sandstone/flysch/molasses
loa/sil	loamy/silty
detri	detrital formations
crys	crystalline rocks
schists	schists
volc	volcanic rocks
oth/org	other/organic

2.2.21. Column “EcoRegio”

The column "EcoRegio" contains information about the ecological region in which the sample site is located. The determination of the ecological region was carried out GIS-based by intersecting the sample points with the map of European ecological regions (DMEER; EEA, 2003) (see Figure 10.13 in Birke et al., 2014a). Please note the remarks on GIS-based attribution in chapter 2.2.23.

Code in the column “EcoRegio“	Ecological region
1	Kola Peninsula tundra
2	Scandinavian montane birch forests and grasslands
3	Scandinavian and Russian taiga
4	Caledon coniferous forests
5	Scandinavian coastal coniferous forests
6	Iberian conifer forests
7	Carpathian montane coniferous forests
8	Alps conifer and mixed forests
9	Pyrenees conifer and mixed forests
10	Sarmatic mixed forests
11	Baltic mixed forests
12	Central European mixed forests
13	Cantabrian mixed forests
14	Pannonian mixed forests
15	Balkan mixed forests
16	Po Basin mixed forests
17	North Atlantic moist mixed forests
18	Northwest Iberian montane forests
19	South Apennine mixed montane forests
20	Dinaric Mountains mixed forests
21	Pindus Mountains mixed forests
22	Rhodope montane mixed forests
23	Northern Temperate Atlantic
24	Southern Temperate Atlantic
25	Celtic broadleaf forests
26	Western European broadleaf forests
27	English Lowlands beech forests
28	Corsican montane broadleaf and mixed forests
29	Illyrian deciduous forests
30	Euxine-Colchic deciduous forest
31	Apennine deciduous montane forests
32	Iberian sclerophyllous and semi-deciduous forests
33	Italian sclerophyllous and semi-deciduous forests
34	Southwest Iberian Mediterranean sclerophyllous and mixed forests
35	Tyrrhenian-Adriatic sclerophyllous and mixed forests
36	Aegean & West Turkey sclerophyllous and mixed forest
37	Southeastern Iberian shrubs and woodlands

Code in the column "EcoRegio"	Ecological region
38	Crimean submediterranean forest complex
39	Northeastern Spain & Southern France Mediterranean
40	Cyprus Mediterranean forests
41	Crete Mediterranean forests
42	East European forest steppe
43	Pontic steppe

2.2.22. Column "PD_2005"

In the original GEMAS project, the population density grids published by the Center for International Earth Science Information Network in 2005 (Gridded Population of the World Version 3 (GPWv3); CIESIN, 2005) were used for the representation of population density (see Fig. 10.14. in Birke et al., 2014a). These data provide the population density from the year 2000 with a resolution of 2.5 arc minutes. Note that reliable data for certain sub-areas, e.g., several of the former Yugoslavian Republics, are scarce. Data availability for Germany and the United Kingdom and further areas is also not sufficient.

Therefore, more up-to-date data sources were also used for the digital provision of the supporting information for the interpretation of the geochemical maps. The version 4 revision 11 of the population density grids (Gridded Population of the World Version 4 Revision 11 (GPWv4.11); CIESIN, 2018) contains population density data for the years 2000, 2005, 2010, 2015 and 2020 with a spatial resolution of 30 arc seconds. This is equivalent to about 0.86 km^2 at the equator and less elsewhere and commonly referred to as '1 km resolution'. The quality of the population density data of GPWv4.11 for the year 2005 is significantly improved compared to the data of GPWv3 for the year 2000. This applies to some Balkan countries, but also to Germany, the United Kingdom and other areas.

To generate the data in column "PD_2005", the sample sites were intersected with the population density grid for the year 2005 (GPWv4.11) using ESRI ArcGIS (method "Extract values to points" of the Spatial Analyst Tools).

2.2.23. Column "PD_2020"

The population density data for the year 2020 in the column "PD_2020" are based on the population density grid for that year contained in the data collection Gridded Population of the World Version 4 Revision 11 (GPWv4.11); CIESIN, 2018; see chapter 2.2.21). To generate the data in column "PD_2020", the sample sites were intersected with the population density grid using ESRI ArcGIS (method "Extract values to points" of the Spatial Analyst Tools).

2.2.24. Notes on GIS-based data collection

GIS-based data collection by intersecting the sample sites with polygon maps is error-prone due to its principle. The polygon maps used to intersect the sample sites of the GEMAS project are mostly small-scale maps on a continental (European) scale (e.g. 1:5 000 000). Therefore, the location of the boundaries of the feature areas is subject to a high uncertainty compared to the accuracy of the location of the sample sites. Furthermore, the positional assignment of the sample sites to the feature areas is always carried out exactly to the area in which the corresponding sample site is located according to its positional co-ordinates. In connection with the described positional uncertainty of the area boundaries, this inevitably leads to a certain number of incorrect assignments, especially if the sample site is located in the border area between two or more feature areas. Moreover, it is impossible with this method to capture the simultaneous geochemical influence from two or more adjacent feature areas (e.g. by several parent materials of soil formation). In this respect, the allocation given in the characteristics is subject to a certain uncertainty and must not be regarded as absolutely correct, but only as a general orientation for statistical evaluations on a European scale.

2.3. GEMAS – AR

2.3.1. One analytics - two data sets

The data set GEMAS - AR contains the analytical results for the elemental contents of 53 chemical elements that can be extracted from the GEMAS samples using aqua regia digestion. The analytical results are contained in the XLSX files “12 GEMAS Ap – AR” and “22 GEMAS Gr – AR” respectively in the corresponding CSV files (cf. tables in chapter 1). Detailed information on the digestion procedure and the analytical methods used (ICP-AES: inductively coupled plasma atomic emission spectrometry, ICP-MS: inductively coupled plasma mass spectrometry) can be found in Birke et al. (2014b). The analytical results of the aqua regia digestions are available for both the Ap and Gr samples. Digestion and analysis were carried out at Acme Analytical Laboratories (Vancouver) Ltd. in Canada.

The XLSX files each contain two data sheets with the analytical results of the aqua regia digestions of the GEMAS samples. The data sheets “Ap AR PDL” and “Gr AR PDL” contain the analytical results taking into account the “practical detection limit” (PDL). The data sheets “Ap AR LDL ACME” and “Gr AR LDL ACME” contain the same analytical results, but taking into account the “official” lower detection limit (LDL) of the AcmeLabs Vancouver laboratory. The data described are also available as CSV files (cf. tables in chapter 1). For a more detailed description of the basics and differences between PDL and LDL, see Demetriades et al. (2014).

The reason for providing the analytical results taking into account of two different detection limits is that for the determination and application of the PDL, the analytical data must be available without any rounding and without cut-off at the laboratory's predetermined detection limit (LDL). Even sub-zero measurements must be recorded and reported for this purpose (Demetriades et al., 2014). This requirements are not met for the data from eastern Ukraine (ID's 2358-2464 and 5481-5585), which were analysed at a later time than the other GEMAS samples. The analytical results of the samples from eastern Ukraine were rounded by the AcmeLabs Vancouver laboratory and the LDL was taken into account. Therefore, there is a break in the analytical results taking into account the PDL ("Ap AR PDL" or "Gr AR PDL") between the data from the main part of the project survey area and the data from eastern Ukraine. However, this only affects the parameters where the PDL is smaller than the LDL. Furthermore, the break only occurs in the samples from eastern Ukraine in which contents below the LDL were observed for one or more parameters.

In the analytical results of the aqua regia digestions of the GEMAS samples taking into account the "official" lower detection limit (LDL) of the AcmeLabs Vancouver laboratory (data sheets "Ap AR LDL ACME" or "Gr AR LDL ACME" in the XLSX files or corresponding CSV files, cf. tables in Chapter 1), this break is not present, because the LDL was applied as the uniform detection limit (LDL) for all samples. However, under certain conditions, this leads to a reduction in the resolution of the data in the content range between PDL and LDL for samples that were not taken in eastern Ukraine.

Therefore, when using the data, for parameters with a deviation of the PDL from the LDL and a frequent occurrence of contents in the range between the PDL and the LDL, it is important to weigh up whether the detailed representation of the content range between the PDL and the LDL in the majority of the project survey area is more important than the correct representation of the data in the lower content ranges in eastern Ukraine.

2.3.2. Detection limits and missing values

The data sheet "DL's" respectively the corresponding CSV file (cf. tables in chapter 1) contain a list of the detection limits for the different parameters. On the one hand, the "official" lower detection limit of the AcmeLabs Vancouver laboratory (LDL), on the other hand the practical detection limit (PDL) is given. As already mentioned, a detailed description of the basics and differences between PDL and LDL can be found in Demetriades et al. (2014). In the data sets, analytical results below the relevant detection limit are indicated with "-DL", whereby the valid numerical value is entered for DL.

Missing values in the data sets are marked by an empty field; there is no separate value for coding the missing values.

2.4. GEMAS – XRF

2.4.1. WD-XRF (Wavelength-dispersive X-ray fluorescence spectrometry)

X-ray fluorescence analysis (XRF) is the preferred method when the total element concentrations in a sample are to be determined. The samples of the GEMAS project were analysed by wavelength dispersive X-ray fluorescence spectrometry (WD-XRF) in the laboratories of the Federal Institute for Geosciences and Natural Resources Hanover, Germany. For a detailed description of the analytical procedure, see Birke et al. (2014b). The contents of 41 elements were analysed.

The analytical results of the wavelength dispersive X-ray fluorescence analysis are available for both the Ap and the Gr samples. They are contained in the XLSX files “13 GEMAS Ap – XRF” and “23 GEMAS Gr – XRF” respectively in the corresponding CSV files (cf. tables in chapter 1). The concentrations of the main components (Si, Ti, Al, Fe, Mn, Mg, Ca, Na, K, P, S, Cl, F) are given in weight-%, whereby the elements, with the exception of Cl and F, are represented by their oxides (SiO_2, TiO_2, Al_2O_3, Fe_2O_3, MnO, MgO, CaO, Na_2O, K_2O, P_2O_5, SO_3). In addition, the loss on ignition (L.O.I.) is given, also in weight-%. These data can be found in the data sheets “Ap WD-XRF Oxides” and “Gr WD-XRF Oxides” respectively in the corresponding CSV files (cf. tables in chapter 1). The contents of all other elements are given in mg/kg and can be found in the data sheets “Ap WD-XRF” and “Gr-WD-XRF” or in the corresponding CSV files (cf. tables in chapter 1). For better comparability, the oxide contents of the main components were converted into element contents and included in these spreadsheets. Here, they are given in mg/kg.

2.4.2. ED-XRF (Energy-dispersive X-ray fluorescence spectrometry)

After publication of the results of the GEMAS project (Reimann et al., 2014a; Reimann et al., 2014b; et al.), further analytical investigations of the sample material were carried out. Among other investigations, the Ap samples were analysed by energy dispersive X-ray fluorescence analysis (ED-XRF) at the laboratories of the Federal Institute for Geosciences and Natural Resources Hanover, Germany. In addition to an improved detection limit for some elements, the concentrations of the elements bromine and iodine could be determined in the Ap samples. The analytical results are documented in the XLSX file "13 GEMAS Ap - XRF" in the data sheet "Ap ED-XRF" respectively in the corresponding CSV file (cf. tables in chapter 1). The contents are given in mg/kg.

2.4.3. Detection limits and missing values

The data sheets “DL's WD-XRF” and “DL's ED-XRF” respectively the corresponding CSV files (cf. tables in chapter 1) contain a list of the detection limits of the different parameters in the according analytical method. In the data sets, analytical results below the relevant detection limit are indicated with “-DL”, whereby the respective valid numerical value is entered for DL.

Missing values in the data records are marked by an empty field, there is no separate value for coding the missing values.

2.5. GEMAS – MMI

In order to determine the environmentally available element fraction, Mobile Metal Ions (MMI®) digestions were carried out on the Ap samples of the GEMAS project. The digestions were analysed for 47 elements. Detailed information on the digestion procedure and the analytical methods used (ICP-OES: optical emission spectrometry with inductively coupled plasma, ICP-MS: mass spectrometry with inductively coupled plasma) can be found in Birke et al. (2014b). Digestion and analysis were carried out by SGS Minerals Services in Canada. The analytical results are documented in the XLSX file “GEMAS Ap – MMI” in the data sheet “Ap MMI” or in the corresponding CSV file (cf. tables in Chapter 1). The contents are given in mg/kg.

The data sheet “DL's” respectively the corresponding CSV file (cf. tables in chapter 1) contains a list of the detection limits for the different parameters. In the data set, analytical results below the relevant detection limit are indicated with “-DL”, whereby the valid numerical value is entered for DL.

Missing values in the data set are marked by an empty field; there is no separate value for coding the missing values.

2.6. GEMAS – Other data

2.6.1. General notes

In addition to the multi-element analyses of the total contents as well as various digestions of the GEMAS samples (see chapters 2.3 - 2.5), the concentrations of some other elements as well as soil properties and parameters were also determined in the samples. In addition, some coefficients and indices were calculated. The results of these analyses and evaluations are contained in the XLSX files “15 GEMAS Ap - Other data” and “25 GEMAS Gr - Other data” respectively in the corresponding CSV files (cf. tables in Chapter 1).

2.6.2. Data Sheet "pH-CEC-C-N-S-χ-CIA"

2.6.2.1. pH values (pH_$CaCl_2$)

The pH values of the soil samples were measured in 0.01 M $CaCl_2$ solution. The measurements were carried out by the Geological Survey of Norway (NGU). A detailed description of the measurement procedure can be found in Birke et al. (2014b). The pH values are available for both the Ap and Gr samples. No pH values were measured on the samples from eastern Ukraine.

2.6.2.2. Effective Cation Exchange Capacity (CEC)

The effective cation exchange capacity (eCEC) was measured on the GEMAS project samples using the silver-thiourea method. The measurements were carried out in the laboratory of the Slovak State Geological Institute (SGIDS). For a detailed description of the measurement procedure, see Birke et al. (2014b) and Appendix 7 of Reimann et al. (2011). Measured values for the effective cation exchange capacity are available for the Ap and Gr samples. They are given in milliequivalent of hydrogen per 100 g (meq/100 g). No measurements of the effective cation exchange capacity were carried out on the samples from eastern Ukraine.

2.6.2.3. Total Carbon (TC, NGU)

Total carbon (TC) was analysed in the GEMAS project samples using a sulphur/carbon analyser LECO SC-444. The measurements were carried out at the laboratory of the Geological Survey of Norway (NGU). For a description of the measurement procedure, see Birke et al. (2014b). Measured values for the total carbon content are available for the Ap and Gr samples. The values are given in weight-%.

2.6.2.4. Total Organic Carbon (TOC)

Total organic carbon (TOC) was determined in the samples of the GEMAS project according to ISO 10694:1995 "Soil quality - Determination of organic and total carbon after dry combustion (elementary analysis)". The measurements were performed by FUGRO Consult Gmbh, now KIWA Control GmbH, using a carbon/sulphur analyser (ELTRA Helios). A description of the measurement procedure can be found in Birke et al. (2014b). Measured values for total organic carbon are available for the Ap and Gr samples. The values are given in weight-%.

2.6.2.5. Total Sulphur (TS, NGU)

Total sulphur (TS) was analysed in the GEMAS project samples using a sulphur/carbon analyser LECO SC-444. The measurements were carried out at the laboratory of the Geological Survey of Norway (NGU). For a description of the measurement procedure, see Birke et al. (2014b). Measured values for the total sulphur are available for the Ap and Gr samples. The values are given in weight-%.

2.6.2.6. Total Carbon (TC, TU BA FG)

For the Ap samples, total carbon was determined again by the laboratory of the Technical University Bergakademie Freiberg. The aim of this repeat analysis was a more precise determination of the contents with a lower detection limit. A description of the measurement procedure can be found in Matschullat et al. (2018). The data are given in weight-%. No analytical results are available for the samples from eastern Ukraine.

2.6.2.7. Total Nitrogen (TN, TU BA FG)

For the Ap samples, total nitrogen was also determined by the laboratory of the Technical University Bergakademie Freiberg. A description of the measurement procedure can be found in Matschullat et al. (2018). The data are given in weight-%. No analytical results are available for the samples from eastern Ukraine.

2.6.2.8. Total Sulphur (TS, TU BA FG)

For the Ap samples, total sulphur was determined again by the laboratory of the Technical University Bergakademie Freiberg. The aim of this repeat analysis was a more precise determination of the contents with a lower detection limit. A description of the measurement procedure can be found in Matschullat et al. (2018). The data are given in weight-%. No analytical results are available for the samples from eastern Ukraine.

2.6.2.9. Susceptibility (χ)

Weight normalised magnetic susceptibility of the Ap samples was measured using a Sapphire instruments SI2B susceptibility meter with an internal coil frequency of about 19.2 kHz. A description of the measurement procedure can be found in Birke et al. (2014b). The measurements were carried out by the Geological Survey of Norway (NGU). The susceptibility is given in 10^{-6} m^3/kg. No measurement results are available for the samples from eastern Ukraine.

2.6.2.10. Chemical Index of Alteration (CIA)

For all samples of the GEMAS project (Ap and Gr samples), the CIA (Chemical Index of Alteration) was calculated according to Nesbitt & Young (1982). The calculation is based on the total contents determined by wavelength dispersive X-ray fluorescence analysis (WD-XRF, see chapter 2.4.1). A more detailed description of the calculation method can be found in Nesbitt & Young (1982) and in chapter 11.2 of Reimann et al. (2014a).

2.6.3. Data sheet "Ap Pb isotope ratios" (Lead isotope ratios)

The Pb isotope ratio measurements were carried out on a high-resolution sector field inductively coupled plasma mass spectrometer (HR-ICP-MS). The measurements were carried out by the Geological Survey of Norway (NGU). A description of the measurement procedure can be found in Birke et al. (2014b). The isotope ratios $^{206}Pb/^{207}Pb$, $^{206}Pb/^{208}Pb$ and $^{207}Pb/^{208}Pb$ were determined. The measurements were taken exclusively on the Ap samples (including the samples from eastern Ukraine) and on the Gr samples from eastern Ukraine.

2.6.4. Data sheet "Ap PSD" (particle size distribution)

The determination of the particle size (grain size) distribution (PSD) in the soil samples was a "special story". A detailed description of the procedure used and the problems encountered can be found in Birke et al. (2014b). Finally, the determination of the particle size distribution was carried out on the basis of the mid infrared spectra (MIR) of the soil samples measured at CSIRO Land and Water Adelaide, Australia, and a prediction model for the European soils developed on the basis of sedigraph measurements. Note that the prediction of the proportions of clay, silt and sand based on the MIR spectra is based on three different models. Therefore, the sum of the three fractions is not always 100 %. The unchanged prediction data of the three fractions are included in the columns "Sand", "Silt" and "Clay". The columns of the same name with the supplement "(norm)" contain the values normalised to 100 %. The particle size distribution was determined for both the Ap and the Gr samples. No data are available for the samples from eastern Ukraine.

2.6.5. Data sheet „Ap Kd values"

The solid-solution partitioning coefficient (K_d value) has a significant influence on the mobility and bioavailability of metals in soils. More information on its determination and importance can be found in Janik et al. (2014). In the GEMAS project, K_d values were determined for 10 metals in both the Ap and Gr samples. No data are available for the samples from eastern Ukraine.

2.6.6. Data sheet „Ap PNEC & RCR“

The PNEC (predicted no effect concentration) and the RCR (risk characterisation ratio) were determined in the GEMAS project for the metals cobalt (Co), copper (Cu) and molybdenum (Mo) in both the Ap and Gr samples. The procedure for determining the two parameters and a discussion of the results can be found in Oorts & Schoeters (2014). No data are available for the samples from eastern Ukraine.

2.6.7. Data sheet „DL’s“

The data sheet “DL's” respectively the corresponding CSV file (cf. tables in chapter 1) contain a list of the detection limits for the different parameters. In the data sets, analytical results below the relevant detection limit are indicated with “-DL”, whereby the valid numerical value is entered for DL.

Missing values in the data sets are marked by an empty field; there is no separate value for coding the missing values.

References

Baritz, R., Fuchs, M., Hartwich, R., Krug, D. & Richter, S. (2005): Soil Regions of the European Union and Adjacent Countries 1:5 000 000 (Version 2.0) - Europaweite thematische Karten und Datensätze. European Soil Bureau Network: 122.

BFW (Bundesforschungs- und Ausbildungszentrum für Wald, Naturgefahren und Landschaft): Einführung in die bodenkundlichen Grundlagen. https://geo.bfw.ac.at/boden/downloads/Einfuehrung_Bodenkartierung.pdf (last accessed 2021-04-06).

Birke, M., Rauch, U. & Reimann, C. (2014a): Supporting Information for Interpretation of Geochemical Maps. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds.): Chemistry of Europe's Agricultural Soils, Part A: Methodology and Interpretation of the GEMAS Data Set. (Geologisches Jahrbuch). B102: 93-102; Stuttgart (Schweizerbart Science Publishers).

Birke, M., Reimann, C. & Fabian, K. (2014b): Analytical Methods Used in the GEMAS Project. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds.): Chemistry of Europe's Agricultural Soils, Part A: Methodology and Interpretation of the GEMAS Data Set. (Geologisches Jahrbuch). B102: 41-46; Stuttgart (Schweizerbart Science Publishers).

CIESIN (Center for International Earth Science Information Network) (2018): Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets. Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H45Q4T5F. (downloaded: 2020-06-07).

CIESIN (Center for International Earth Science Information Network), CIAT (Centro Internacional de Agricultura Tropical) (2005): Gridded Population of the World Version 3 (GPWv3): Population Density Grids. Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, NY. http://sedac.ciesin.columbia.edu/gpw. (downloaded: 2012-01-08).

Demetriades, A., Reimann, C. & Filzmoser, P. (2014): Evaluation of GEMAS Project Quality Control Results. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds.): Chemistry of Europe's Agricultural Soils, Part A: Methodology and Interpretation of the GEMAS Data Set. (Geologisches Jahrbuch). B102: 47-60; Stuttgart (Schweizerbart Science Publishers).

Dürr, H.H., Meybeck, M. & Dürr, S.H. (2005): Lithologic composition of the Earth's continental surfaces derived from a new digital map emphasizing riverine material transfer. Global Biogeochemical Cycles, 19, 4: 1-22. DOI: https://doi.org/10.1029/2005GB002515.

EEA (European Environmental Agency, 2003): Digital map of European ecological regions (DMEER). http://www.eea.europa.eu/data-and-maps/data/ds_resolveuid/A770EED0-74AA-4007-8DA5-19B2834E0F5E. (downloaded: 2013-01-15, last accessed: 2021-04-29).

Fick, S.E. & Hijmans, R.J. (2017): WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37, 12: 4302-4315. DOI: https://doi.org/10.1002/joc.5086.

Filzmoser, P, Reimann, C. & Birke. M. (2014): Univariate Data Analysis and Mapping. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds.): Chemistry of Europe's Agricultural Soils, Part A: Methodology and Interpretation of the GEMAS Data Set. (Geologisches Jahrbuch). B102: 67-81; Stuttgart (Schweizerbart Science Publishers).

Günther, A., Eeckhaut, M., Reichenbach, P., Hervás, J., Malet, J.-P., Foster, C. & Guzzetti, F. (2013): New Developments in Harmonized Landslide Susceptibility Mapping over Europe in the Framework of the European Soil Thematic Strategy. (In: Margottini, C., Canuti, P. & Sassa, K. (Eds.): Landslide Science and Practice: Volume 1: Landslide Inventory and Susceptibility and Hazard Zoning). 297-301; Berlin, Heidelberg (Springer). DOI: https://doi.org/10.1007/978-3-642-31325-7_39.

Hartwich, R., Baritz, R., Fuchs, M., Krug, D. & Thiele, S. (2005): Soil Regions Map of the European Union and Adjacent Countries 1:5.000.000 (Version 2.0) – Associations of dominant parent material, EU catalogue number S.P.I.05.134. European Soil Bureau Research. European Soil Bureau Research Reports: 122; Luxembourg.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005): Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 15: 1965-1978. DOI: https://doi.org/10.1002/joc.1276.

Janik, L.J., Forrester, S., Kirby, J.K., McLaughlin, M.J., Soriano-Disla, J.M. & Reimann, C. (2014): Prediction of Metal and Matalloid Partitioning Coefficients (Kd) in Soil Using Mid-Infrared Diffuse Reflectance Spectroscopy. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds.): Chemistry of Europe's Agricultural Soils. Part B: General Background Information and Further Analysis of the GEMAS Data Set). Geologisches Jahrbuch Reihe B: Regionale Geologie Ausland und Angewandte Geowissenschaften, B 103: 183-188; Stuttgart (Schweizerbart Science Publishers).

Matschullat, J., Reimann, C., Birke, M. & Carvalho, D. (2018): GEMAS: CNS concentrations and C/N ratios in European agricultural soil. Science of the Total Environment, 627: 975-984. DOI: https://doi.org/10.1016/j.scitotenv.2018.01.214.

Nesbitt, H.W. & Young, G.M. (1982): Early Proterozoic climates and plate motions inferred from major element chemistry of lutites. Nature, 299, 5885: 715-717. DOI: https://doi.org/10.1038/299715a0.

Oorts, K. & Schoeters, I. (2014): Use of Monitoring Data for Risk Assessment of Metals in Soil under the European REACH Regulation. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds.): Chemistry of Europe's Agricultural Soils. Part B: General Background Information and Further Analysis of the GEMAS Data Set). Geologisches Jahrbuch Reihe B: Regionale Geologie Ausland und Angewandte Geowissenschaften, B 103: 189-202; Stuttgart (Schweizerbart Science Publishers).

Reimann, C. (2014): Sample Material, Site Selection and Sampling. In: Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds.): Chemistry of Europe's Agricultural Soils, Part A: Methodology and Interpretation of the GEMAS Data Set. (Geologisches Jahrbuch). B102: 31-36; Stuttgart (Schweizerbart Science Publishers).

Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds.) (2014a): Chemistry of Europe's Agricultural Soils, Part A: Methodology and Interpretation of the GEMAS Data Set. (Geologisches Jahrbuch). B102: 523; Stuttgart (Schweizerbart Science Publishers).

Reimann, C., Birke, M., Demetriades, A., Filzmoser, P. & O'Connor, P. (eds.) (2014b): Chemistry of Europe's Agricultural Soils, Part B: General Background Information and Further Analysis of the GEMAS Data Set. (Geologisches Jahrbuch). Vol. B 103: 352; Stuttgart (Schweizerbart Science Publishers).

Reimann, C., Demetriades, A., Eggen, O.A., Filzmoser, P. & EuroGeoSurveys Geochemistry Working Group (2011): The EuroGeoSurveys geochemical mapping of agricultural and grazing land soils project (GEMAS) – Evaluation of quality control results of total C and S, total organic carbon (TOC), cation exchange capacity (CEC), XRF, pH, and particle size distribution (PSD) analysis. NGU Report 2011.043. Geological Survey of Norway: 92; Trondheim. http://www.ngu.no/upload/publikasjoner/rapporter/2011/2011_043.pdf (last accessed: 2021-07-14).

UCDAVIS (University of California, Davis campus, 2020): https://biogeo.ucdavis.edu/data/worldclim/v2.1/base/wc2.1_30s_bio.zip. (downloaded: 2021-04-07, last accessed: 2021-04-07).

UCDAVIS (University of California, Davis campus, 2005a): https://biogeo.ucdavis.edu/data/climate/worldclim/1_4/grid/cur/bio1-9_30s_bil.zip. (downloaded: 2012-10-23, last accessed: 2021-05-26).

UCDAVIS (University of California, Davis campus, 2005b): https://biogeo.ucdavis.edu/data/climate/worldclim/1_4/grid/cur/bio10-19_30s_bil.zip. (downloaded: 2012-10-23, last accessed: 2021-05-26).

WorldClim.org (2020): WorldClim 2.1 climate data for 1970-2000. https://worldclim.org/data/worldclim21.html. (downloaded: 2021-04-07, last accessed 2021-07-04).

WorldClim.org (2005): WorldClim 1.4 historical climate conditions for 1960-1990. https://www.worldclim.org/data/v1.4/worldclim14.html. (downloaded: 2012-10-23, last accessed: 2021-04-07).