Applied Computing and Geosciences 6 (2020) 100024

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.journals.elsevier.com/applied-computing-and-geosciences

Investigating the influence of environmental factors on the incidence of renal disease with compositional data analysis using balances

Jennifer M. McKinley [a,*], Ute Mueller [b], Peter M. Atkinson [a,c], Ulrich Ofterdinger [a], Chloe Jackson [a], Siobhan F. Cox [a], Rory Doherty [a], Damian Fogarty [d], J.J. Egozcue [e], V. Pawlowsky-Glahn [f]

[a] *School of Natural and Built Environment, Queen's University Belfast, Northern Ireland, UK*
[b] *School of Science, Edith Cowan University, Perth, Western Australia, Australia*
[c] *Lancaster Environment Centre, Lancaster University, Lancaster, UK*
[d] *Belfast Health Trust, Belfast, Northern Ireland, UK*
[e] *Dept. Civil and Environmental Engineering, U. Politécnica de Cataluña (UPC), Barcelona, Spain*
[f] *Dep. Computer Sciences, Applied Mathematics, and Statistics, University of Girona, Spain*

ARTICLE INFO

Keywords:
Geochemistry
Chronic kidney disease
Uncertain aetiology

ABSTRACT

This research uses an urban soil geochemistry database of elemental concentration to examine the potential relationship between Standardised Incidence Rates (SIRs) of Chronic Kidney Disease (CKD) of uncertain aetiology (CKDu), and cumulative low level geogenic and diffuse anthropogenic contamination of soils with PTEs. A compositional data analysis approach was applied to determine the elemental balance(s) of the geochemical data showing the greatest association with CKDu. The research concludes that both anthropogenic and geogenic factors may be contributing influences to explain high incidences of CKDu, up to 12 times greater in some Super Output Areas (SOAs) than would be expected for the average population. The role of As, Cr, Cu, Pb, Sb and Mo was highlighted, which may be linked to anthropogenic sources such as historical industrial sources, atmospheric pollution deposition and brake emissions. Geogenic factors were shown to be important in areas with elevated relative concentrations of naturally occurring potentially toxic elements (PTEs).

1. Introduction

An increasing amount of digital spatial data is available through geoscience surveys including sub-continental scale and regional scale ground-based geochemical baseline surveys and an ongoing global geochemical baseline survey. These have been used for a range of purposes including geological and soil mapping, baseline environmental quality documentation, mineral prospecting, soil resource management, human health risk evaluation, environmental education, land use planning and detection of contaminants. Such datasets enable the exploration of relationships between health outcomes and naturally occurring elements in soil and water such as potentially toxic elements (PTEs), particularly to inform our understanding of the links between chronic disease and cumulative exposure.

The data provided by geochemical surveys pose many challenges for accurate analysis and interpretation. The most well-known of these is the fact that geochemical data are compositional in nature and, as such, convey relative information. As a result, correlations between raw geochemical compositional data are spurious, prone to artefacts and potentially unrelated to any natural processes (Aitchison, 1986; Buccianti and Pawlowsky-Glahn, 2005). Compositional data analysis (CoDA) methods are frequently used to extract information from geochemical data by treating with log-ratio scores instead of analysing the raw values (Pawlowsky-Glahn and Egozcue, 2001; Tolosana-Delgado, 2012; McKinley et al., 2013). However, the results obtained can be difficult to interpret as the use of compositionally-compliant methods for geochemical surveys invalidates some of the traditional concepts of defining background and identifying anomalies which may be pertinent for understanding the relationships between elevated incidences of disease and exposure to PTEs.

For this research an urban soil geochemistry database of elemental concentrations was used to examine the potential relationship between Standardised Incidence Rates (SIRs) of Chronic Kidney Disease (CKD) of uncertain aetiology (CKDu) and cumulative low level soil contamination,

* Corresponding author.
E-mail address: j.mckinley@qub.ac.uk (J.M. McKinley).

https://doi.org/10.1016/j.acags.2020.100024
Received 30 October 2019; Received in revised form 6 March 2020; Accepted 25 March 2020
Available online 31 March 2020

whether of anthropogenic or geogenic origin. A forward-selection method (Rivera-Pinto et al., 2018) was applied to find those elemental balances of the geochemical data which had the greatest association with CKDu. Our goal was to find a compositionally-compliant method which provides an interpretable output. Balances offer very interpretable versions of log-contrasts. This approach acknowledges the compositional nature of the data and offers the opportunity to identify the PTEs with the relative abundances most associated with elevated incidences of CKD.

1.1. Compositional data analysis

The relative amounts of components, frequently expressed as constraints of constant sum, i.e. as closed data, have implications for the analysis of geochemical data. This realisation was first presented by Chayes (1960) who described the closure constraint. Common ways used to deal with this include the use of ternary or quaternary diagrams and summing components to form subsets. These methods have the same effect as the initial percentages or may even strengthen the closure effect (a closed ternary system). The use of ratios does eliminate the initial closure effect, but ratios are still bounded below and the problem of "spurious" correlations is not resolved through their use. Closure forces at least one negative correlation.

Log-ratios between components are unbounded and unaffected by constant sum closure effects caused by the relative nature of geochemical data (Buccianti and Pawlowsky-Glahn, 2005, 2011). Several families of log-ratio transformations exist. Aitchison (1986) introduced the pairwise log-ratio transformation (pwlr), the additive log-ratio transformation (alr) and the centred log-ratio transformation (clr). Egozcue et al. (2003) proposed the family of isometric log-ratio transformations (ilr), each based on the choice of an orthonormal coordinate system. None is inherently better than the other, and each has advantages and disadvantages.

1.2. Selection of balances

Accurate and understandable interpretation of log-ratios has been the topic of much debate. For example, the interpretation of the clr is not simple in general because the clr transformation involves all of the components in the sample. The clr log-scores must always be interpreted in relation to all components. This is particularly relevant and complex for the interpretation of relations between environmental factors and incidences of disease. A key concept in CoDa is the concept of balance between two groups of components of a composition (Egozcue et al., 2003; Egozcue and Pawlowsky-Glahn, 2005). A balance can be defined as corresponding to the normalized difference in means of the log-transformed abundances between two sub-compositions. The use of balances offers the opportunity to identify components (or geochemical elements) whose relative abundances are associated with elevated incidences of CKD.

Rivera-Pinto et al. (2018) have proposed a forward-selection algorithm called *selbal*, for the identification of two groups of variables whose relative abundance, or balance, is associated with the response variable of interest. It is important to acknowledge that it is possible to select balances using a knowledge-driven approach within a compositionally compliant context, for systems with many parts such as soil geochemistry (Kynčlová et al., 2016). However, the *selbal* technique can be used to predict the most relevant balances. The *selbal* approach is a model selection procedure that searches for a signature in terms of a balance between two groups of parts that adequately explains the response variable of interest. Starting with the balance composed of only two components which has the largest positive association with the response, the *selbal* algorithm performs a stepwise multiple linear regression with forward selection and, at each step, a new variable is added to the existing balance so that the specified association criterion is maximized. The algorithm stops when none of the possible additions increases the current association. A cross-validation (CV) procedure provides the association or discrimination value for each balance. To identify the set of balances an n-fold CV procedure is used. The dataset is subdivided into a training set comprised of (n-1) folds and the final fold is used as a test set. For this study a 5-fold CV procedure was used. The training set is used to determine the "best" balance based on the coefficient of determination and the associated regression equation is used to calculate the mean response for the test set based on the balances identified in the parameter estimation step.

The mean squared error (MSE), as a function of the number of components included in the balance, highlights the "optimal number" of variables identified in the forward selection process whose relative abundance is associated with the response variable of interest. Once the "optimal number" of variables has been determined, the approach selects the global balance, that is, the *selbal* algorithm is applied to the whole dataset with the specification that the maximum number of variables in the balance complies with the "optimal number". The CV result determines the robustness of the global balance. The relative frequency of different balances obtained in the CV process and the proportion of times that each component has been included into a balance is summarized in the form of a table. The CV process is also used to obtain the CV accuracy.

1.3. Data

1.3.1. Geochemical data

In the last decade several comprehensive regional and national soil sampling programmes have been completed. The Tellus Survey (Young and Donald, 2013) is an example of a ground-based geochemical survey which generated multivariate datasets. The datasets comprise soil, stream sediment and stream water samples from rural areas and soil samples from urban areas. This research uses the Tellus urban geochemical database consisting of 1000 sample points which have been analysed for 58 elements with XRF elemental analysis (Fig. 1A).

The Belfast urban area displays a diverse geological underlay (Fig. 1A; Mitchell, 2004; Young and Donald, 2013). Silurian greywacke and shales formed as part of the Southern Uplands-Down-Longford Terrane, comprise the oldest rocks, overlain by Permo-Triassic sandstones and mudstones. These successions are overlain by Cretaceous sandstone and chalk in the west which are covered by Palaeocene basalts along the north west boundary of the city. Previous work (Barsby et al., 2012; Cox et al., 2013; McKinley et al., 2013; McIlwaine et al., 2014, 2015; Palmer et al., 2015) has indicated that bedrock geology provides a strong geogenic source for PTEs in soils across Northern Ireland with observed concentrations for a range of PTEs including arsenic (As), chromium (Cr), cobalt (Co), copper (Cu), molybdenum (Mo), nickel (Ni), lead (Pb), vanadium (V) and zinc (Zn). Silurian shales and sandstone show relatively elevated As and Mo while Palaeogene basalts exhibit the strongest control over the distribution of Cr, Co, Ni, V and Zn. The superficial geology of Belfast is dominated by till, glacial sands and gravels with alluvium found in the vicinity of the main river, the River Lagan.

Urbanisation in the form of anthropogenic contamination has been found to have an increasing impact on soils resulting in increased PTEs. A recent study by McIlwaine et al. (2017) investigated the use of PTEs as 'urbanisation tracers' and found that PTEs could be split into two main groups explained by geogenic (Co, V, Cr and Ni) and anthropogenic (Zn, tin (Sn), Pb, antimony (Sb), As and Mo) factors. An anthropogenic source of Zn and Cu has been attributed to brass production (Herting et al., 2008) and As, Mo, Sn, Sb and Pb have been linked to atmospheric pollution deposition (Carrero et al., 2013) including traffic pollution. Researchers have reported Fe, Cu, Zn and Pb to be the most abundant metals in the brake lining (review by Grigoratos and Martini, 2015). Brake wear emissions have been also cited as a potentially important source of Sb and Mo (Grigoratos and Martini, 2015). The industrial legacy of ship building and development of Belfast City airport have been cited as potential anthropogenic sources for Sn, Sb and Pb (McIlwaine et al., 2017). The relevance of this for research into environmental influences on the incidence of CKDu is that studies have shown that

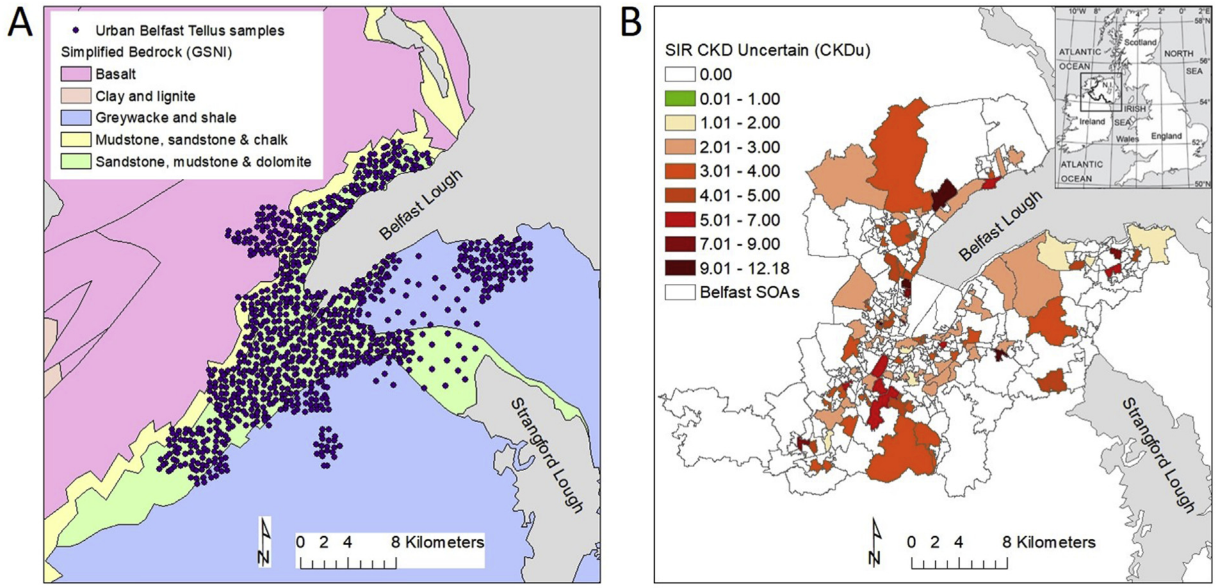


Fig. 1. A) Tellus Urban sample locations (Young and Donald, 2013) plotted on a simplified geological map (provided by Geological Survey of Northern Ireland (GSNI), Mitchell, 2004); B) Map of UKRR SIR for CKDu (uncertain aetiology) with insert showing location of Northern Ireland.

ultrafine particles (including Pb, Mo and Sb) may become blood-borne and translocate to other tissues such as the liver, brain and kidney (Geiser and Kreyling, 2010; Oberdörster et al., 2005).

1.3.2. *Chronic Kidney Disease (CKD)*

The United Kingdom Renal Registry (UKRR) collects data on all patients with advanced CKD on dialysis or with a kidney transplant (Renal Replacement Therapy (RRT)) on a regular basis across the UK. For this study, the UKRR provided SIRs for patients starting RRT between 2006 and 2016, by Super Output Area (SOAs) administrative wards. Created from the 2011 census outputs, SOAs are the smallest statistical geography boundaries within Northern Ireland (NISRA, 2013). In total there are 890 SOA administrative wards in Northern Ireland including 265 urban SOAs in Belfast. Data were provided in age brackets (16–39, 40–64 and 65+, all ages >16 and for uncertain aetiology (CKDu) for 2006–2016). A SIR for a SOA is a measure that quantifies the relationship between the expected incidence based on that of Northern Ireland as a whole and the actual incidence in the SOA. SIRs of exactly 1 indicate that the incidence for RRT for a SOA is equal to that expected based on Northern Ireland's average age-specific incidence rates. In summary, the value is equal to 1, if the expected and actual incidence match. SIRs above 1 indicate that the incidence for a SOA for RRT is greater than expected based on the Northern Ireland's average age-specific incidence rates.

CKD is a collective term for many causes of progressive renal failure which is increasing worldwide due to ageing, obesity and diabetes (Gilg et al., 2012; Lewis, 2012; UK Renal Registry, 2019). However, these factors cannot explain the environmental clusters of renal disease that are known to occur, such as the Balkan, Sri Lankan and Mesoamerican nephropathies, as discussed in Weaver et al. (2015). As this research aims to examine the potential relationship between incidence rates of CKD and cumulative low level soil contamination, we concentrated on SIRs for uncertain aetiology (CKDu). However, factors that cause CKDu may also be relevant to the heterogeneity of progressive CKD in diabetes and hypertension.

To demonstrate the forward-selection balance approach, SIR data for CKDu were selected for the SOAs in Belfast. Of the 265 SOAs within the greater Belfast area, 92 SOAs included SIRs for CKDu. Spatial autocorrelation of the SOAs for CKDu and the potential influence of the SOAs with no recorded CKDu were tested using Global Moran's *I*. We found a *z*-score of -0.8952 (p value $= 0.3707$) suggesting that the patterns do not appear to be significantly different to random. Within the Belfast urban area 15 SOAs showed SIRs for CKDu five times larger, with two SOAs showing SIRs for CKDu at least 10 times larger, than expected given Northern Ireland's average incidence rates (Fig. 1B). This statistic highlights the need for the current research in that these elevated incidence rates cannot be completely explained by known causes of CKD.

1.4. *Method*

Using the finding by McIlwaine et al. (2017) that the PTEs of Co, V, Cr, Ni, Zn, Sn, Pb, Sb, As and Mo represent 'urbanisation tracers' with which to determine geogenic and anthropogenic contamination, a knowledge-driven approach was used for this research using a sub-composition of these elements. Using the *selbal* CV option, log-transformed response variables (SIRs) and geochemical variables were used to identify the elemental balance with the best association with CKDu (unknown aetiology) in the 2006–2016 UKRR SIR data for the Greater Belfast urban area. The main goal of the analysis was to find the balance most associated with CKDu. A secondary aim was to investigate whether an elemental balance was able to discriminate between geogenic and anthropogenic sources potentially linked to CKDu. Initially, the analysis was conducted using the full urban dataset tagging the geochemical soil data to the urban SOAs with SIR data for CKDu. It was recognised by Aitchison (1986) that the arithmetic mean is not the best estimate for the centre of a dataset. Therefore, the analysis was run investigating using the geometric mean geochemical soil values for the SOAs. 92 SOAs were available for this analysis. Following this, SIR CKDu data were tagged to the soil geochemistry to provide stratification according to a simplified geology. In total, this provided 340 sample points which were further classified into those overlying different bedrock: basalt, sandstone, mudstone and lithic arenites (Fig. 1A). Since the areas with underlying limestone geology were very limited, these areas were not used for this stage of the analysis.

2. Results and discussion

Table 1 provides an overview of the results for the balances (Global Balance 1) identified as having the greatest association with CKDu for the different datasets. Using the geometric mean of the geochemical data for the SOAs, according to the CV procedure, the optimal number of

Table 1
Overview of results for the 92 Urban SOAs using the geometric mean, values of the soil data for each SOA. The results using SIRs for CKDu tagged to the soil geochemical points for different geology are also shown. The global balance is explained in the Selection of Balances section above. The results are discussed below.

Geochemical components for SIR_CKDu regression modelling	Global Balance	Estimate	Std. Error	t value	Pr (>|t|)	R-Squared
Belfast Urban area (92 SOAs) SIR_CKDu using geometric mean of soil data for SOAs	Intercept	1.7575	0.3593	4.892	4.34E-06	
	Co/Ni	0.6569	0.4389	1.497	0.138	0.02428
Belfast Urban area tagged SIR_CKDu to soil data (340 sample points)	Intercept	1.06227	0.06448	16.473	<2e-16	
	As/Mo	0.26841	0.14767	1.818	0.07	0.00968
Belfast Urban area tagged SIR_CKDu to soil data for basalt geology (17 sample points)	Intercept	2.726	0.653	4.175	0.000813	
	Mo/Co	0.448	0.2036	2.201	0.043834	0.2441
Belfast Urban area tagged SIR_CKDu to soil data for sandstone geology (132 sample points)	Intercept	1.23653	0.20909	5.914	2.78E-08	
	Co/Ni	0.01343	0.08072	0.166	0.868	0.0002128
Belfast Urban area tagged SIR_CKDu to soil data for mudstones (121 sample points)	Intercept	1.24016	0.07416	16.722	<2E-16	
	As/Mo	0.05135	0.06299	0.815	0.417	0.005554
Belfast Urban area tagged SIR_CKDu to soil data for lithic arenites (60 sample points)	Intercept	1.755	0.3755	4.674	1.81E-05	
	Sb/Mo	0.2635	0.1394	1.891	0.0637	0.05805

components to be included in the balance was two (Table 1; Fig. 2a). The global balance shows the relative importance of Ni and Co in the association with CKDu with Ni appearing 62% and Co appearing 31% of the time (Fig. 2b). This is an indicator of robustness for the proposed global balance. Further potential secondary and tertiary balances were indicated involving Mo appearing 38% of the time, with As and Cr (appearing 35% and 19% respectively). Although the optimal number of components was identified as two, the significance of three components, as identified in the MSE plot (Fig. 2a), can be interpreted by the important role of Ni, in addition to Mo and As in balances. A possible interpretation is that the larger balance scores indicate that there are two dominant influences for the Belfast urban area: an anthropogenic control with larger relative abundances of As with respect to Mo and a geogenic influence identified with larger relative abundances of Co and Cr with respect to Ni.

When the SIRs for CKDu were tagged to the soil geochemistry data the overall findings remain consistent with the above results (Table 1; Fig. 2c and d). However, since the regression has now been carried out with repeated observations, that is each SOA response (the SIR) appears as many times as geochemical analyses are available in the SOA, this will have captured more efficiently the dispersion or variability of geochemistry over the area. The balance most frequently identified in the CV procedure, as being most associated with CKDu, was Mo (appearing 68% of the time) versus As (appearing 58% of the time) with secondary balances identified between Ni (appearing 32% of the time) with Co (appearing 22% of the time) and Mo with Cu (appearing 10% of the time). Although the optimal number of components as identified by the *selbal* algorithm is two, the significance of three components as identified in the MSE plot is most likely explained by the important role of Ni, in addition to Mo and As in the balances.

The findings suggest that both anthropogenic and geogenic factors may be important in understanding the relationship between the

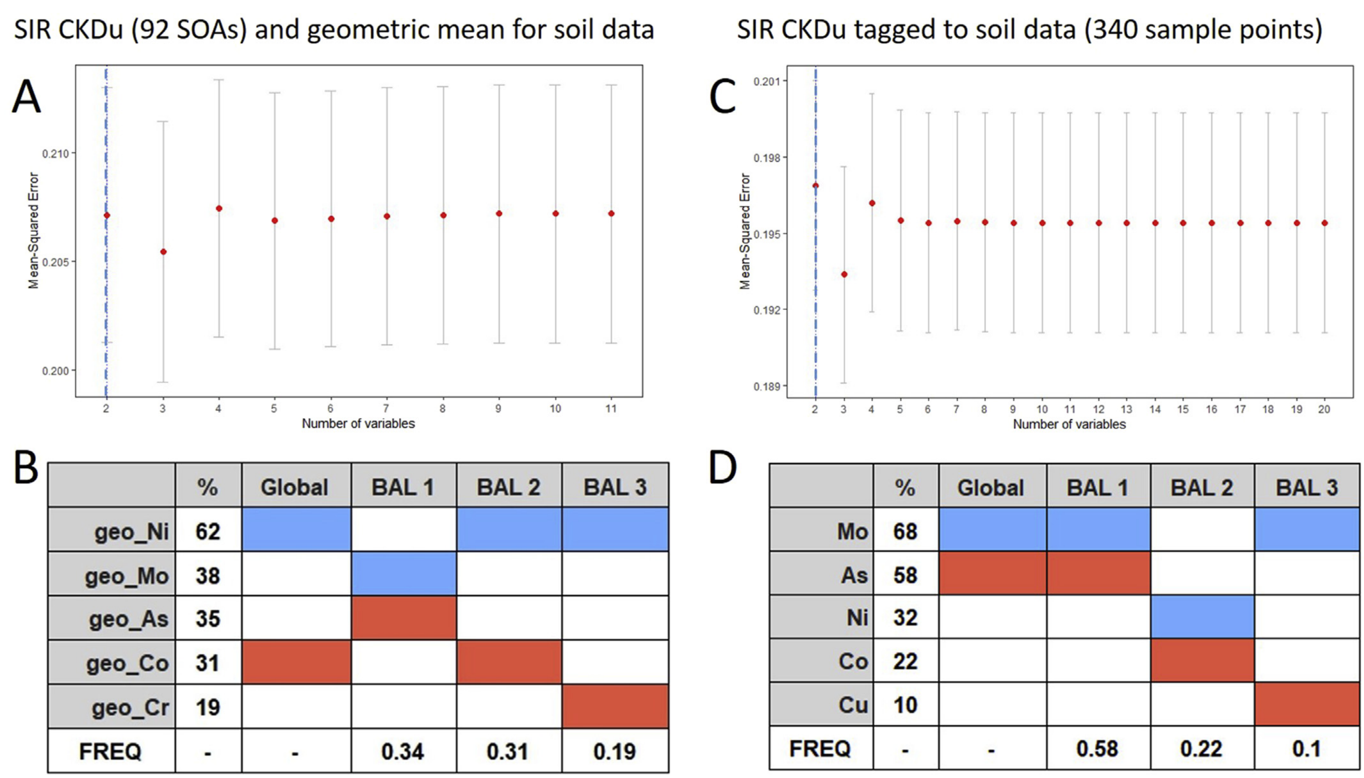


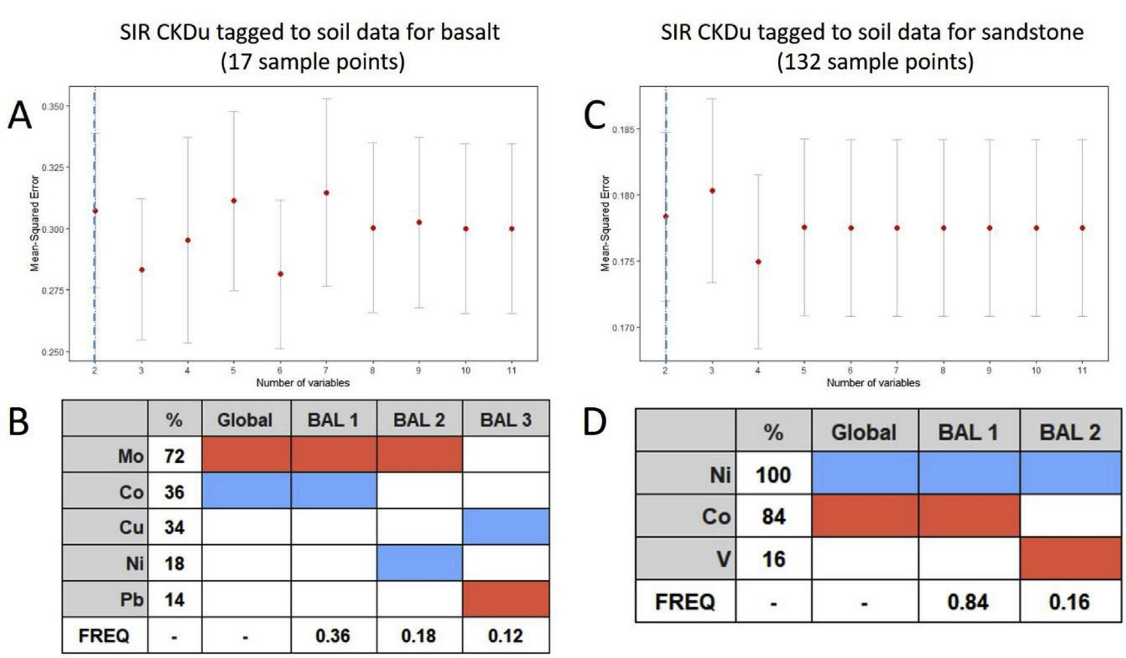
Fig. 2. Results for Belfast urban area (92 SOAs) SIR CKDu using the geometric mean values of soil data for each SOA. A) Mean squared error (MSE) as a function of the number of components included in the balance - the optimal number of components is highlighted with a vertical dashed line; and B) The global balance most frequently identified in the CV procedure, using the whole dataset, along with other balances identified in the CV procedure. C) MSE results for SIR CKDu tagged to soil data (340 sample points). D) Global and other balances identified in CV procedure relating to (C). Red indicates the numerator and blue the denominator in balances. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

environment and incidences of CKD of unknown aetiology. While the influence of anthropogenic factors may be important in understanding the environmental association with increased incidences of CKDu, the role of geogenic influences in the environment should not be overlooked. Geogenic factors controlled by the underlying bedrock geology may be especially relevant in areas with elevated relative concentrations of naturally occurring elements such as Ni, Co and Cr. The findings suggest a dominant geogenic influence most associated with CKDu with larger relative abundances of Co and Cr with respect to Ni due to elevated concentrations in areas of Belfast underlain by Palaeogene basalts and glacial deposits accumulated from basalt bedrock. Stratification by geology allows us to consider elemental balances which discriminate between anthropogenic and geogenic sources potentially associated with CKDu (Table 1; Figs. 3 and 4). The results consistently show the influence of both anthropogenic and geogenic factors in that the balances Mo/Co, As/Mo and Sb/Mo are most frequently identified in the CV procedure for basalt, mudstone and lithic arenite, respectively. Secondary and tertiary balances included Pb/Cu, Mo/Pb and Pb/Sb for basalt, mudstone and lithic arenite, respectively A geogenic influence is demonstrated most strongly for sample points tagged to an underlying sandstone bedrock, as Ni appears 100% of the time in the balances as the most associated with CKDu. This finding requires further consideration as the PTEs of Ni, Co and V, are known to be present in elevated relative quantities in Palaeocene basalts rather than Sherwood sandstone. These findings indicate the importance of superficial deposits in addition to bedrock geology and, in particular, glacial processes producing a glacial overprint of PTEs, originating from basalt bedrock, over areas underlain by sandstone in certain areas of urban Belfast.

This research identified a potential link between CKDu and the relative abundances of PTEs in the Greater Belfast urban area. A link between CKDu and anthropogenic sources of PTEs in urban areas has not previously been demonstrated. Moreover, this research demonstrates the potential importance of both anthropogenic *and* geogenic factors in relation to the incidence of CKDu. The results consistently suggested the importance of Mo and As, as these were the most frequently identified in the balances most associated with CKDu using the CV procedure. The role of Mo, Sb and Pb in balances was also shown. The PTE of Pb, although not observed in global balances, was found to appear 14–20% of the time in secondary and tertiary balances (Figs. 3 and 4). As Pb is a well-known nephrotoxic, this finding, indicating the significant role of Pb in balances, is important in understanding the relationship between CKDu and environmental factors. For the Greater Belfast area, the industrial legacy of ship building and the more recent development of Belfast City airport have been cited as potential anthropogenic sources for both Sb and Pb (McIlwaine et al., 2017). This needs to be explored further but this research has identified a potential link between CKDu and the relative abundances of these PTEs. Atmospheric pollution deposition, including traffic pollution, has been cited as a source for the PTEs of Mo, Sb, As and Pb (Carrero et al., 2013; Grigoratos and Martini, 2015). Moreover, brake wear emissions have been cited as a potentially important source of Sb and Mo (Grigoratos and Martini, 2015). Research indicating a link between air pollution and kidney disease is more recent (Afsar et al., 2019). Studies have shown that ultrafine particles (including Pb, Mo and Sb) may become blood-borne and translocate to other tissues such as the liver, brain and the kidneys (Geiser and Kreyling, 2010; Oberdörster et al., 2005), supporting a plausible pathway between anthropogenic PTEs and CKDu, and underlining the potential importance of this research. This research has shown that soils show evidence of air pollution deposition and the potential impact of modern pollutants on health. The implications are that PTEs in urban soils may be indicative of the availability of nephrotoxins for human intake from environmental pollution.

3. Conclusions

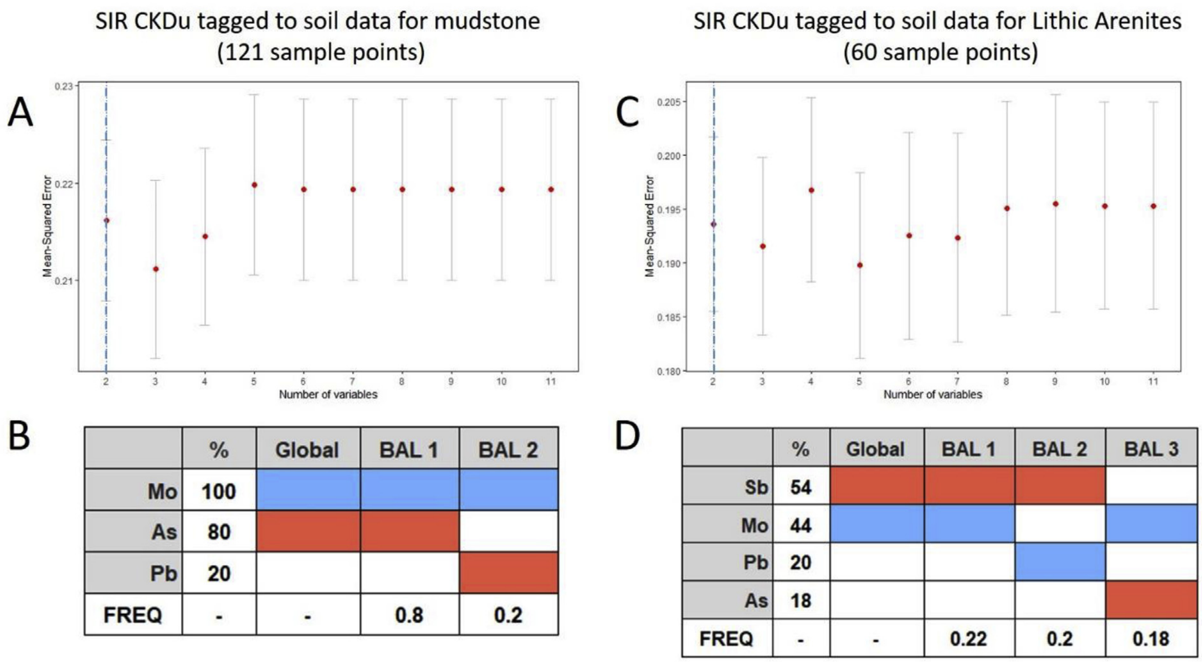
This research used an urban soil geochemistry database of total elemental concentrations to examine the potential relationship between Standardised Incidence Rates (SIRs) of Chronic Kidney Disease (CKD) of uncertain aetiology (CKDu) and cumulative low level PTEs found in soils, whether of anthropogenic or geogenic origin. A forward-selection method was applied to determine the elemental balances of PTEs which had the greatest association with CKDu.



	%	Global	BAL 1	BAL 2	BAL 3
Mo	72				
Co	36				
Cu	34				
Ni	18				
Pb	14				
FREQ	-	-	0.36	0.18	0.12

	%	Global	BAL 1	BAL 2
Ni	100			
Co	84			
V	16			
FREQ	-	-	0.84	0.16

Fig. 3. Results for Belfast urban area SIR CKDu tagged to soil data for each SOA for basalt and sandstone. For basalt A) Mean squared error (MSE) as a function of the number of components included in the balance - the optimal number of components is highlighted with a vertical dashed line; and B) The balance most frequently identified in the CV procedure, using the whole dataset, along with other balances identified in CV procedure. For sandstone C) MSE results and D) Global and other balances identified in CV procedure. Red indicates the numerator and blue the denominator in balances. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)



	%	Global	BAL 1	BAL 2
Mo	100			
As	80			
Pb	20			
FREQ	-	-	0.8	0.2

	%	Global	BAL 1	BAL 2	BAL 3
Sb	54				
Mo	44				
Pb	20				
As	18				
FREQ	-	-	0.22	0.2	0.18

Fig. 4. Results for Belfast urban area SIR CKDu tagged to soil data for each SOA for mudstone and lithic arenite. For mudstone A) Mean squared error (MSE) as a function of the number of components included in the balance - the optimal number of components is highlighted with a vertical dashed line; and B) The balance most frequently identified in the CV procedure, using the whole dataset, along with other balances identified in CV procedure. For lithic arenite C) MSE results; and D) Global and other balances identified in CV procedure. Red indicates the numerator and blue the denominator in balances. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

We found that PTEs commonly attributed to both anthropogenic and geogenic factors were associated with high incidences of CKDu, with CKDu incidence up to 12 times greater in some SOAs than would be expected for the average population across the greater Belfast area. The key findings from the research are:

- Soils showed evidence of PTEs from air pollution deposition, which have known potential impacts on human health.
- Mo and As, commonly attributed to air pollution deposition, were consistently found in the elemental balances most frequently identified as the most associated with CKDu.
- The significance of the ratio of Mo, Sb, and Pb in balances was also shown which may be linked to air pollution deposition and traffic pollution.
- Geogenic factors controlled by the underlying bedrock geology are especially important in areas with elevated relative concentrations of naturally occurring elements such as Ni, Cr, Co and V. Superficial deposits resulting from glacial processes can produce a glacial overprint of PTEs.

The conclusion is that CKDu incidence is shown here to be associated with environmental PTEs using a balances approach, and that the approach was sufficiently clear to support inferences regarding characteristics of PTEs for different areas.

Declaration of competing interest

This is to confirm that we have no conflict of interest for the manuscript of the above title.

CRediT authorship contribution statement

Jennifer M. McKinley: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Writing - review & editing. **Ute Mueller:** Conceptualization, Methodology, Software, Validation, Data curation, Writing - original draft, Writing - review & editing. **Peter M. Atkinson:** Conceptualization, Methodology, Validation, Writing - review & editing. **Ulrich Ofterdinger:** Validation, Writing - review & editing. **Chloe Jackson:** Conceptualization. **Siobhan F. Cox:** Validation, Writing - review & editing. **Rory Doherty:** Validation, Writing - review & editing. **Damian Fogarty:** Conceptualization, Methodology, Validation, Writing - review & editing. **J.J. Egozcue:** Conceptualization, Methodology, Software, Validation, Writing - review & editing. **V. Pawlowsky-Glahn:** Conceptualization, Methodology, Software, Validation, Writing - review & editing.

Acknowledgements

The Tellus Project was funded by the Department of Enterprise, Trade and Investment (NI) and by the 'Building Sustainable Prosperity' fund of the EU Regional Development Programme. The study received ethics approval in March 2018 from the NHS National Research Ethics Committee REC reference: 15/EM/0366. We thank all the UK Renal Centres for providing data to the UK Renal Registry. The views and opinions expressed in this article are those of the authors and do not reflect the views of the UK Renal Registry or UK Renal Association. We appreciate the helpful comments of two anonymous reviewers.

References

Aitchison, J., 1986. The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London, UK, p. 416.

Afsar, B., Elsurer Afsar, R., Kanbay, A., Covic, A., Ortiz, A., Kanbay, M., 2019. Air pollution and kidney disease: review of current evidence. Clin. Kidney J. 12 (1), 19–32. https://doi.org/10.1093/ckj/sfy111. https://www.ncbi.nlm.nih.gov/pubmed/30746128.

Barsby, A., McKinley, J.M., Ofterdinger, U., Young, M., Cave, M.R., Wragg, J., 2012. Bioaccessibility of trace elements in soils in Northern Ireland. Sci. Total Environ. 433, 398–417. https://doi.org/10.1016/j.scitotenv.2012.05.099.

Buccianti, A., Pawlowsky-Glahn, V., 2005. New perspectives on water chemistry and compositional data analysis. Math. Geol. 37 (7), 703–727.

Buccianti, A., Pawlowsky-Glahn, V. (Eds.), 2011. Compositional Data Analysis: Theory and Applications. Wiley, Chichester.

Carrero, J.A., Arrizabalaga, I., Bustamante, J., Goienaga, N., Arana, G., Madariaga, & J.M., 2013. Diagnosing the traffic impact on roadside soils through a multianalytical data analysis of the concentration profiles of traffic-related elements. Sci. Total Environ. 458–460, 427–434. https://doi.org/10.1016/j.scitotenv.2013.04.047.

Chayes, F., 1960. On correlation between variables of constant sum. J. Geophys. Res. 65 (12), 4185–4193.

Cox, S., Chelliah, M., McKinley, J.M., Palmer, S., Ofterdinger, U., Young, M., Cave, M.R., Wragg, J., 2013. The importance of solid-phase distribution on the oral bioaccessibility of Ni and Cr in soils overlying Palaeogene basalt lavas, Northern Ireland. Environ. Geochem. Health 35 (5), 553–567. https://doi.org/10.1007/s10653-013-9539-6.

Egozcue, J.J., Pawlowsky-Glahn, V., 2005. Groups of parts and their balances in compositional data analysis. Math. Geol. 37 (7), 795–828.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. Math. Geol. 35 (3), 279–300.

Geiser, M., Kreyling, W.G., 2010. Deposition and biokinetics of inhaled nanoparticles. Particle Fiber Toxicol. 7 (2). https://doi:10.1186/1743-8977-7-2.

Gilg, J., Castledine, C., Fogarty, D., 2012. UK Renal Registry 14th Annual Report: Chapter 1 UK RRT Incidence in 2010: National and Centre-Specific Analyses (Southampton: Karger AG, Basel).

Grigoratos, T., Martini, G., 2015. Brake wear particle emissions: a review. Environ. Sci. Pollut. Res. 22, 2491–2504 https://doi.10.1007/s11356-014-3696-8.

Herting, G., Goidanich, S., Odnevall Wallinder, I., Leygraf, C., 2008. Corrosion induced release of Cu and Zn into rainwater from brass, bronze and their pure metals. A 2-year field study. Environ. Monit. Assess. 144, 455–461. https://doi.org/10.1007/s10661-007-0008-5.

Kynčlová, P., Filzmoser, P., Hron, K., 2016. Compositional biplots including external non-compositional variables. Statistics 50 (5), 1132–1148. https://doi:0.1080/02331888.2015.1135155.

Lewis, R., 2012. Understanding Chronic Kidney Disease: A Guide for the Non-specialist. M&K Update Ltd, England.

Mitchell, W. (Ed.), 2004. The Geology of Northern Ireland, second ed. Geological Survey of Northern Ireland.

McKinley, J.M., Ofterdinger, U., Young, M., Barsby, A., Gavin, A., 2013. Investigating local relationships between trace elements in soils and cancer data. Spatial Stat. 5, 25–41. https://doi.org/10.1016/j.spasta.2013.05.001.

McIlwaine, R., Cox, S., Doherty, R., Palmer, S., Ofterdinger, U., McKinley, J.M., 2014. Comparison of methods used to calculate typical threshold values for potentially toxic elements in soil. Environ. Geochem. Health 36 (5), 953–997. https://doi.org/10.1007/s10653-014-9611-x.

McIlwaine, R., Cox, S., Doherty, R., 2015. When are total concentrations not total? Factors affecting geochemical analytical techniques for measuring element concentrations in soil. Environ. Sci. Pollut. Control Ser. 22 (8), 6364–6371.

McIlwaine, R., Doherty, R., Cox, S., Cave, M., 2017. The relationship between historical development and potentially toxic element concentrations in urban soils. Environ. Pollut. 220 (B), 1036–1049. https://doi.org/10.1016/j.envpol.2016.11.040.

Northern Ireland Statistical Research Agency (NISRA), 2013. Modifications for 2001 Super Output Areas for the 2011 Census Data. Accessed 25.2.2020. ttps://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/SOAPaperJan2013.pdf.

Oberdörster, G., Oberdörster, E., Oberdörster, J., 2005. Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. Environ. Health Perspect. 113, 823–839.

Palmer, S., McIlwaine, R., Ofterdinger, U., Cox, S.F., McKinley, J.M., Doherty, R., Wragg, J., Cave, M., 2015. The effects of lead sources on oral bioaccessibility in soil and implications for contaminated land risk management. Environ. Pollut. 198, 161–171. https://doi.org/10.1016/j.envpol.2015.01.004.

Pawlowsky-Glahn, V., Egozcue, J.J., 2001. Geometric approach to statistical analysis on the simplex. Serra Or 15 (5), 384–398.

Rivera-Pinto, J., Egozcue, J.J., Pawlowsky-Glahn, V., Paredes, R., Noguera-Julian, M., Calle, M.L., 2018. Balances: a new perspective for microbiome analysis. mSystems. https://doi.org/10.1101/219386.

Tolosana-Delgado, R., 2012. Uses and misuses of compositional data in sedimentology, Sedimentology. Sediment. Geol. 280, 60–79.

UK Renal Registry, 2019. UK renal Registry 21st annual report – data to 31/12/2017, Bristol, UK. Available from: https://www.renalreg.org/publications-reports/.

Weaver, V.M., Fadrowski, J.J., Jaar, B.J., 2015. Global dimensions of chronic kidney disease of unknown etiology (CKDu): a modern era environmental and/or occupational nephropathy? BMC Nephrol. 16, 145. https://doi.org/10.1186/s12882-015-0105-6. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4539684/.

Young, M.E., Donald, A.W. (Eds.), 2013. A Guide to the Tellus Data. Geological Survey of Northern Ireland (GSNI), Belfast. http://nora.nerc.ac.uk/509171/.