London Atlas: Materials and methods III: data analysis

From Earthwise

Jump to navigation Jump to search

Ferreira, A, Johnson, C C, Appleton, J D, Flight, D M A, Lister, T R, Knights, K V, Ander, L, Scheib, C, Scheib, A, Cave, M, Wragg, J, Fordyce, F and Lawley, R. 2017. London Region Atlas of Topsoil Geochemistry. *British Geological Survey*.

☐

Contents

The subsets SEEN and LOND

The London Region Topsoil Dataset (**LRD**) is composed of the **LOND** and **SEEN** data subsets, which were collected using two different sampling densities (**LOND** 1 per 0.25 km^2, **SEEN** 1 per 2 km^2), and broadly correspond to two regions, which are very different in terms of population density and historical record of human occupancy. About 95% of **LOND** and **SEEN** samples are located respectively inside and outside of the GLA. **LOND** samples collected outside GLA (307 out of 6801) are grouped into four urban clusters located to the east of the GLA, while **SEEN** samples collected inside GLA (80 out of 1599) are mainly located in the northern and north-eastern London outskirts, always close to the GLA border (**Figure 2**). Given the different sampling densities between **LOND** and **SEEN** subsets, **LRD** statistics are necessarily biased towards the values observed inside the GLA, thus not adequately representing the ***area*** of the London Region Atlas of Topsoil Geochemistry. The effect of this bias is particularly important for elements for which concentrations are expected to be much higher (or much lower) inside the GLA, such as those commonly classified as anthropogenic.

Accordingly, it was decided to include statistics and graphs of **LOND** and **SEEN** subsets separately in addition to the statistics and graphs of the London Region Topsoil Dataset as a whole (**LRD**). By showing the summary statistics and distribution of the two subsets (**SEEN** and **LOND**), and taking into account that they mainly represent the area outside and inside the GLA respectively, an unbiased view of topsoil geochemistry within the built up area of London relative to the outskirts is obtained. This enables a better understanding of soil quality in the LRA area, the distinctive anthropogenic geochemical signal observed within and outside the GLA becoming particularly clear.

The term *urban domain* in Appleton et al. (2013)[1], one of the key references of the present atlas, is used as in Ander et al. (2011)[2] and Ander et al. (2013)[3], while *rural domain* includes both *rural* and *semi-urban* areas defined in the same document. This *urban domain* definition is based on a urbanisation index after the Generalised Land Use Database (GLUD) Statistics for England, 2005 (Department for Communities and Local Government, 2007[4]), a document from the Office of National Statistics (ONS). This definition is such that its complementary *rural* includes large openspace areas within central London, such as Richmond Park and Wimbledon Common in south-west London. For the present atlas, *urban* and *rural* are rather used to define the sampling survey type (**LOND** and **SEEN**, respectively) or when referring to the 'GLA area' or 'central London' (urban) relative to 'outwith GLA area' or 'London outskirts' (rural).

Univariate statistics and graphics

A preliminary outline of the 46 variables (44 elements, LOI, pH) in the **LRD** dataset and in the two subsets (**SEEN** and **LOND**) is given in **Table 9**. This was built in *Microsoft Excel*®, and includes some of the most commonly used distribution measures reporting univariate descriptive statistics, namely percentiles and some non-parametric statistics in attempting to obtain robust statistics, less dependent on outliers (Reimann et al., 2008)[5]. It includes the lower limit of detection (**LLD**), 9 **percentiles** (or quantiles, **Q02, Q05, Q10, Q25, Q75, Q90, Q95, Q98, Q99**), the minimum (**Min**) and the maximum (**Max**); measures of central tendency, namely, the median (**Mdn**), the geometric mean (**GM**) and the arithmetic mean (**AM**); some measures of dispersion, namely, the interquartile range (**IQR**), the median absolute deviation (**MAD**), the standard deviation (**SD**) and the geometric standard deviation (**GSD**); two other measures of dispersion, the coefficient of variation (**CV%**) and the robust coefficient of variation (**CVR%**), expressed in percent, are also shown, as they have the advantage of being independent of the magnitude of the data (Reimann et al., 2008[5]). The CV% is defined as the SD divided by the AM, while the CVR% is defined as the MAD divided by the Mdn. The MAD is a robust equivalent of the SD measuring the average deviation from a central value, in this case the median. **Powers**, also a measure of dispersion, are defined as the decimal logarithm of the ratio between the Max and the Min, thus showing the orders of magnitude of the variation. Finally, the skewness is also shown, both for normal (**Skew**) and log-transformed (**SkewLOG**) data. Skewness is a measure of asymmetry of the distribution, indicating if the tails on both sides of the AM (or the GM) balance out or not. This statistic, however, must always be evaluated together with other information, such as a histogram or other graphical representation of the distribution, as has been done for the present work. These measures of asymmetry can be helpful in deciding whether to use a linear or a logarithmic scale in the graphical representation of element concentrations, and, together with the descriptive graphics, can be used to decide the most appropriate parametric statistics, in case these are required.

Table 9 Summary table providing a selection of statistical parameters to describe analytes determined on topsoils from the London region. Number of samples is LRD = 8400 (7928 for LOI and 7929 for pH), SEEN = 1599 (1128 for LOI and pH) and LOND = 6801 (6800 for LOI). For further explanation go to Univariate statistics and graphics

Analytes units	Dataset	LLD	Min	Q02	Q05	Q10	Q25	Mdn	Q75	Q90	Q95	Q98	Q99	Max	IQR	MAD	CVR%	AM	SD	CV%	GM	GSD	Powers	Skew	Skew LOG
Al_2O_3 %	**LRD**	0.2	0.8	3.0	3.9	4.7	5.9	**7.6**	10.0	12.0	13.2	14.4	15.3	25.5	4.1	2.8	37.1	8.0	2.9	35.9	7.5	1.47	1.50	**0.52**	-0.71
	SEEN	0.2	0.8	2.3	3.0	3.8	5.7	**8.2**	10.7	12.2	13.4	14.6	15.6	19.9	4.9	3.7	45.2	8.2	3.3	39.7	7.4	1.62	1.40	**0.09**	-1.11
	LOND	0.2	0.8	3.4	4.3	4.9	5.9	**7.5**	9.7	11.9	13.2	14.4	15.2	25.5	3.8	2.7'	35.6'	8.0	2.8	34.9	7.5	1.43	1.50	0.66	**-0.44**
CaO %	**LRD**	0.05	<0.05	0.27	0.37	0.50	0.74	**1.18**	2.23	4.31	6.95	15.45	23.26	48.97	1.49	0.84	71.3	2.30	3.97	172.7	1.34	2.51	3.09	5.58	**0.62**
	SEEN	0.05	<0.05	0.11	0.18	0.29	0.55	**0.80**	1.44	5.75	13.14	28.04	32.48	47.47	0.89	0.49	61.4	2.64	5.97	226.0	0.99	3.33	3.07	4.20	**0.84**
	LOND	0.05	0.22	0.34	0.43	0.56	0.81	**1.31**	2.31	4.20	6.15	12.84	18.54	48.97	1.49	0.91	69.1	2.22	3.33	149.9	1.44	2.29	2.35	5.86	**0.76**
Fe_2O_3 %	**LRD**	0.01	0.13	1.56	2.09	2.48	3.10	**3.80**	4.70	5.52	6.09	6.66	7.14	15.59	1.60	1.17	30.7	3.93	1.27	32.4	3.71	1.43	2.08	**0.87**	-1.42
	SEEN	0.01	0.13	0.91	1.51	2.00	2.80	**3.65**	4.53	5.25	5.90	6.62	7.31	15.59	1.73	1.29	35.3	3.70	1.42	38.3	3.39	1.59	2.08	**1.06**	-1.83
	LOND	0.01	0.15	1.81	2.24	2.56	3.15	**3.82**	4.76	5.57	6.11	6.66	7.12	15.37	1.61	1.14'	29.7'	3.98	1.23	30.9	3.79	1.39	2.00	**0.85**	-0.89
K_2O %	**LRD**	0.01	0.12	0.57	0.74	0.88	1.08	**1.32**	1.71	2.07	2.28	2.48	2.58	3.47	0.63	0.43'	32.6'	1.41	0.47	33.2	1.33	1.42	1.46	**0.52**	-0.72
	SEEN	0.01	0.15	0.48	0.63	0.79	1.05	**1.36**	1.83	2.15	2.32	2.51	2.65	3.47	0.78	0.55'	40.3'	1.43	0.52	36.6	1.32	1.52	1.36	**0.26**	-1.01
	LOND	0.01	0.12	0.63	0.78	0.90	1.09	**1.32**	1.68	2.06	2.27	2.46	2.57	3.33	0.59	0.42'	31.4'	1.40	0.45	32.2	1.33	1.40	1.44	0.59	**-0.57**

MgO %	LRD	0.3	<0.3	0.3	0.4	0.4	0.6	0.8	1.1	1.4	1.6	2.0	2.2	4.6	0.5	0.3'	37.1'	0.9	0.4	48.5	0.8	1.59	1.36	1.65	-0.16
	SEEN	0.3	<0.3	<0.3	0.3	0.5	0.7	0.8	1.0	1.2	1.4	1.8	2.0	4.2	0.3	0.3'	37.1'	0.9	0.3	40.6	0.8	1.53	1.32	1.78	-0.82
	LOND	0.3	<0.3	0.3	0.4	0.4	0.6	0.8	1.1	1.4	1.7	2.0	2.3	4.6	0.5	0.3'	37.1'	0.9	0.4	50.2	0.8	1.61	1.36	1.62	-0.05
Analytes units	Dataset	LLD	Min	Q02	Q05	Q10	Q25	Mdn	Q75	Q90	Q95	Q98	Q99	Max	IQR	MAD	CVR%	AM	SD	CV%	GM	GSD	Powers	Skew	Skew LOG
MnO %	LRD	0.005	<0.005	0.013	0.022	0.030	0.043	0.056	0.075	0.108	0.147	0.204	0.264	0.697	0.032	0.022	39.6	0.066	0.047	70.7	0.056	1.79	2.24	4.01	-0.35
	SEEN	0.005	<0.005	0.008	0.016	0.025	0.043	0.071	0.108	0.162	0.205	0.261	0.290	0.573	0.065	0.047	66.9	0.085	0.062	73.6	0.065	2.21	2.16	1.98	-0.74
	LOND	0.005	<0.005	0.016	0.024	0.031	0.043	0.055	0.070	0.093	0.120	0.170	0.226	0.697	0.027	0.019	35.2	0.062	0.041	66.6	0.054	1.67	2.24	5.23	-0.27
Na_2O %	LRD	0.3	<0.3	<0.3	<0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	10.0	0.2	0.1	34.2	0.4	0.2	47.5	0.4	1.39	1.70	21.43	0.24
	SEEN	0.3	<0.3	<0.3	<0.3	<0.3	0.3	0.4	0.6	0.7	0.7	0.8	0.9	1.1	0.3	0.1	37.1	0.4	0.2	37.8	0.4	1.47	0.74	0.65	-0.12
	LOND	0.3	<0.3	<0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	10.0	0.2	0.1	34.2	0.4	0.2	49.5	0.4	1.38	1.70	23.28	0.40
P_2O_5 %	LRD	0.05	<0.05	0.12	0.15	0.18	0.22	0.30	0.40	0.54	0.66	0.88	1.09	4.49	0.18	0.12	39.5	0.34	0.22	64.3	0.30	1.60	2.05	5.82	0.34
	SEEN	0.05	<0.05	0.09	0.12	0.14	0.18	0.24	0.30	0.39	0.47	0.59	0.77	3.26	0.12	0.09	37.0	0.26	0.17	64.3	0.24	1.57	1.91	8.38	0.00
	LOND	0.05	<0.05	0.13	0.16	0.19	0.24	0.31	0.42	0.56	0.69	0.91	1.13	4.49	0.18	0.13	43.1	0.36	0.23	62.7	0.32	1.58	2.05	5.67	0.50
SiO_2 %	LRD	0.1	4.6	34.9	48.0	54.2	60.2	66.6	73.3	79.2	82.7	86.4	89.1	100.0	13.1	9.8	14.7	66.1	11.4	17.3	64.8	1.25	1.34	-1.04	-3.48
	SEEN	0.1	5.9	20.6	39.0	52.0	61.6	69.7	76.8	83.4	86.9	90.4	93.5	100.0	15.2	11.3	16.2	67.7	14.7	21.6	65.3	1.37	1.23	-1.37	-3.34
	LOND	0.1	4.6	39.3	49.1	54.5	59.9	66.0	72.6	78.1	81.4	84.6	87.1	100.0	12.7	9.3	14.2	65.7	10.5	16.0	64.7	1.22	1.34	-0.90	-3.17
TiO_2 %	LRD	0.01	0.17	0.33	0.37	0.41	0.47	0.57	0.70	0.82	0.88	0.93	0.97	1.18	0.23	0.16	28.3	0.59	0.16	26.5	0.57	1.31	0.83	0.44	-0.21
	SEEN	0.01	0.18	0.29	0.34	0.38	0.49	0.63	0.75	0.86	0.90	0.95	0.98	1.06	0.26	0.19	30.6	0.62	0.17	28.1	0.59	1.36	0.78	-0.05	-0.73
	LOND	0.01	0.17	0.34	0.38	0.42	0.47	0.56	0.68	0.81	0.87	0.93	0.96	1.18	0.21	0.15	26.2	0.59	0.15	25.9	0.57	1.29	0.83	0.57	-0.04
Analytes units	Dataset	LLD	Min	Q02	Q05	Q10	Q25	Mdn	Q75	Q90	Q95	Q98	Q99	Max	IQR	MAD	CVR%	AM	SD	CV%	GM	GSD	Powers	Skew	Skew LOG
Ag mg/kg	LRD	0.5	<0.5	<0.5	<0.5	<0.5	0.5	0.5	0.5	1.0	2.2	6.2	11.5	268.8	0.0	0.0	0.0	1.1	5.8	552.0	0.6	1.95	2.95	30.49	3.35
	SEEN	0.5	<0.5	<0.5	<0.5	<0.5	0.5	0.5	0.5	0.5	1.4	4.2	7.3	237.2	0.0	0.0	0.0	0.9	6.2	681.6	0.5	1.74	2.90	34.46	4.35
	LOND	0.5	<0.5	<0.5	<0.5	<0.5	0.5	0.5	0.5	1.1	2.5	6.6	12.4	268.8	0.0	0.0	0.0	1.1	5.7	525.9	0.6	1.99	2.95	29.22	3.19
As mg/kg	LRD	2.4	<2.4	6.8	8.6	10.0	12.2	14.8	18.2	22.9	27.5	36.9	46.5	160.9	6.0	4.3	29.1	16.3	8.2	50.7	15.0	1.46	1.91	5.42	0.31
	SEEN	2.4	<2.4	4.5	6.0	7.6	10.2	12.7	15.2	17.8	20.4	25.4	29.3	111.8	5.0	3.7	29.2	13.1	5.6	43.1	12.1	1.48	1.75	5.36	-0.65
	LOND	2.4	<2.4	8.1	9.5	10.7	12.8	15.4	18.8	24.0	28.5	38.8	48.8	160.9	6.0	4.3	27.9	17.0	8.6	50.4	15.8	1.43	1.91	5.46	0.76
Ba mg/kg	LRD	1	139	229	257	283	324	371	417	490	577	738	885	3475	93	70	18.8	389	140	36.0	374	1.30	1.40	6.48	1.26
	SEEN	1	139	198	228	248	298	340	376	405	422	443	518	1850	78	58	17.0	339	91	26.7	330	1.24	1.12	6.88	0.40
	LOND	1	143	242	271	293	331	380	429	512	602	763	920	3475	98	73	19.1	401	147	36.6	386	1.30	1.39	6.43	1.41
Bi mg/kg	LRD	0.3	<0.3	<0.3	<0.3	<0.3	<0.3	<0.3	0.6	1.2	2.2	4.2	7.5	70.5	0.4	0.0	0.0	0.8	2.5	338.5	0.4	2.42	2.55	14.43	1.72
	SEEN	0.3	<0.3	<0.3	<0.3	<0.3	<0.3	<0.3	0.4	0.7	1.0	1.9	3.2	59.3	0.2	0.0	0.0	0.5	1.9	381.4	0.3	1.94	2.47	23.74	2.17
	LOND	0.3	<0.3	<0.3	<0.3	<0.3	<0.3	<0.3	0.6	1.4	2.4	4.7	8.1	70.5	0.4	0.0	0.0	0.8	2.7	328.9	0.4	2.51	2.55	13.41	1.61
Br mg/kg	LRD	0.8	1.1	5.3	6.4	7.4	9.2	11.4	14.5	18.5	22.1	27.4	33.6	241.1	5.3	3.9	33.8	12.7	6.9	54.5	11.6	1.48	2.34	8.76	0.34
	SEEN	0.8	1.1	4.2	5.5	6.4	7.9	9.5	11.6	14.6	17.7	23.3	28.2	94.8	3.7	2.7	28.1	10.4	5.6	53.5	9.6	1.47	1.94	6.35	0.28
	LOND	0.8	1.5	5.7	6.8	7.8	9.7	12.0	15.0	19.0	22.5	28.5	34.1	241.1	5.3	3.9	32.1	13.2	7.1	53.7	12.1	1.47	2.21	9.22	0.42
Analytes units	Dataset	LLD	Min	Q02	Q05	Q10	Q25	Mdn	Q75	Q90	Q95	Q98	Q99	Max	IQR	MAD	CVR%	AM	SD	CV%	GM	GSD	Powers	Skew	Skew LOG
Cd mg/kg	LRD	0.5	<0.5	<0.5	<0.5	<0.5	<0.5	0.6	0.8	1.3	1.9	3.8	6.8	165.2	0.5	0.3	49.4	0.9	3.5	375.2	0.6	1.95	2.74	30.07	1.81
	SEEN	0.5	<0.5	<0.5	<0.5	<0.5	<0.5	0.5	0.6	0.9	1.3	2.1	3.4	135.8	0.3	0.3	59.3	0.8	4.7	605.7	0.5	1.81	2.66	25.08	2.81
	LOND	0.5	<0.5	<0.5	<0.5	<0.5	0.5	0.6	0.8	1.4	2.1	4.2	7.3	165.2	0.3	0.3	49.4	1.0	3.1	325.4	0.6	1.95	2.74	31.43	1.72
Ce mg/kg	LRD	1	16	29	33	37	43	50	59	68	74	85	101	238	16	11.9	23.7	51.8	14.2	27.5	50.1	1.29	1.17	1.93	0.11
	SEEN	1	16	24	29	33	44	55	65	75	86	106	117	165	21	14.8	27.0	55.7	18.3	32.8	52.8	1.39	1.01	1.04	-0.38
	LOND	1	18	31	34	38	43	50	57	65	70	79	91	238	14	10.4	20.8	50.9	12.9	25.4	49.5	1.26	1.12	2.27	0.27
Co mg/kg	LRD	1.5	<1.5	3.6	5.3	6.7	9.0	11.4	14.4	18.0	20.8	26.0	31.6	85.2	5.4	3.9	33.8	12.2	5.5	45.5	11.1	1.57	1.93	2.63	-0.90
	SEEN	1.5	<1.5	2.1	3.4	4.9	8.1	11.4	14.6	19.3	23.7	29.1	36.4	69.9	6.5	4.9	42.9	12.1	6.6	55.0	10.4	1.82	1.84	1.95	-1.04
	LOND	1.5	<1.5	4.3	5.8	7.1	9.2	11.4	14.3	17.7	20.2	24.7	30.6	85.2	5.1	3.7	32.5	12.2	5.2	43.0	11.3	1.50	1.93	2.91	-0.59
Cr mg/kg	LRD	3	9	38	46	52	61	73	88	104	116	140	172	2094	27	19	26.4	78	45	58.3	73	1.38	2.37	20.75	0.79
	SEEN	3	9	32	39	47	61	74	88	104	118	133	155	718	27	19	26.0	77	35	45.3	72	1.42	1.90	8.39	-0.22
	LOND	3	15	41	47	52	61	72	89	104	115	142	177	2094	28	19	26.8	78	48	60.8	74	1.37	2.14	21.51	1.15
Cs mg/kg	LRD	1	1.0	1.0	1.5	2.0	2.0	3.0	4.0	5.0	6.0	6.9	7.0	11.2	2.0	1.5	49.4	3.2	1.4	44.5	2.9	1.55	1.05	1.12	-0.14
	SEEN	1	1.0	1.0	2.0	2.0	2.0	3.0	4.7	5.8	6.9	7.0	8.0	11.2	2.7	1.5	49.4	3.5	1.6	45.0	3.2	1.58	1.05	0.92	-0.18
	LOND	1	1.0	1.0	1.0	2.0	2.0	3.0	4.0	5.0	6.0	6.0	7.0	11.0	2.0	1.5	49.4	3.1	1.3	43.6	2.8	1.54	1.04	1.14	-0.15
Analytes units	Dataset	LLD	Min	Q02	Q05	Q10	Q25	Mdn	Q75	Q90	Q95	Q98	Q99	Max	IQR	MAD	CVR%	AM	SD	CV%	GM	GSD	Powers	Skew	Skew LOG
Cu mg/kg	LRD	1.3	3	10	13	17	24	38	67	114	167	285	433	5326	43	25.7	67.7	63.5	132	207.8	41.7	2.23	3.25	19.83	0.71
	SEEN	1.3	3	7	9	11	17	22	29	41	55	96	180	1055	13	8.3	38.4	29.2	54	186.2	22.1	1.83	2.55	12.97	1.16
	LOND	1.3	3	13	16	20	28	45	75	126	184	315	461	5326	47	29.7	65.9	71.6	143	199.9	48.4	2.15	3.25	18.91	0.76
Ga mg/kg	LRD	1.0	1.0	6.5	7.2	7.9	9.1	10.6	12.9	15.1	16.3	18.0	18.9	44.1	3.8	2.7	25.2	11.1	2.9	25.9	10.8	1.29	1.64	0.92	-0.04
	SEEN	1.0	3.1	5.8	6.3	7.2	9.0	11.1	13.5	14.9	16.2	17.7	18.9	34.7	4.5	3.4	30.7	11.2	3.1	27.7	10.8	1.34	1.05	0.48	-0.43
	LOND	1.0	1.0	6.9	7.5	8.0	9.1	10.6	12.7	15.1	16.4	18.1	18.9	44.1	3.6	2.5	23.8	11.1	2.8	25.5	10.8	1.28	1.64	1.04	0.11
Ge mg/kg	LRD	0.5	<0.5	<0.5	0.5	0.7	1.0	1.4	2.1	3.1	4.1	5.9	7.6	38.7	1.1	0.7	53.0	1.8	1.6	91.4	1.4	1.92	2.11	7.17	0.06
	SEEN	0.5	<0.5	<0.5	<0.5	0.5	0.8	1.1	1.4	1.7	1.9	2.4	2.9	38.7	0.6	0.4	40.4	1.2	1.2	99.9	1.0	1.65	2.11	23.26	-0.26
	LOND	0.5	<0.5	<0.5	0.5	0.7	1.1	1.5	2.3	3.4	4.3	6.3	8.1	32.7	1.2	0.7	49.4	1.9	1.7	87.6	1.5	1.93	2.04	6.06	-0.03
Hf mg/kg	LRD	1.0	1.1	3.8	4.5	5.2	6.2	7.4	9.1	11.3	12.8	14.3	15.8	40.7	2.9	2.1	28.0	7.9	2.7	34.5	7.5	1.38	1.57	2.02	-0.03
	SEEN	1.0	1.2	3.7	4.7	5.9	7.6	9.5	11.7	13.5	14.5	16.5	18.7	31.2	4.1	3.1	32.8	9.7	3.2	33.5	9.1	1.43	1.41	0.94	-0.86
	LOND	1.0	1.1	3.8	4.5	5.1	6.0	7.2	8.6	10.1	11.5	13.4	14.6	40.7	2.6	1.9	26.8	7.5	2.4	32.1	7.2	1.34	1.57	2.58	0.03
I mg/kg	LRD	0.5	<0.5	1.2	1.5	1.8	2.4	3.1	4.2	6.6	8.8	12.3	16.7	79.9	1.8	1.2	38.3	3.9	3.4	86.5	3.3	1.72	2.30	7.12	0.65
	SEEN	0.5	<0.5	1.1	1.4	1.8	2.7	3.9	5.7	8.2	10.0	12.6	15.5	32.9	3.0	2.1	53.2	4.6	3.1	68.1	3.9	1.82	1.92	3.03	-0.12
	LOND	0.5	<0.5	1.2	1.5	1.8	2.4	3.0	3.9	5.8	8.2	12.2	16.8	79.9	1.5	1.0	34.6	3.7	3.4	91.2	3.2	1.68	2.30	8.02	0.87
Analytes units	Dataset	LLD	Min	Q02	Q05	Q10	Q25	Mdn	Q75	Q90	Q95	Q98	Q99	Max	IQR	MAD	CVR%	AM	SD	CV%	GM	GSD	Powers	Skew	Skew LOG
La mg/kg	LRD	1	3	12	14	16	20	25	30	36	40	49	59	134	10	7.4	29.7	25.9	10	37.0	24.4	1.41	1.65	2.41	-0.19
	SEEN	1	3	9	12	16	22	28	34	41	47	59	73	134	12	8.9	31.8	29.0	12	41.5	26.7	1.53	1.65	1.94	-0.80
	LOND	1	3	12	14	16	20	24	29	34	38	45	55	130	9	7.4	30.9	25.1	9	34.7	23.9	1.37	1.64	2.49	0.00

Mo mg/kg	LRD	0.2	<0.2	0.4	0.6	0.7	1.0	1.4	2.0	2.8	3.6	5.5	7.5	561.2	1.0	0.7	53.0	1.9	8.1	432.6	1.4	1.83	3.75	52.74	0.56
	SEEN	0.2	<0.2	0.3	0.4	0.5	0.7	0.9	1.2	1.5	1.8	2.6	3.8	192.2	0.5	0.3	32.9	1.1	4.8	424.7	0.9	1.70	3.28	38.41	0.48
	LOND	0.2	<0.2	0.5	0.7	0.8	1.1	1.5	2.1	3.0	3.8	5.9	7.8	561.2	1.0	0.7	49.4	2.0	8.7	424.7	1.6	1.77	3.75	51.32	0.79
Nb mg/kg	LRD	1.0	4.7	8.3	9.3	9.9	11.0	12.6	15.1	16.9	17.8	18.5	19.0	146.7	4.1	2.8	22.4	13.1	3.1	23.5	12.8	1.23	1.49	10.08	0.01
	SEEN	1.0	4.7	7.2	8.2	9.2	11.3	14.2	16.4	17.7	18.3	18.8	19.1	23.4	5.1	3.7	26.1	13.8	3.2	23.3	13.3	1.29	0.70	-0.35	-0.86
	LOND	1.0	5.3	8.8	9.4	10.0	11.0	12.5	14.6	16.6	17.5	18.4	18.8	146.7	3.6	2.5	20.2	12.9	3.0	23.4	12.6	1.22	1.44	13.26	0.36
Nd mg/kg	LRD	4.0	<4.0	8.0	10.4	12.9	17.0	21.8	26.9	32.5	36.7	45.5	54.8	172.7	9.9	7.3	33.3	22.7	9.7	42.5	21.0	1.49	1.69	3.05	-0.47
	SEEN	4.0	<4.0	5.8	8.4	11.4	17.8	24.2	30.7	36.7	41.7	55.1	71.2	172.7	12.9	9.6	39.8	25.1	12.6	50.3	22.3	1.66	1.69	3.08	-0.83
	LOND	4.0	<4.0	8.6	11.0	13.2	16.9	21.3	26.1	31.1	34.7	42.0	51.4	122.8	9.2	6.7	31.3	22.2	8.7	39.4	20.7	1.45	1.55	2.70	-0.34
Ni mg/kg	LRD	1.3	<1.3	8.0	11.3	13.9	18.7	24.6	32.2	41.3	49.6	62.3	78.3	505.6	13.5	9.8	39.8	27.2	16.8	61.8	24.2	1.62	2.62	8.77	-0.40
	SEEN	1.3	<1.3	3.7	6.6	9.6	15.2	21.1	30.0	38.7	45.6	56.8	74.6	469.4	14.8	10.4	49.2	24.1	20.0	82.7	20.1	1.86	2.59	10.41	-0.73
	LOND	1.3	2.3	10.0	12.7	15.1	19.5	25.4	32.6	41.8	50.2	63.7	79.0	505.6	13.1	9.3	36.8	27.9	15.9	57.0	25.3	1.54	2.34	8.05	0.16
Analytes units	Dataset	LLD	Min	Q02	Q05	Q10	Q25	Mdn	Q75	Q90	Q95	Q98	Q99	Max	IQR	MAD	CVR%	AM	SD	CV%	GM	GSD	Powers	Skew	Skew LOG
Pb mg/kg	LRD	1.3	10	32	38	46	70	138	284	531	775	1222	1668	25206	214	122	88.	251	477	190.3	149	2.59	3.40	22.98	0.46
	SEEN	1.3	10	25	30	34	40	55	81	126	184	325	484	1914	41	25	45.8	79	105	133.4	61	1.82	2.28	8.95	1.39
	LOND	1.3	11	39	49	60	94	175	332	593	845	1323	1864	25206	238	145	83.0	291	520	178.4	184	2.45	3.36	21.74	0.42
Rb mg/kg	LRD	1.0	6.5	25.5	33.1	38.4	45.9	56.5	72.5	89.5	98.0	107.4	113.5	157.2	26.6	18.7	33.1	60.3	20.1	33.4	57.0	1.41	1.38	0.64	-0.50
	SEEN	1.0	6.5	21.1	27.3	35.4	47.2	62.8	80.0	95.7	103.0	111.0	114.1	157.2	32.8	24.2	38.5	63.8	23.0	36.1	59.1	1.52	1.38	0.21	-1.05
	LOND	1.0	8.9	28.4	34.4	38.9	45.8	55.8	70.8	87.7	96.4	106.5	112.9	148.6	25.0	17.3	31.1	59.5	19.3	32.4	56.5	1.39	1.22	0.77	-0.26
Sb mg/kg	LRD	0.5	<0.5	0.5	0.7	0.9	1.3	2.4	4.5	8.2	12.7	22.0	33.1	612.3	3.2	1.9	80.3	4.5	13.2	294.8	2.6	2.48	3.18	26.04	0.67
	SEEN	0.5	<0.5	<0.5	0.5	0.6	0.8	1.0	1.4	2.2	3.1	5.7	9.9	47.1	0.6	0.4	44.5	1.4	2.1	148.4	1.1	1.81	2.07	11.40	1.42
	LOND	0.5	<0.5	0.7	1.0	1.2	1.7	2.9	5.2	9.3	14.2	24.9	38.2	612.3	3.5	2.1	71.6	5.2	14.6	279.9	3.2	2.34	3.18	23.91	0.76
Sc mg/kg	LRD	3.0	<3.0	<3.0	3.0	4.0	5.8	7.8	10.2	12.6	14.0	15.7	16.9	35.6	4.4	3.3	41.8	8.1	3.4	41.5	7.4	1.60	1.25	0.61	-0.83
	SEEN	3.0	<3.0	<3.0	<3.0	<3.0	4.6	7.3	10.0	12.3	13.7	15.6	17.3	35.6	5.4	4.0	54.8	7.5	3.8	50.6	6.4	1.81	1.25	0.65	-0.65
	LOND	3.0	<3.0	<3.0	3.5	4.5	6.0	7.9	10.3	12.6	14.1	15.7	16.8	33.3	4.3	3.1	39.4	8.3	3.2	39.2	7.6	1.54	1.22	0.66	-0.73
Se mg/kg	LRD	0.2	<0.2	0.2	0.2	0.3	0.4	0.5	0.7	1.0	1.2	1.7	2.3	19.6	0.3	0.3	59.3	0.6	0.6	92.6	0.5	1.72	2.29	14.13	0.09
	SEEN	0.2	<0.2	<0.2	0.2	0.2	0.3	0.4	0.6	0.7	0.9	1.3	1.7	15.4	0.3	0.1	37.1	0.5	0.5	101.4	0.4	1.69	2.19	18.10	0.14
	LOND	0.2	<0.2	0.2	0.2	0.3	0.4	0.6	0.8	1.0	1.3	1.8	2.3	19.6	0.4	0.3	49.4	0.7	0.6	90.1	0.6	1.71	2.29	13.75	0.09
Analytes units	Dataset	LLD	Min	Q02	Q05	Q10	Q25	Mdn	Q75	Q90	Q95	Q98	Q99	Max	IQR	MAD	CVR%	AM	SD	CV%	GM	GSD	Powers	Skew	Skew LOG
Sn mg/kg	LRD	0.5	0.9	2.6	3.1	3.7	5.6	10.9	22.6	44.2	67.2	124.4	169.1	1041.5	17.0	9.6	88.4	21.5	40.0	186.2	12.1	2.64	3.06	9.24	0.62
	SEEN	0.5	1.2	2.2	2.5	2.7	3.4	4.5	7.3	11.9	18.9	38.1	71.8	466.5	3.9	2.2	49.4	8.1	19.7	243.5	5.3	2.00	2.59	13.61	1.80
	LOND	0.5	0.9	3.2	4.0	4.7	7.3	13.6	25.7	50.0	75.9	133.5	181.1	1041.5	18.4	11.4	83.9	24.6	42.8	173.9	14.6	2.52	3.06	8.83	0.59
Sr mg/kg	LRD	1	11	29	38	45	57	73	93	125	153	217	272	601	36	25.2	34.5	82.2	44.4	54.0	74.2	1.55	1.74	3.24	0.34
	SEEN	1	11	24	29	35	48	63	80	115	171	277	330	576	32	22.2	35.3	75.1	54.9	73	64.3	1.68	1.72	3.53	0.71
	LOND	1	12	33	41	48	60	76	96	126	152	198	254	601	36	25.2	33.2	83.9	41.4	49.3	76.8	1.50	1.70	3.09	0.34
Th mg/kg	LRD	0.7	<0.7	3.1	3.8	4.4	5.4	6.8	8.4	9.5	10.1	10.6	11.0	456.5	3.0	2.2	32.7	7.0	5.3	76.1	6.6	1.38	2.88	73.02	-0.69
	SEEN	0.7	<0.7	2.2	3.1	3.9	5.7	7.7	9.1	10.0	10.4	10.8	11.0	15.7	3.4	2.4	30.8	7.3	2.3	31.6	6.8	1.49	1.42	-0.49	-1.49
	LOND	0.7	<0.7	3.3	4.0	4.5	5.4	6.6	8.2	9.4	9.9	10.6	11.0	456.5	2.8	2.1	31.4	6.9	5.8	83.9	6.5	1.35	2.88	69.56	-0.29
U mg/kg	LRD	0.5	<0.5	1.1	1.4	1.6	1.9	2.2	2.6	2.9	3.1	3.4	3.6	11.5	0.7	0.5	24.7	2.2	0.6	26.1	2.2	1.34	1.46	0.88	-1.55
	SEEN	0.5	<0.5	1.1	1.3	1.5	1.9	2.3	2.6	2.9	3.1	3.2	3.4	7.3	0.7	0.5	23.6	2.2	0.6	24.9	2.2	1.31	1.26	0.43	-1.00
	LOND	0.5	<0.5	1.1	1.4	1.6	1.9	2.2	2.6	2.9	3.2	3.4	3.6	11.5	0.7	0.5	24.2	2.2	0.6	26.4	2.1	1.34	1.46	0.96	-1.65
V mg/kg	LRD	3	9	36	44	50	61	75	98	122	135	151	163	531	37	25.2	33.6	81.4	29.7	36.4	76.5	1.43	1.77	1.41	-0.24
	SEEN	3	9	25	33	41	54	71	91	113	128	146	161	352	37	26.7	37.6	74.5	30.1	40.4	68.6	1.52	1.59	1.24	-0.59
	LOND	3	16	40	47	53	62	76	100	124	137	151	163	531	38	25.2	33.2	83.1	29.3	35.3	78.4	1.40	1.52	1.50	0.04
Analytes units	Dataset	LLD	Min	Q02	Q05	Q10	Q25	Mdn	Q75	Q90	Q95	Q98	Q99	Max	IQR	MAD	CVR%	AM	SD	CV%	GM	GSD	Powers	Skew	Skew LOG
W mg/kg	LRD	0.6	<0.6	<0.6	0.8	1.1	1.6	2.1	2.6	3.3	3.9	5.6	8.4	316.8	1.0	0.7	35.3	2.4	5.8	241.8	2.0	1.70	2.90	37.19	0.21
	SEEN	0.6	<0.6	<0.6	<0.6	0.8	1.6	2.4	3.0	3.5	4.0	4.8	5.5	184.6	1.4	1.0	43.2	2.5	4.8	193.1	2.0	1.83	2.66	35.32	-0.70
	LOND	0.6	<0.6	<0.6	0.9	1.1	1.6	2.0	2.5	3.2	3.9	6.2	8.7	316.8	0.9	0.7	37.1	2.4	6.1	252.4	2.0	1.67	2.90	36.98	0.55
Y mg/kg	LRD	1	3	9	12	14	17	21	25	29	32	42	50	158	8.0	5.9	28.2	21.6	8.5	39.3	20.3	1.40	1.72	4.16	0.01
	SEEN	1	3	7	9	12	18	23	28	32	39	51	69	158	10.5	7.4	32.2	23.5	11.3	48.1	21.4	1.55	1.72	3.47	-0.41
	LOND	1	5	10	12	14	17	20	24	28	30	37	48	133	7.0	5.9	29.7	21.1	7.6	35.9	20.1	1.35	1.42	4.24	0.18
Zn mg/kg	LRD	1.3	<1.3	27	44	58	81	130	225	371	521	801	1120	10095	144	89	68.4	195	269	138.0	137	2.21	4.00	13.90	0.14
	SEEN	1.3	<1.3	18	28	42	60	77	98	136	176	254	446	2505	38	27	34.7	94	120	127.8	76	1.84	3.40	11.83	-0.42
	LOND	1.3	<1.3	35	51	65	95	152	252	412	563	842	1214	10095	157	102	67.3	219	288	131.7	158	2.15	4.00	13.57	0.15
Zr mg/kg	LRD	1	35	153	182	204	240	288	350	424	478	536	580	1488	110	79	27.3	304	99	32.4	290	1.36	1.63	2.02	-0.14
	SEEN	1	48	137	182	219	286	351	434	504	537	596	646	1062	148	107	30.4	360	115	32.1	340	1.43	1.34	0.60	-1.14
	LOND	1	35	156	182	202	235	277	330	391	437	496	552	1488	95	70	25.2	291	89	30.7	279	1.32	1.63	2.71	0.05
LOI %	LRD	0.01	0.7	3.2	3.9	4.5	5.7	7.1	8.9	10.9	12.5	15.5	18.3	72.1	3.2	2.4	33.4	7.6	3.3	43.8	7.1	1.45	2.01	4.27	0.10
	SEEN	0.01	1.2	2.8	3.3	3.8	4.8	6.0	8.0	10.6	12.8	15.9	18.7	72.1	3.2	2.1	34.6	6.9	4.3	61.7	6.2	1.53	1.78	6.31	0.72
	LOND	0.01	0.7	3.3	4.0	4.7	5.8	7.3	9.1	11.0	12.5	15.5	18.3	58.7	3.3	2.4	32.5	7.8	3.2	40.6	7.3	1.43	1.92	3.46	0.02
pH log[H+] mol/L	LRD	0.01	2.8	3.4	4.0	4.6	5.5	6.5	7.0	7.3	7.4	7.5	7.6	8.4	1.5	0.9	14.1	6.2	1.1	17.3	6.1	1.22	0.48	-0.96	-1.38
	SEEN	0.01	2.9	3.2	3.6	4.3	5.0	6.2	7.1	7.4	7.5	7.5	7.6	7.8	2.1	1.5	24.5	5.9	1.2	20.6	5.8	1.26	0.44	-0.53	-0.95
	LOND	0.01	2.8	3.5	4.1	4.7	5.6	6.6	7.0	7.3	7.4	7.5	7.6	8.4	1.4	0.8	12.9	6.2	1.0	16.6	6.1	1.21	0.48	-1.03	-1.45

Following **Table 9**, a set of box-plots of the 44 elements are shown in **Figure 7**, **Figure 8**, and **Figure 9** for **LRD**, **SEEN** and **LOND** respectively. These were generated running R (Rx64 3.1.0) in RStudio (Version 0.98.977) with an R script derived mainly from functions and scripts in StatDA R package (Reimann et al., 2008[5]). The elements are ordered by abundance in the **SEEN** subset. Each figure is composed of two graphics, one (on the left) showing the 10 most abundant elements (expressed as oxides in wt%) and another (on the right) showing the less abundant elements (in mg/kg); Mn is included in both graphics allowing a visual calibration of the concentration scale (Y axis). The concentration scales of these three figures are log-decimal, thus allowing visualisation of elements with orders of magnitude difference in content on the same graphic. This also has advantage when showing elements with a highly positive skewed distribution.

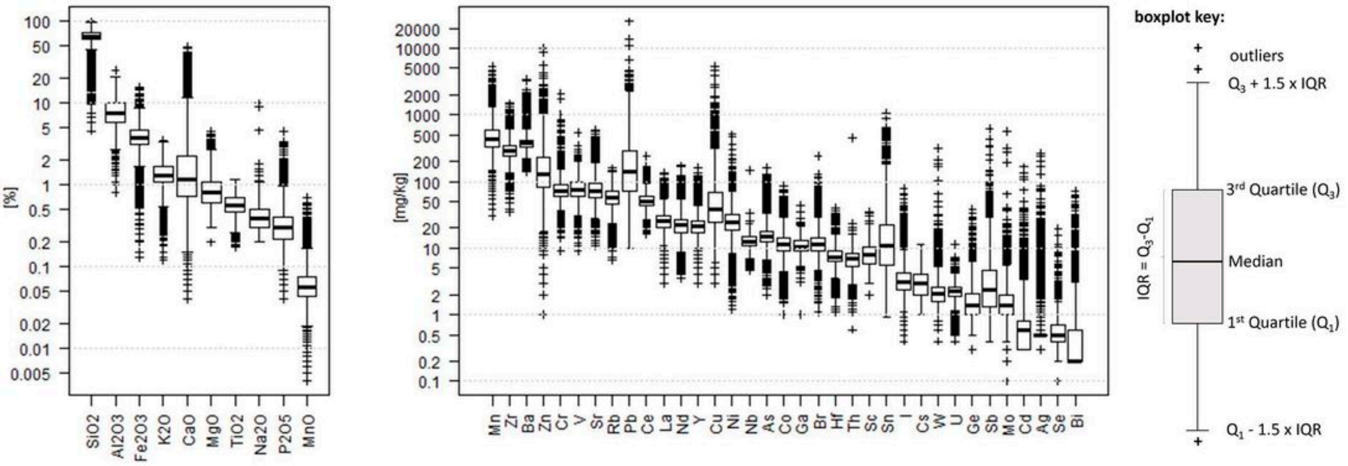

Figure 7 Box-plot of element concentrations in topsoil in the **LRD** dataset (8400 samples), shown by order of abundance in the **SEEN** subset. The 10 most abundant elements (expressed as oxides in wt%) on the left and the less abundant (in mg/kg) on the right. Manganese is shown as MnO and as Mn for reference between both graphs. (P929864).

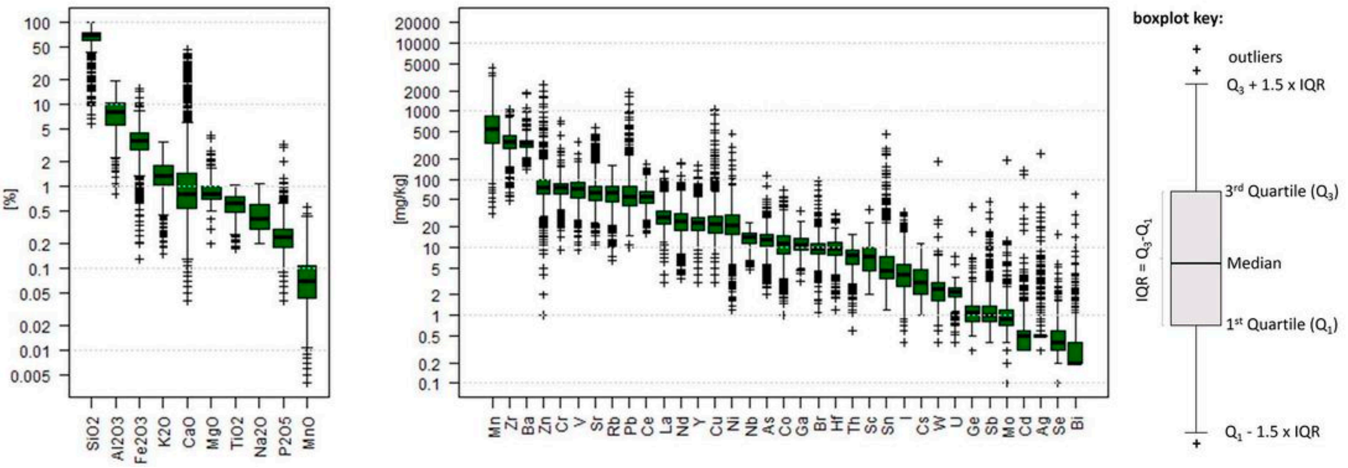

Figure 8 Box-plot of element concentrations in topsoil in the **SEEN** subset (1599 samples), shown by order of abundance in the **SEEN** subset. The 10 most abundant elements (expressed as oxides in wt%) on the left and the less abundant (in mg/kg) on the right. Manganese is shown as MnO and as Mn for reference between both graphs. (P929865).

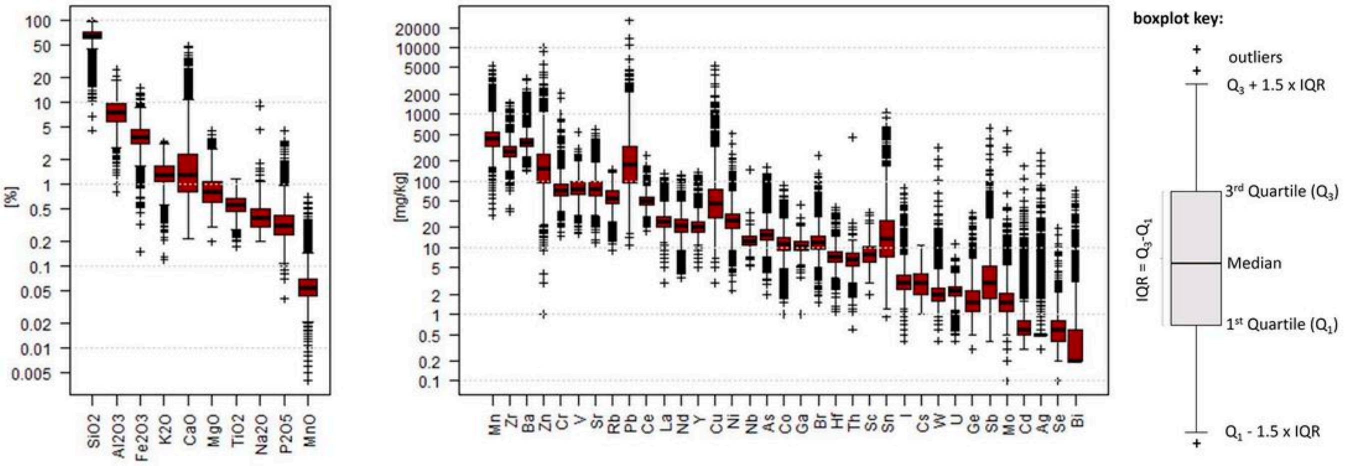

Figure 9 Box-plot of element concentrations in topsoil in the **LOND** subset (6801 samples), shown by order of abundance in the **SEEN** subset. The 10 most abundant elements (expressed as oxides in wt%) on the left and the less abundant (in mg/kg) on the right. Manganese is shown as MnO and as Mn for reference between both graphs. (P929866).

Adjacent to each element map, a univariate table of statistics (e.g.: **Figure 10**) and a set of statistical graphics are shown. These were also generated running R (Rx64 3.1.0) in RStudio (Version 0.98.977) using an R script derived from functions and scripts from StatDA[6] R package (Reimann et al., 2008[5]).

Where appropriate the **LRD** dataset, and **SEEN** and **LOND** subsets are represented in **black/white**, **dark green** and **dark red** respectively. **Quaternary**, **Palaeogene** and **Cretaceous** geological time periods are represented by **soft yellow**, **orange** and **lime green** colours respectively.

Graphics shown are (i) a histogram of **LRD** data distribution with the density lines of the **SEEN** and **LOND** subsets; (ii) a Tukey boxplot of the **LRD**, **SEEN** and **LOND** dataset distributions and (iii) a cumulative probability plot of the **LRD**, **SEEN** and **LOND** datasets (e.g.: **Figure 11**); (iv) the **LOND** and **SEEN** topsoil parameter concentrations over each geological unit (e.g.: **Figure 12**); (v) the **LRD** topsoil parameter concentrations over each geological unit (e.g.: **Figure 13** and **Figure 14**).

The **histogram**, is used to show the **LRD** topsoil parameter distribution. The **SEEN** and **LOND** subsets are shown as **density traces**, as these allow overlap of the three data distributions (**LRD**, **SEEN**, **LOND**), thus directly comparing their shapes, independently of the number of soil samples. The **Tukey boxplots** (Tukey, 1977[7]) are built around the median (line dividing the box in two parts); the box representing the interquartile range (Q_3-Q_1=IQR), that is, 50% of the data, from the lower (Q_1) to the upper (Q_3) quartiles; the lowest end of the whiskers is 1.5 times the length of the box (IQR) starting from the lower quartile, while the highest one is 1.5 times the IQR starting from the upper quartile; in this graphical representation, data below and above the whiskers (plotted with the symbol **+**) are considered as outliers. The boxplots shown also include a notch around the median. This gives an approximate indication as to whether the medians are different or not. If notches of different boxplots do not overlap, there is some evidence of a statistical difference between the medians. The **cumulative probability (CP) plots** (Sinclair, 1976[8]) are built with a non-linear probability scale on the Y axis (from >0 to <100%). The quartiles are close together near the median (probability = 50%) and stretch out symmetrically moving away from the median. If the variable follows a normal (or lognormal) distribution, then its values (or log transformed values) fall in a straight line from the bottom left to the top right of the graph; inflexion points on the curve suggest that the distribution is made up of multiple data populations. CP plots are a good complement to the previous graphs in understanding and comparing the distribution of several datasets.

The Tukey style of boxplots used to display the **LRD**, **SEEN** and **LOND** datasets were used to show element concentrations over each geological parent material unit also. In the **LOND** and **SEEN** comparison, two boxplots (**LOND** above **SEEN**) are shown, together with the number of soil samples collected over each geological unit (n, on the right of the graphic) (**Figure 12**). For some parent material classes (e.g.: **LOND** over Gault Fm.), the boxplot is not shown as there are no soil samples over this rock type (**Table 7**). The **LRD** geological parent material comparison is presented in two formats. The upper set of boxplots show topsoil element concentrations over each geological unit (**Figure 13**), whereas the lower set of boxplots show topsoil element concentrations classified by parent material geological period (**Figure 14**). The number of soil samples in each geological time zone is shown in brackets and the boxplots are coloured according to the geological time period. The last boxplot refers to **LRD** and is shown for reference.

Al2O3 [%] **statistical parameters**

	LRD	SEEN	LOND
N	8400	1599	6801
%<LLD (0.2)	0	0	0
minimum	0.8	0.8	0.8
Q25	5.9	5.8	5.9
median	**7.6**	**8.2**	**7.5**
Q75	10	10.7	9.7
Q95	13.2	13.4	13.2
Q99	15.3	15.6	15.2
maximum	25.5	19.9	25.5
MAD	**2.8**	**3.7**	**2.7**
CVR%	37.1	45.2	35.6
mean	8	8.2	8
SD	2.9	3.3	2.8
CV%	35.9	39.7	34.9
GM	7.5	7.4	7.5
GSD	1.47	1.62	1.43
powers	1.5	1.4	1.5
skewness	0.52	0.09	0.66

Figure 10 Example of a univariate table of statistics for **LRD**, **SEEN** and **LOND** shown next to the respective map. The case for Al_2O_3. These and other statistical parameters are shown in **Table 9** for other elements also. For further explanation go to Univariate statistics and graphics (P929875).

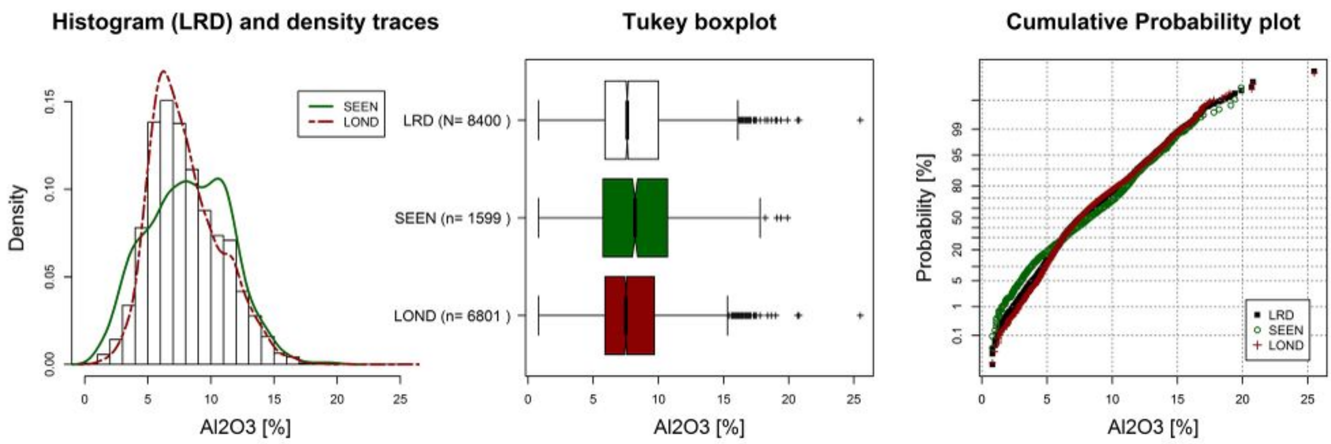

Figure 11 Example of a histogram, a Tukey boxplot and a cumulative probability plot shown for **LRD**, **SEEN** and **LOND** next to the respective map. The case for Al_2O_3. For further explanation go to Here (P929874).

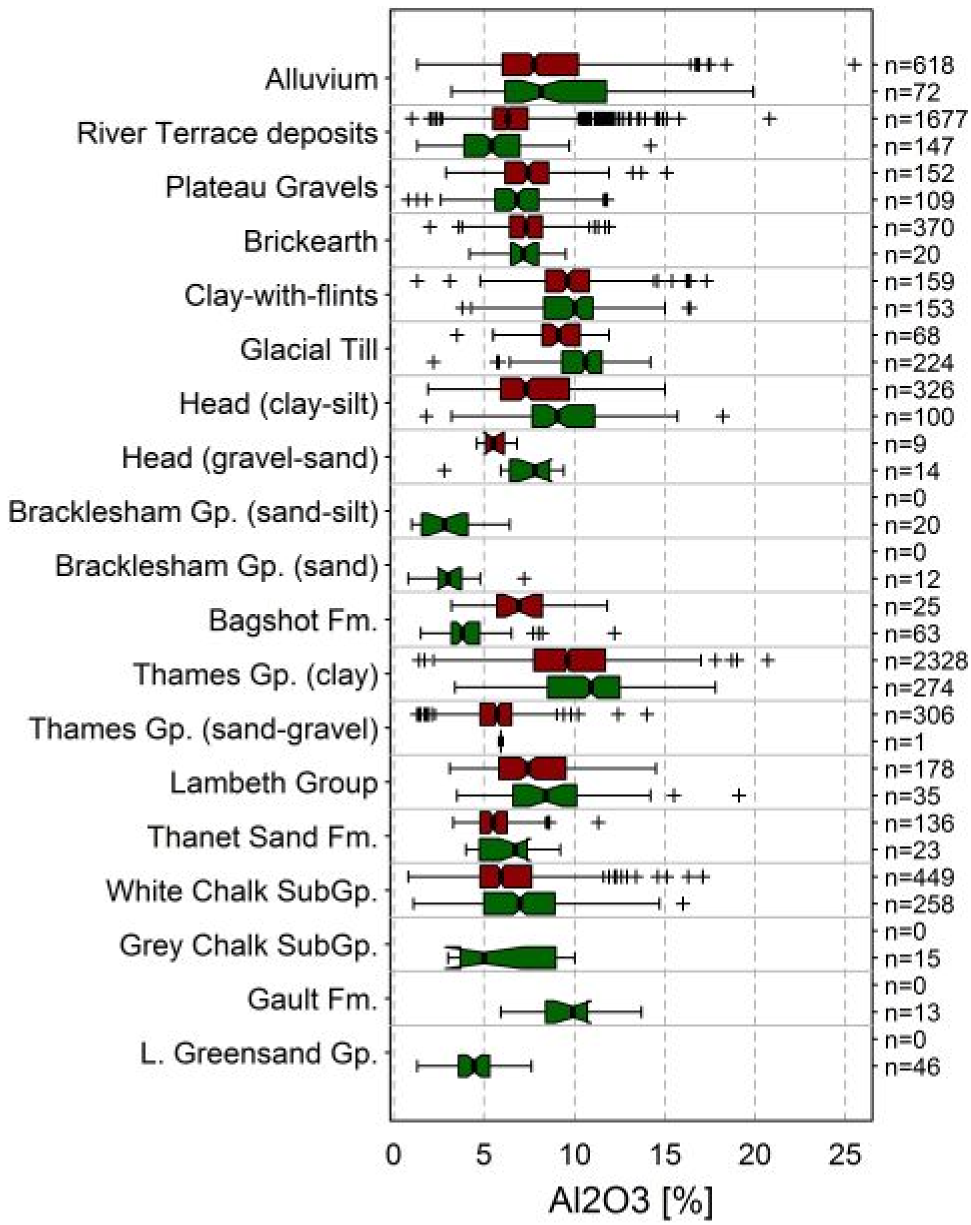

Figure 12 Example of a **LOND** and **SEEN** topsoil parameter concentrations over each geological unit shown next to the respective

map. The case for Al_2O_3. For further explanation go to Here (P929876).

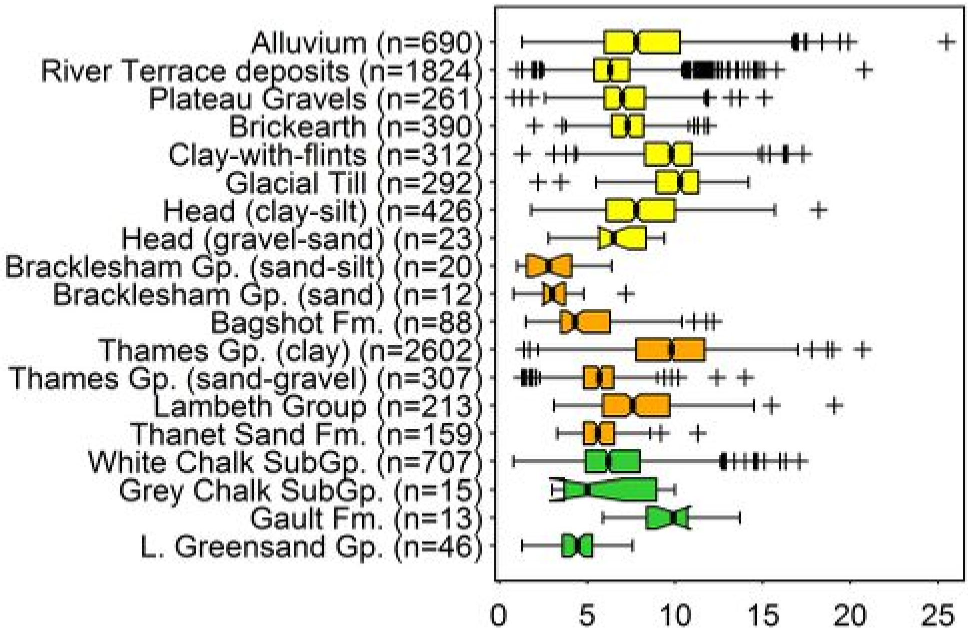

Figure 13 Example of the **LRD** topsoil parameter concentrations over each geological unit (parent material) shown next to the respective map. The case for Al_2O_3. For further explanation go to Here (P929877).

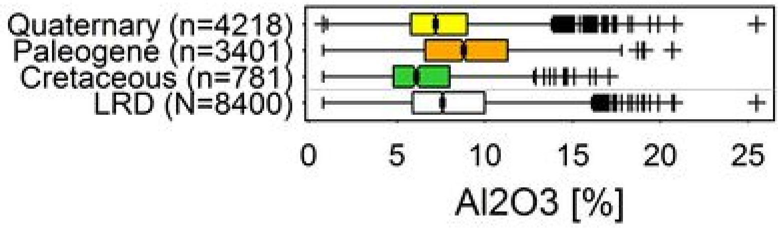

Figure 14 Example of the **LRD** topsoil parameter concentrations over each geological time period shown next to the respective map. The case for Al_2O_3. For further explanation go to Here (P929878).

Compositional data analysis (CoDA)

Geochemical data have an intrinsic compositional nature (Aitchison, 1986[9]), as the sum of all parts in a sample necessarily sum up to a constant (frequently 1, 100% or 1 000 000 mg/kg). These datasets are called *compositional data* or *closed data*. For closed data the concentration obtained for one part (element) does not vary independently from the others, thus the information is not absolute but only relative. As pointed out by Pearson (1897)[10], applying classical methods of statistical data analysis to this type of dataset may lead to wrong results, as for example, in the form of spurious correlations when performing a bivariate analysis. This problem was first tackled by Aitchison (1986)[9], who found that compositional data is better encompassed, not in the usual Euclidean space (on which classical statistical methods have been developed), but in the Aitchison geometry on the simplex. As the statistical evaluation and interpretation of a certain element concentration should take into account the remaining parts, the statistical methods and techniques developed in this context are mainly based on three different logratio transformations. The *alr* - additive logratio and the *clr* – centred logratio transformation, both proposed by Aitchison (1986)[9], and the *ilr* – isometric logratio transformation (Egozcue et al., 2003[11]). The idea of these transformations is that the data should be transformed into the correct geometry first, after which classical methods can be fully applied.

However these transformations lead to dimensionless values, which constitute a serious drawback when the main task is to document and to study the spatial distribution of element concentrations (Reimann et al., 2012[12]). This is particularly evident in the production of geochemical atlases where the primary aim is to present single element distribution maps and associated statistics (Reimann et al., 2014[13]).

As this work is a geochemical atlas, the uni-element maps and associated statistics are shown as prevoiusly explained, as the absolute values are of interest. However, readers, especially earth science practitioners, are advised to take into account that the geochemical dataset presented here is clearly a case of compositional data. Thus a conflict with the CoDA approach may exist, particularly for the statistics typically used to compare variables such as the measures of dispersion, despite of some extenuatory reasoning as referred by Filzmoser et al. (2009)[14] and Reimann et al. (2014)[13].

As pointed out above, uni-element and bivariate statistics are not among the best practices when statistically analysing compositions, because these datasets are multivariate in nature. Still, the *center* or *compositional mean* (equation 4.1 in van den Boogaart and Tolosana-Delgado, 2013[15]) is meaningful in the CoDA context, namely with respect to translation and scaling operations, referred as perturbation and powering in CoDA (van den Boogaart and Tolosana-Delgado, 2013[15]). This central tendency CoDA statistic is computed for the **LRD** dataset and **SEEN** and **LOND** subsets after the line command *mean(x)*, with *x* being a composition, in *compositions*[16] R package (**Table 10**); the composition is based on all the 44 chemical elements and in all the topsoil samples of this atlas (8400 for **LRD**, 1599 for **SEEN** and 6801 for **LOND**). The *compositional mean* was computed after converting the trace elements to percent (%, i.e.: concentration in mg/kg / 10000); the obtained values were then back transformed to the original units (% for the 10 major and mg/kg for the 34 less abundant elements).

By computing the *compositional mean* in **LOND** and in **SEEN**, a preliminary overview of the geochemical signatures in urban versus rural areas can be obtained. The comparison of **LOND** and **SEEN** is better achieved after the (Log(**LOND**/**SEEN**) in **Table 10**, which gives an indication of element enrichment (positive) / depletion (negative) in the urban environment relative to the rural environment, and the degree (distance to zero) of this enrichment / depletion; the elements are ordered from the highest to the lowest Log(**LOND**/**SEEN**) value in order to facilitate comparison of the elements. Pb (closely followed by Sb, Sn, Cu, Zn) is the element showing the highest relative enrichment in **LOND**, contrary to Hf (closely followed by I, Zr, MnO) which shows the highest relative depletion in **LOND**; in this sense the major Al_2O_3 and K_2O are the elements showing the lowest enrichment or depletion in **LOND** relative to **SEEN** as their Log(**LOND**/**SEEN**) values are the closest to zero among all elements. This sort of approach, i.e., the use of ratios to estimate enrichment / depletion (enrichment factors, EFs), has been used previously to examine urban data relative to rural data (e.g.: Fordyce et al., 2005[17]; Flight and Scheib, 2011[18], among many others). This and other indexes, such as the I_{geo}-index (Müller, G, 1979[19]), have been widely used in searching for anthropogenic impact on the surface environment since their introduction in the 1970s (e.g.: Chester and Stoner, 1973[20]; Zoller et al., 1974[21]).

Table 10 Compositional mean based on the concentration of 44 chemical elements in 8400, 1599 and 6801 topsoil samples respectively for LRD, SEEN and LOND. Log(LOND/SEEN) corresponds to the logarithm of column 6 (ratio between the LOND and the SEEN compositional mean values).

Elements are ordered from the highest to lowest log(LOND/SEEN) value.

Chemical element	Units	LRD	SEEN	LOND	LOND/SEEN	Log(LOND/SEEN)
Pb	mg/kg	183.9	75.4	226.7	3.01	0.478
Sb	mg/kg	3.20	1.37	3.90	2.86	0.456
Sn	mg/kg	14.92	6.59	18.08	2.74	0.438
Cu	mg/kg	51.50	27.35	59.74	2.18	0.339
Zn	mg/kg	169.6	93.8	194.8	2.08	0.318
Mo	mg/kg	1.73	1.10	1.93	1.75	0.243
Ge	mg/kg	1.75	1.25	1.89	1.51	0.179
CaO	%	1.66	1.23	1.78	1.44	0.160
Cd	mg/kg	0.75	0.57	0.80	1.39	0.143
P_2O_5	%	0.37	0.29	0.40	1.35	0.131
Se	mg/kg	0.67	0.53	0.70	1.33	0.125
As	mg/kg	18.62	15.11	19.55	1.29	0.112
Br	mg/kg	14.35	11.91	14.99	1.26	0.100
Bi	mg/kg	0.46	0.38	0.48	1.25	0.098
Ni	mg/kg	29.97	25.00	31.26	1.25	0.097
Sr	mg/kg	91.70	79.72	94.73	1.19	0.075
Sc	mg/kg	9.10	7.95	9.39	1.18	0.072
Ba	mg/kg	462.6	409.2	476.0	1.16	0.066
V	mg/kg	94.47	85.04	96.80	1.14	0.056
Fe_2O_3	%	4.59	4.20	4.68	1.11	0.047
Co	mg/kg	13.71	12.86	13.91	1.08	0.034
Ag	mg/kg	0.70	0.67	0.71	1.06	0.027
Cr	mg/kg	90.53	89.00	90.86	1.02	0.009
Na_2O	%	0.51	0.50	0.51	1.02	0.006
Al_2O_3	%	9.26	9.23	9.26	1.00	0.002
K_2O	%	1.642	1.640	1.642	1.00	0.001
U	mg/kg	2.66	2.67	2.65	1.00	-0.002
Ga	mg/kg	13.31	13.37	13.29	0.99	-0.002
SiO_2	%	80.1	80.9	79.8	0.99	-0.006
W	mg/kg	2.45	2.51	2.43	0.97	-0.014
MgO	%	0.96	0.98	0.95	0.97	-0.015
Th	mg/kg	8.14	8.45	8.07	0.95	-0.020
Rb	mg/kg	70.48	73.30	69.81	0.95	-0.021
TiO_2	%	0.71	0.74	0.70	0.95	-0.023
Nb	mg/kg	15.77	16.53	15.59	0.94	-0.026
Y	mg/kg	25.14	26.50	24.83	0.94	-0.028
Ce	mg/kg	61.91	65.50	61.07	0.93	-0.030
Nd	mg/kg	26.06	27.78	25.67	0.92	-0.034
La	mg/kg	30.15	33.11	29.48	0.89	-0.050
Cs	mg/kg	3.55	3.98	3.45	0.87	-0.062
MnO	%	0.069	0.081	0.067	0.83	-0.080

Zr	mg/kg	358.1	421.2	344.5	0.82	-0.087
I	mg/kg	4.06	4.79	3.90	0.81	-0.089
Hf	mg/kg	9.34	11.39	8.91	0.78	-0.107

References

1. ↑ Appleton, J D, Johnson, C C, Ander, E L, and Flight, D M A. 2013. Geogenic signatures detectable in topsoils of urban and rural domains in the London region, UK, using parent material classified data. Applied Geochemistry, Vol. 39, 169–180. DOI 10.1016/j.apgeochem.2013.07.010
2. ↑ Ander, E L, Cave, M R, Johnson, C C, and Palumbo-Roe, B. 2011. Normal background concentrations of contaminants in the soils of England. Available data and data exploration. British Geological Survey Commissioned Report, CR/11/145.
3. ↑ Ander, E L, Johnson, C C, Cave, M R, Palumbo-Roe, B, Nathanail, C P, and Lark, R M. 2013. Methodology for the determination of normal background concentrations of contaminants in English soil. Science of The Total Environment, Vol. 454-455, 604-618. 10.1016/j.scitotenv.2013.03.005
4. ↑ Communities and Local Government. 2007. Generalised Land Use Database Statistics for England 2005. Department for Communities and Local Government. Product Code 06CSRG04342. February 2007
5. ↑ [5.0] [5.1] [5.2] [5.3] Reimann, C, Filzmoser, P, Garrett, R G, and Dutter, R. 2008. Statistical data analysis explained. Applied environmental statistics with R. John Wiley & Sons Ltd., Chichester, England. http://www.statistik.tuwien.ac.at/StatDA/R-scripts/
6. ↑ StatDA: Statistical Analysis for Environmental Data https://cran.r-project.org/web/packages/StatDA/index.html
7. ↑ Tukey, J W. 1977. Exploratory data analysis. Addison-Wesley, Reading, Massachusetts, USA. 506pp.
8. ↑ Sinclair, A J. 1976. Applications of probability graphs in mineral exploration. Special Volume 4. Association of Exploration Geochemists. Toronto, Canada. 95pp.
9. ↑ [9.0] [9.1] [9.2] Aitchison, J. 1986. The statistical analysis of compositional data. Monographs on statistics and applied probability. London: Chapman & Hall (Reprinted in 2003 with additional material by The Blackburn Press), 416 pp.
10. ↑ Pearson, K. 1897. Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proceedings of the Royal Society of London, Vol. LX, 489–502.
11. ↑ Egozcue, J J, Pawlowsky-Glahn, V, Mateu-Figueras, G, and Barcelo-Vidal, C. 2003. Isometric logratio transformations for compositional data analysis. Mathematical Geology, Vol. 35(3), 279–300.
12. ↑ Reimann, C, Filzmoser, P, Fabian, K, Hron, K, Birke, M, Demetriades, A, Dinelli, E, Ladenberger, A, and the GEMAS Project Team. 2012. The concept of compositional data analysis in practice — total major element concentrations in agricultural and grazing land soils of Europe. Science of the Total Environment, Vol. 426, 196–210
13. ↑ [13.0] [13.1] Reimann, C, Birke, M, Demetriades, A, Filzmoser, P, and O'Connor, P (eds.). 2014. Chemistry of Europe's agricultural soils. Part A: methodology and interpretation of the GEMAS dataset. Geologisches Jahrbuch, B 102, 528 pp., 358 figs., 86 Tables, 1 DVD; Hannover, Germany.
14. ↑ Filzmoser, P, Hron, K, and Reimann, C. 2009. Univariate statistical analysis of environmental (compositional) data: problems and possibilities. Science of The Total Environment, Vol. 407, Issue 23, 6100–6108. DOI 10.1016/j.scitotenv.2009.08.008

15. ↑ [15.0] [15.1] van den Boogaart, K G, and Tolosana-Delgado, R. 2013. Analyzing compositional data with R. Use R! Springer-Verlag Berlin Heidelberg. DOI 10.1007/978-3-642-36809-7 1
16. ↑ van den Boogaart, K. G. & Tolosana-Delgado, R. (2008). Compositions: a unified R package to analyze compositional data. Computers and Geosciences, 34(4), 320–338.
17. ↑ Fordyce, F M, Brown, S E, Ander, E L, Rawlins, B G, O'Donnell, K E, Lister, T R, Breward, N, and Johnson, C C. 2005. GSUE: urban geochemical mapping in Great Britain. Geochemistry: Exploration, Environment, Analysis. Vol. 5(4), 325-336 (12).
18. ↑ Flight, D M A, and Scheib, A J, 2011. Soil geochemical baselines in UK urban centres: the G-BASE project. In: Johnson, C, Demetriades, A, Locutura, J, and Ottesen, R T (eds.). Mapping the chemical environment of urban areas. Wiley-Blackwell, Oxford, pp. 186-206.
19. ↑ Müller, G, 1979. Schwermetalle in den Sedimenten des Rheins - Veränderungen seit 1971. Umschau 79, 778–783.
20. ↑ Chester, R, and Stoner, J H. 1973. Pb in particulates from the lower atmosphere of the eastern Atlantic. Nature 245, 27–28.
21. ↑ Zoller, W H, Gladney, E S, and Duce, R A. 1974. Atmospheric concentrations and sources of trace metals at the South Pole. Science, Vol. 183(4121), 198–200

Navigation menu

Personal tools

- Not logged in
- Talk
- Contributions
- Log in
- Request account

Namespaces

- Page
- Discussion

☐

Variants

Views

- Read
- Edit
- View history
- PDF Export