Chronic kidney disease of uncertain aetiology and its relation with waterborne environmental toxins: An investigation via compositional balances

Jennifer M. McKinley[1*], Ute Mueller[2], Peter M. Atkinson[1,3], Ulrich Ofterdinger[1], Siobhan F. Cox[1], Rory Doherty[1,] Damian Fogarty[4] and J.J. Egozcue[5]

[1] School of Natural and Built Environment, Queen's University Belfast, Northern Ireland

[2] School of Science, Edith Cowan University, Perth, Western Australia

[3] Lancaster Environment Centre, Lancaster University, UK

[4] Belfast Health Trust, Belfast, Northern Ireland

[5] Dept. Civil and Environmental Engineering, U. Politécnica de Cataluña (UPC), Barcelona, Spain

Abstract

The occurrence of environmental clusters of chronic kidney disease of uncertain aetiology (CKDu), where there is no known cause for the onset of kidney dysfunction, is a concern globally. Waterborne exposure pathways in the environment may result in indirect or direct ingestion of trace elements with potential health risks. This research examines the relationship between standardized incidence rates (SIRs) of CKDu and the log-ratio balances of potentially toxic elements (PTEs) in regional stream water. Compositional elemental balances were created and regression was used to identify the balances most associated with log-transformed SIR CKDu. At the regional scale, a statistically significant relationship was found between log(CKDu SIR) and the elemental balance Al/As which effectively delineates different geological domains across Northern Ireland. Following stratification by basalt bedrock (the dominant bedrock geology for SIR CKDu), the balance Al/Fe was identified as significantly associated with log(CKDu SIR). Superficial deposits, dissolved organic carbon and pH may act as controls on the balance of Al and Fe. With a high proportion of private water supplies registered in these areas, this research highlights the importance of considering bedrock geology and superficial deposits in understanding multi-element interactions of waterborne environmental toxins and potential links with environmental clusters of CKDu.

We would like to acknowledge and thank our esteemed colleague and friend Vera Pawlowsky-Glahn who has been instrumental in the development of this research, pioneering the importance of compositional data analysis (CoDA) in the study of health and the environment. For several of us, Vera provided our first introduction to CoDA and over many years has aided our shared understanding and increased awareness of the need to address the compositional nature of data such as soil and water geochemistry in our research. We are pleased to have this work on the use of compositional balances in exploring the explanatory environmental factors related to chronic kidney disease of uncertain aetiology included in this Zestschrift. It represents an exciting field of innovative research where an acknowledgement of CoDA principles is vital if we are to understand fully the critical relationship between our health and the environment.

Introduction

The World Health Organization [1] divides naturally occurring elements into three groups based on their nutritional significance in humans: 1) essential elements (Iodine (I), zinc (Zn), selenium (Se), copper (Cu), molybdenum (Mo) and chromium (Cr)); 2) elements which are probably essential (manganese (Mn), silicon (Si), nickel (Ni), boron (B) and vanadium (V)); and 3) potentially toxic elements (PTEs), some which may have essential functions at low levels (lead (Pb), cadmium (Cd), mercury (Hg), arsenic (As), aluminum (Al), lithium (Li) and tin (Sn)). However, individual elements or single component values may not be the most appropriate way to evaluate the risk to human health. An inherent assumption is that elemental data, such as geochemical data, represent absolute abundances. The WHO [1] recognizes that multi-element interactions can increase the bioavailability or modify the metabolism of both essential and potentially toxic trace elements (PTEs). Therefore, a multivariate approach is more suitable to investigate the relationship between human health and elements in the environment. However, because of the compositional nature of geochemical data single elements cannot be compared directly with one another. An alternative approach is presented here, using compositional data analysis (CoDA). In particular, the relation between standardized incidence rates (SIRs) of chronic kidney disease of uncertain aetiology (CKDu) and the log-ratio balances of PTEs in regional stream water is explored.

Several factors are known to cause chronic renal disease (CKD) including age, ethnicity and pre-existing medical conditions [2,3,4,5,6] However, reporting of environmental clusters of CKDu has increased globally to such an extent that a WHO special task force was set up to investigate the incidence of CKDu without known risk causes [7,8,9,10,11,12,13,14]. The UK renal registry (UKRR) reports data for nine different identifiable primary renal diseases (PRD) - diabetes, glomerulonephritis, hypertension, polycystic kidney disease, pyelonephritis, renal vascular disease, other, missing and uncertain aetiology. Variation between renal centres cannot account for the significant percentage of cases of uncertain aetiology recorded (e.g. 29.4% of patients' in 2017 were identified to diabetes and 14.9% to uncertain aetiology [15]. The environmental causes of CKDu, require investigation to address the global issue of geographic clusters of CKD where there is no known cause for the onset of kidney dysfunction. The causes of CKDu are most likely multifactorial [13]. Environmental factors that may cause CKDu, including the role of environmental nephrotoxins such as As, Hg and Pb [16] and toxic organic compounds from coal [17], require further exploration as these factors may also contribute to the heterogeneity of other causes of progressive CKD in diabetes and hypertension.

Bedrock and superficial geology have been shown previously to be important potential geogenic sources of PTEs in soils and stream waters across Northern Ireland [18,19,20,21]. Previous research using the UK Renal Register (UKRR) and environmental data is limited [22,23,24]. Jackson et al. [22] used a compositional Poisson regression approach and highlighted the interaction between essential elements, PTEs and potential associations with CKDu. McKinley et al. [23] investigated the use of compositional balances in a study of urban soils and found geogenic and anthropogenic signatures related to CKDu, while McKinley et al. [24] explored the underlying causes of CKDu using social deprivation and environmental factors. In this research, compositional balances were used to explore the water borne exposure pathways using geochemical stream water elemental data to represent potential health risks associated with indirect or direct ingestion of trace elements locally.

Study Area and Data

Study Area

Northern Ireland straddles three geological terrains referred to as the Central Highlands (Grampian) Terrane, Midland Valley and the Southern Uplands-Down-Longford Terrane (Fig. 1A; [25]). Mid-to-late Neoproterozoic rocks of the Dalradian supergroup including clastic marine sediments, basaltic magmatism, volcaniclastic sediments and turbidites, comprise most of the Central Highlands (Grampian) Terrane. These rocks experienced regional metamorphism. The Midland Valley Terrane comprises distinct geological units bounded by faulting, including an early Palaeozoic igneous complex and Proterozoic psammites and pelitic rocks, which have experienced different folding phases and metamorphism. Continental Devonian red bed sediments and marine Carboniferous sediments overlie the Midland Valley Terrane in the south of the country. The Southern Uplands-Down-Longford Terrane is composed of Ordovician and Silurian rocks with a younger igneous intrusive complex. Permian, Triassic, Jurassic and Cretaceous rocks are overlain in the North West of the country by Palaeogene basalt lava. At least 80% of the bedrock is covered by superficial deposits including glacial till, post-glacial alluvium and peat (summarized from [25,26]).

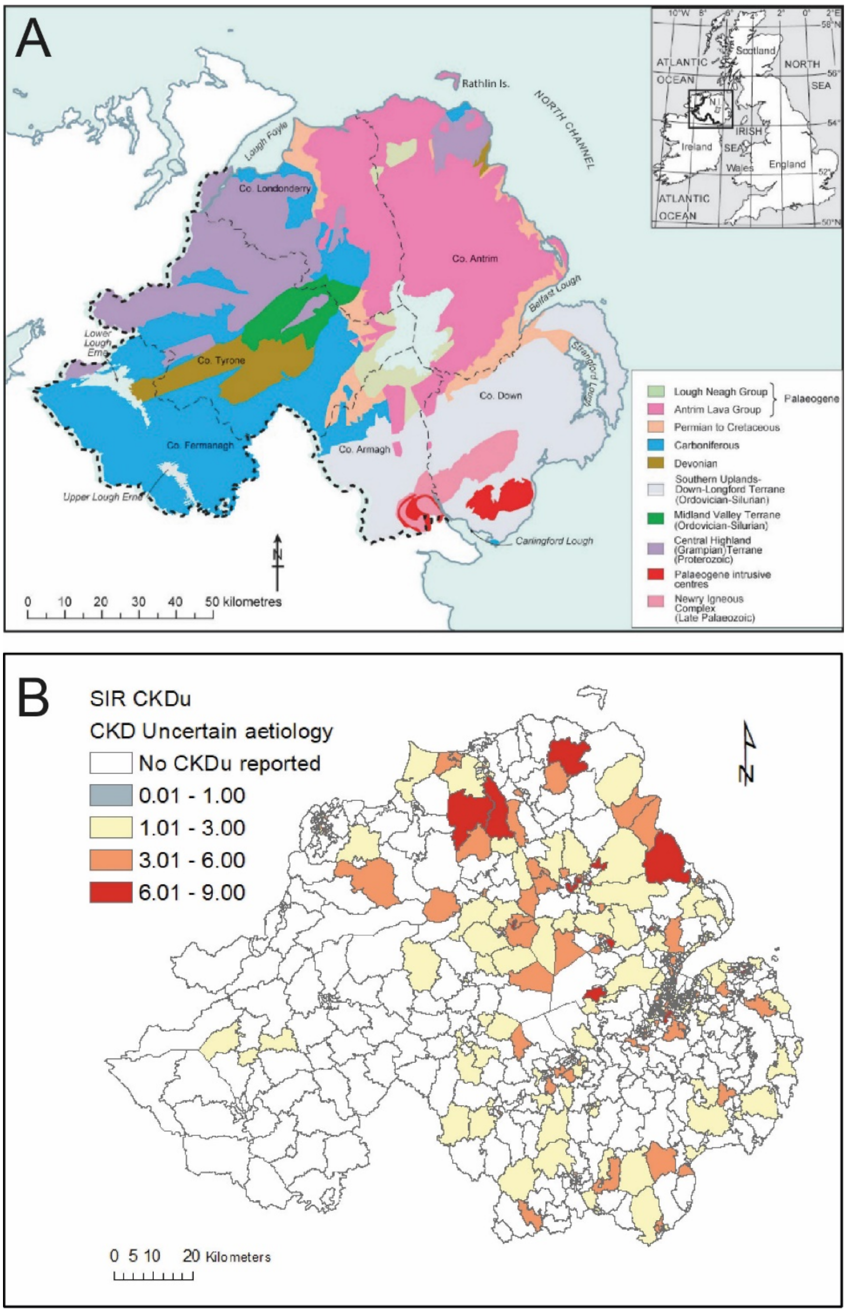

Fig.1 A) Simplified geology of Northern Ireland (provided by Geological Survey of Northern Ireland (GSNI), [26]). B) Standardised Incidence Rates (SIRs) for patients starting Renal Replacement Therapy (RRT) between 2006-2016, by Super Output Areas (SOAs) for Chronic Kidney Disease of uncertain aetiology (CKDu).

Chronic Kidney Disease of uncertain aetiology (CKDu)

The UKRR provided Standardised Incidence Rates (SIRs) for patients starting Renal Replacement Therapy (RRT) between 2006-2016, by Super Output Area (SOA). Data were grouped by age brackets and for uncertain aetiology (CKDu; Fig. 1B) for the 890 SOA administrative wards across Northern Ireland. The SIR for a SOA is a measure that quantifies the relationship between the actual incidence in the SOA and the expected incidence based on that of Northern Ireland as a whole. SIRs of exactly 1 indicate that the incidence of CKD for a SOA is equal to that expected based on Northern Ireland's average age-specific incidence rates. SIRs above 1 indicate that the incidence of CKD for a SOA is greater than expected and SIRs below1 indicate that the incidence for a SOA is lower than expected based on the Northern Ireland's average age-specific incidence rates.

Stream water geochemistry

This research used a regional geochemical survey of Northern Ireland's stream waters, completed by the Geological Survey of Northern Ireland (GSNI) and the British Geological Survey (BGS). Sampling and quality control were undertaken according to the G-BASE protocol of BGS. Stream waters were sampled in 1994-96 (2,908 sites in the west) and in 2004-06 (2,966 sites in the east). The sampling distribution averaged one site per 2.4 km^2. The samples were analysed by ICP MS (ug l^{-1}; trace element analysis) and ICP-AES (mg l^{-1}; anion analysis). The GSNI and BGS confirm that the complete stream water dataset can be used with confidence that element variability is attributable to natural or anthropogenic sources as opposed to analytical instrumentation or temporal variations [27]. For this study all stream water data were converted into ug l^{-1}. The GBase protocol was followed where a sample below the detection limit (DL) was replaced with a value of half the DL [27]

Methodology

In CoDA, a balance can be defined as corresponding to the normalized difference in means of the log-transformed abundances between two sub-compositions [28,29]. The concept of balances between two groups of components of a composition provides an interpretable way to identify components (or geochemical elements) whose relative abundances may be associated with a response variable. Several approaches to select balances are available, which may be grouped as data- or knowledge-driven approaches within a compositionally-compliant context [30,28,29,32,33]. Previous research has investigated the application of CoDA to water chemistry [31] and complex aqueous geochemical systems [34,35,36,37]. A forward-selection method, called the *selbal* algorithm has been proposed by Rivera-Pinto et al. [38]. A description of the *selbal* procedure is given here but for full details readers are referred to studies on the development of this approach [38,39].

The *selbal* algorithm performs a stepwise multiple linear regression using forward selection where variables are included sequentially at each step, starting with a balance including 2 variables to the maximum number of variables possible. 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. An n-fold cross-validation (CV) procedure provides the strength of association or discrimination value for each balance. In the CV procedure, at each iteration (20 iterations were used in this research), the data are subdivided into n disjoint subsets whose union is the entire set and each of the n subsests is used as a test set with the associated training set comprised of the remaining $(n - 1)$ folds. Then the *selbal* algorithm is applied to the training dataset and the optimal balance from the maximum number of variables is obtained.

The classification accuracy (mean squared error (MSE) in this case) of these balances is measured on the test dataset. Since there are 20 iterations and n folds for each number of variables of the balance, the mean MSE and standard error can be computed. The “optimal number” of elements to be included in the balance is defined as the least number of variables for which the mean MSE of the balance is within 1 standard error of the minimum mean MSE. As described by Rivera-Pinto [39], this is similar to the cross-validation process in LASSO for finding the optimal penalisation parameter lambda [40]. Once the “optimal number” of elements has been determined, this defines what is termed the global balance. The CV results are used to explore the robustness of the global balance, that is, the selbal algorithm is applied to the whole dataset with the specification that the maximum number of components in the balance is the “optimal number” of elements as defined above by the MSE. All of the balances, which meet the MSE requirement, obtained in the cross-validation process are compared then with the global balance. Output tables provide the relative frequency of the different balances obtained in the CV process and the proportion of times that each element has been included into a balance.

This approach has been used previously to identify urban soil geochemical elements whose relative abundances indicate geogenic and anthropogenic sources of PTEs associated with elevated incidences of CKDu [23] and demonstrated for CKD data with urban soil geochemistry and social deprivation indices [24].

Data preparation

The stream water geochemistry data were tagged to the SIRs of CKDu and the geometric mean was calculated for each Super Output Area (SOA). Initially, the relationship between CKDu and a subcomposition of the stream water geochemistry was explored for all SOAs with CKDu across Northern Ireland (109 SOAs). The subcomposition comprised 14 PTEs identified in the WHO report (Al, As, Cd, Cr, Cu, K, Mn, Mo, Ni, Pb, Se, U, V and Zn) and the anions (Ca, Fe, Mg, Na and Si). As the incidence of CKDu is spatially variable across the country (Fig. 1b), the stream water geochemical data were tagged to geology using a simplified geology scheme provided by GSNI. As a high proportion of SIRs of CKDu were found to overly basalt bedrock (58 SOAs), this class only was used for further analysis. The other geological domains were not explored further in this research due to limited SIRs of CKDu reported for the SOAs.

Balance approach

Two approaches were used. The first approach was the forward-selection method using the *selbal* algorithm. A *5*-fold cross-validation (CV) procedure was used to identify the set of balances. An associated generalised linear regression model (GLM) was used to calculate the mean response based on the balances identified in the parameter estimation step. The second approach used the isometric log-ratio (ilr) of the subcomposition based initially on a sequential binary partition (SBP) and secondly using an ilr balanced approach, both using the R package *compositions* [41]. GLM regression was used to examine any potential relationship between CKDu SIRs and the stream water geochemistry.

The two compositional balance approaches (log-transformed using the selbal cross-validation option; and ilr transformed using an SBP approach and ilr balanced approach) and the environmental elemental variables were used to identify the elemental balance with the greatest association with log-transformed CKDu SIR using the 2006-2016 UKRR SIR data.

Spatial Dependence

Regression models assume independence between the observations. However, this assumption may not be valid for spatial data [42, 43]. To assess the nature of the impact of spatial autocorrelation in this research Moran's *I* test statistic was used to test for spatial autocorrelation in the residuals computed from the regression models for the regional geology and basalt case studies.

Table 1. Summary of regression results for Standardised Incidence Rates (SIRs) for Chronic Kidney Disease of uncertain aetiology (CKDu) and elemental balances. Only significant statistical results are shown.

Dependent variable		Estimate	Std. Error	t value	Pr(>|t|)	p value
Regional study log(CKDu SIR)						
Selbal	Intercept	0.7015	0.0987	7.106	1.34E-10	<0.001
	Global Balance Al/As	0.1587	0.04164	3.81	0.00023	<0.001
SPB ilr base	Intercept	-1.3510	1.6997	-0.795	0.4288	
	ilr1 Al/As	-0.1788	0.0958	-1.865	0.0654	0.1
ilr balanced	Intercept	-1.3510	1.6997	-0.795	0.4288	
	ilr 1 Al/As	-0.1788	0.0958	-1.865	0.0654	0.1
Basalt bedrock log(CKDu SIR)						
Selbal	Intercept	1.5525	0.1457	10.654	4.38E-15	<0.001

	Global Balance Al/Fe	0.2610	0.0847	3.081	0.0032	0.001
SPB ilr base	Intercept	-5.6935	4.2771	-1.331	0.1909	
	ilr17 Na/(Al, As, Cd, Cr, Cu, K, Mn, Mo, Ni, Pb, Se, U, V, Zn, Ca, Fe, Mg)	1.4652	0.6212	2.359	0.0234	0.05
ilr balanced	ilr	No balances found to be significant				
Basalt bedrock (reordered elements) log(CKDu SIR)						
Selbal	Intercept	1.5525	0.1457	10.654	4.38E-15	<0.001
	Global Balance Al/Fe	0.26099	0.0847	3.081	0.0032	0.001
SPB ilr base	Intercept	-5.6935	4.2771	-1.331	0.1909	
	ilr3 Na/(Al,Fe,Cr)	1.2603	0.5515	2.285	0.0278	0.01
	ilr4 Mo/(Al,Fe,Cr,Na)	-0.5833	0.2511	-2.323	0.0255	0.01
ilr balanced	Intercept	-5.6935	4.2771	-1.331	0.1909	
	ilr2 (Na/Cr)	1.0755	0.4870	2.208	0.0332	0.05

Results

Regional geology study

For the regional study using the *selbal* approach, the elements most frequently identified as the global balance in the CV procedure, and therefore most associated with log(CKDu SIR) were As appearing 90% and Al appearing 74% of the time (Fig. 2A and b). In addition to the global balance of Al/As, the balances of Na/As and Na/Mo were also identified in the CV procedure. Only the global balance Al/As was found to be statistically significant (Table 1, Fig. 3: significance level <0.001). The balance between As and Al was also identified using the SBP (ilr base) and the ilr balanced approaches (Table 1; both showed a significance level of 0.1). To assess the impact of spatial autocorrelation on the regression models, Moran's *I* test statistic was used to test the residuals computed from the regression models. The Moran's Index for the residuals was found to be not significantly different from random (Moran's Index -0.008782; *z*-score: 0.009158; *p*-value: 0.992693).

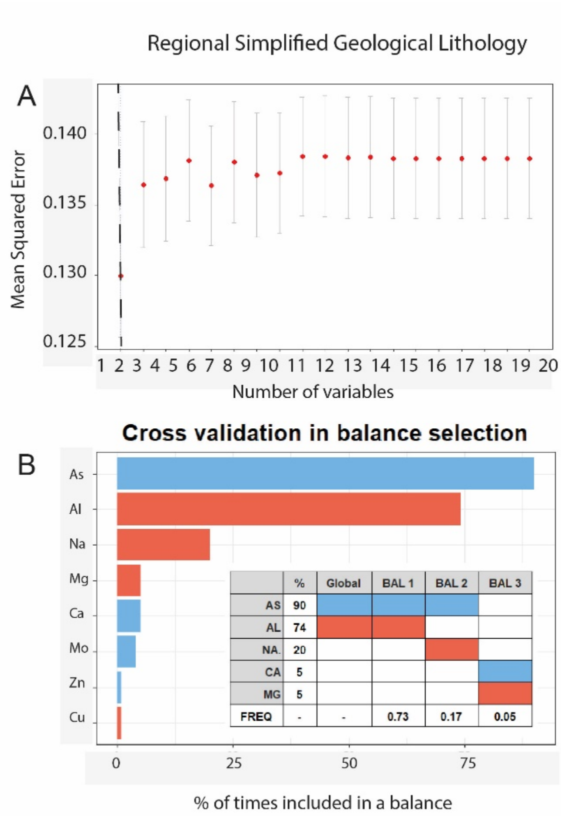

Fig. 2. Results of the forward-selection method using the *selbal* algorithm for (Super Output Areas (SOAs)) Standardised Incidence Rates (SIRs) of Chronic Kidney Disease of uncertain aetiology (CKDu), for regional study A) Mean squared error (MSE) as a function of the number of components included in the balance. The optimal number of components is highlighted with a vertical dashed line; B) The balance identified with the whole dataset is the most frequently identified in the cross-validation (CV) procedure and global balance and other balances identified in CV procedure.

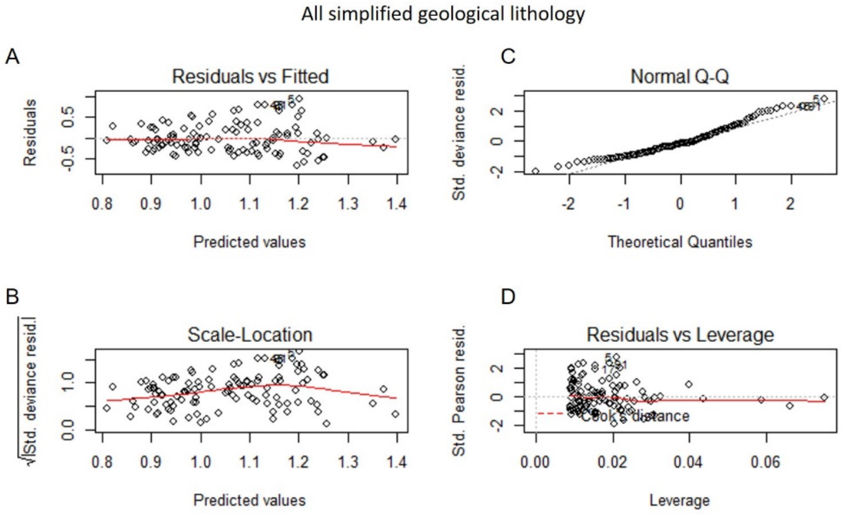

Fig. 3. Generalized Linear model regression (GLM) summary for log(SIR_CKDu) with regional stream water data (all simplified geological lithology) for global balance Al/As; A) Residual vs Fitted; B) Normal Q-Q plot; C) Scale Location and D) Residuals vs Leverage.

Stratified study for basalt bedrock

Following stratification by basalt bedrock geology, a different association is indicated between log(CKDu SIR) and stream water geochemistry. The results for log(CKDu SIR) identified the balances Al/Fe (global balance with Al appearing 92% and Fe appearing 91% of the time, respectively), Pb/Fe and Na/Mo most frequently in the CV procedure (Fig.4C and D). Regression (using GLM) suggested a correlation between log(CKDu SIR) and the global balance Al/Fe (Table 1, Fig 5; significance level of 0.001). To assess the impact of spatial autocorrelation on the regression models, Moran's *I* test statistic was used to test the residuals computed from the regression models. The Moran's Index for the residuals was found to be not significantly different from random (Moran's Index -0.008782; *z*-score: 0.009158; *p*-value: 0.992693).

Using the SBP (ilr base) approach only a large balance which included all elements against Na was found to be significant (level 0.05) and no statistically significant relationship was found between log(CDKu SIR) and any elemental balances using the ilr balanced approach. The composititional dendrograms shown for the SBP (ilr base) and ilr balanced approach (Fig 6A and 6B respectively) show the difference between the two ilr approaches on partitioning of the elements.

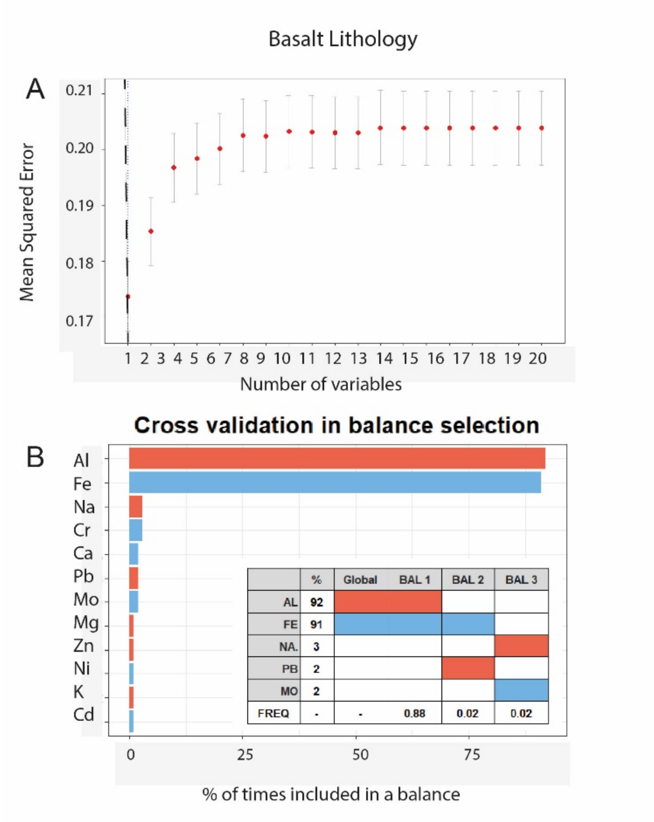

Fig. 4. Results of the forward-selection method using the *selbal* algorithm for (Super Output Areas (SOAs)) Standardised Incidence Rates (SIRs) of Chronic Kidney Disease of uncertain aetiology (CKDu), for basalt bedrock A) Mean squared error (MSE) as a function of the number of components included in the balance. The optimal number of components is highlighted with a vertical dashed line; B) The balance identified with the whole dataset is the most frequently identified in the cross-validation (CV) procedure and global balance and other balances identified in CV procedure.

To test the impact of the ordering of elemental variables in the analysis, the balances identified using the *selbal* approach were used (a knowledge-driven approach) to reorder the elements and all CoDA balance approaches were investigated. The elements were re-ordered as Al, Fe, Cr, Na, Mo, Cu, K, Mn, Ni, Pb, Se, U, Zn, V, As, Cd, Ca, Mg and Si. The results obtained using the *selbal* approach did not change with all balances and associated statistical significance remaining the same (significance level 0.001 for global balance Al/Fe). The dendrograms shown for the SBP (ilr base) and ilr balanced approach (Fig 6C and 6D respectively) demonstrate the impact for reodering the elements on the partitioning process. Reordering the elements using the SBP approach identified a statistically significant relationship with log(CKDu SIR) for the balances $\sqrt{\frac{3}{4}}\log\left(\frac{Na}{(Al\times Fe\times Cr)^{1/3}}\right)$ and $\sqrt{\frac{4}{5}}\log\left(\frac{Mo}{(Al\times Fe\times Cr\times Na)^{1/4}}\right)$ (Table 1 SBP ilr 3 and ilr 4; both showed a significance level of 0.05). Using the ilr balanced approach, only the balance of Na/Cr had a significant relationship with log(CKDu SIR) (significance level 0.05).

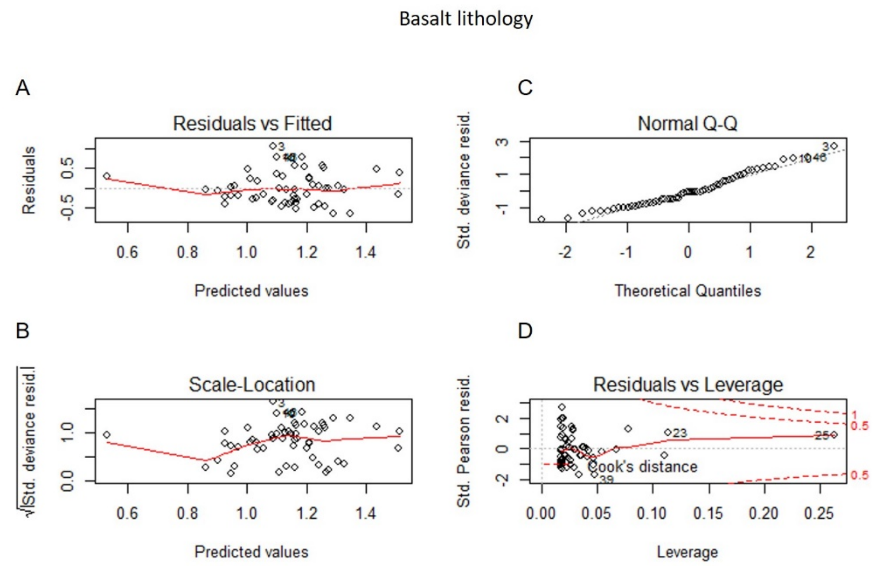

Fig. 5. Generalized Linear model regression (GLM) summary for log(SIR_CKDu) with stream water data for basalt lithology (bedrock) for global balance Al/Fe; A) Residual vs Fitted; B) Normal Q-Q plot; C) Scale Location and D) Residuals vs Leverage.

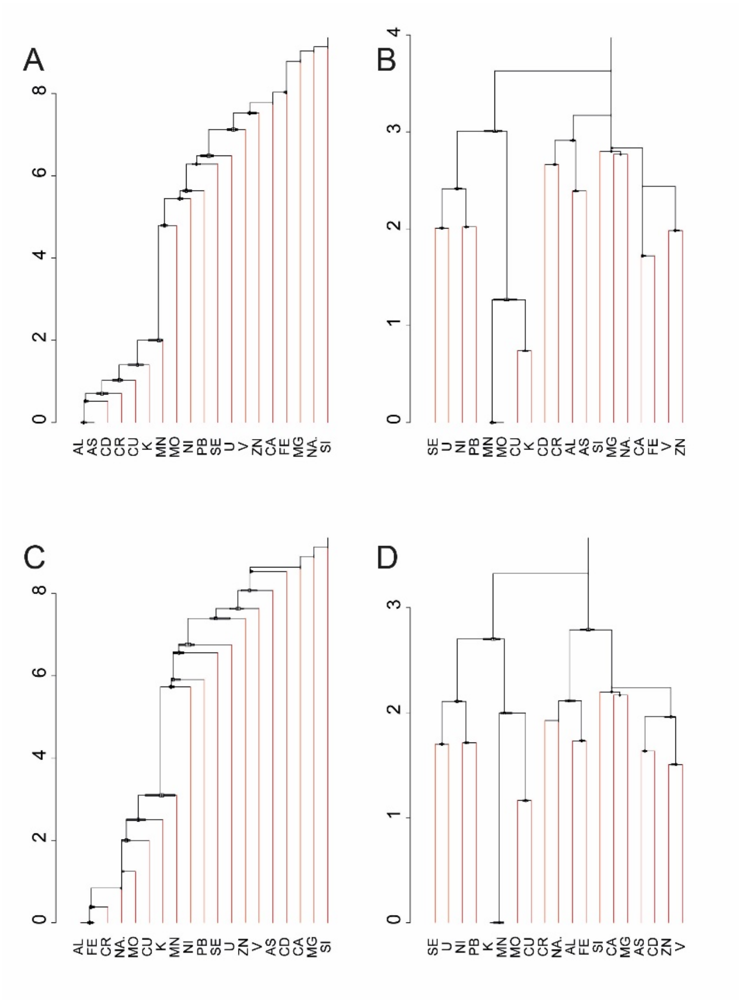

Fig. 6. Compositional dendrograms for stream water geochemistry data for basalt bedrock showing A) SBP (ilr base), B) balanced ilr approach; for reordered elements C) SBP (ilr base) and D) balanced ilr approach.

Discussion

The regional distribution of Al in Northern Ireland is described as directly controlled by geology and terrain [27]. High relative concentrations of Al are associated with pH values >7 and high levels of dissolved organic carbon, both promoting Al in solution in stream waters. Superficial deposits of peat with lower pH values and high dissolved organic carbon are likely to be responsible for enhancing the solubility of Al. Likewise, the relative concentration of As in stream waters is strongly controlled by bedrock geology and the composition of superficial deposits [26]. Higher relative amounts of As are found in the Dalradian supergroup and over Devonian and Carboniferous bedrock (Fig.1). Lower relative amounts are found associated with the Palaeogene Mourne mountain complex and Antrim Lava group. The elemental balance identified in the *selbal* approach categorizes the diversity in the geological foundation of Northern Ireland and the impact of superficial deposits. Further research into the association between CKDu SIR, bedrock geology and superficial deposits through these preliminary findings cannot be confirmed without controlling for other effects (e.g., soil PTEs and socio-economic factors; [24]). However, the significant association found between log(CKDu SIR) and the elemental balance of Al/As in stream waters may indicate that, where indirect or direct ingestion of trace elements occurs through waterborne exposure pathways, there is a potential health risk in terms of CKDu. A major factor in waterborne exposure pathways is whether households have access to public water supply though the water mains network or via private supplies [44]. The distribution of registered private water supplies as reported by NIEA [44] coincides well with basalt bedrock. The distribution of registered private water supplies across Northern Ireland and especially in the northwest of the country suggests that in these areas there is greater potential for indirect or direct ingestion of PTEs. Groundwater flow, residence time, topography, climate and land use are all controls on stream water chemistry [45]. However, the chemical composition of the bedrock and superficial geology of the different environments, through which the streams pass, will impact stream waters. Therefore, it is a reasonable assumption that stream water geochemistry can provide an indication of the chemistry of private wells. An elemental balance of Al/Fe in stream waters is identified as significantly associated with log(CKDu SIR) for areas underlain by basalt bedrock. These elements are considered essential for human health at different levels (Al can be considered essential at low levels). However soluble Al is also known to be a risk for human health including patients on dialysis. The Palaeocene Antrim lava Group on the Antrim Plateau shows a wide variation in relative Al concentrations in stream waters, whereas Fe concentrations are found to be higher over the Upper Basalt Formation. Higher relative concentrations for both Al and Fe are recorded for streams draining the outcrop of Dalradian metasediments in northeast Antrim. The pH of soils and land through which the waters pass and the organic carbon of the cover are major factors in the solubility and mobility of elements [46]. Higher relative concentrations of Al on elevated areas are coincident with upland peat cover with lower pH values which enhances the solubility of Al in stream waters. Fe has relatively low mobility under most environmental conditions but is affected by redox conditions [47]. This can result in very high dissolved Fe loadings in streams draining peatland or other acid reducing situations [46]. Iron coatings on minerals can exert controls on the surface geochemistry of other elements including As and Mo [48]. Previous studies have considered the chemical factors controlling Al bioavailability and toxicity in aquatic environments and the importance of multi-element interactions [49]. The role of superficial deposits (with peat providing higher dissolved organic carbon and reducing pH and Quaternary glacial tills moderating the solubility of Al) as a control on the balance of Al and Fe and any potential link with CKDu needs to be explored further.

There has been much debate on how to interpret compositional balances [23,24,30,31,33,34,35]. Several considerations on the interpretation of compositional balances have arisen from this research. It is widely recognized that while a data-driven approach for exploratory data analysis may be helpful, the ordering of elements for SBP and an ilr balanced approach does impact the results. The results from this research suggest that one should consider carefully the ordering of elements and use an informed knowledge driven approach when using a SBP ilr base approach. Subtle changes in the dendrogram were shown in the balanced ilr approach (Fig 6). This research found the *selbal* forward selection approach to provide consistent and interpretable results to identify 'the most significant balances' most associated with log-transformed SIR CKDu. While it is acknowledged that further work needs to be done controlling for other factors, such as exposure pathways, spatial dependence, environment and social economic demographics, the balance approach shown provides a direction for further research. The approach provides a framework for investigating the importance of multi-element interactions of both essential and potentially toxic trace elements in understanding the environmental clusters of CKDu.

Conclusions

The relationship between standardized incidence rates (SIRs) of CKDu and stream water geochemistry was explored through the use of log-ratio balances of potentially toxic elements (PTEs). Several CoDA approaches using balances (a forward selection procedure using a selbal algorithm, SBP and ilr balanced approaches) were used to identify the elemental balances most associated with log-transformed SIR CKDu. At the regional scale, a statistically significant relationship was found between log(CKDu SIR) and the elemental balance Al/As. This elemental balance characterizes different geological domains across Northern Ireland. Following stratification by basalt bedrock (the dominant bedrock geology for SIR CKDu), the balance Al/Fe was identified as significantly associated with log(CKDu SIR). Controls on the relative amounts of Al and Fe in stream waters include bedrock geology, superficial deposits, amount of dissolved organic carbon and pH. The findings suggest that bedrock geology and superficial deposits play an important role as controls in the availability of environmental toxins with a potential health risk in terms of CKDu, where indirect or direct ingestion of trace elements occurs through waterborne exposure pathways.

Acknowledgments

The study received ethics approval in March 2018, 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. 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. Chemical analysis of the urban soil samples was co-financed by the Northern Ireland Environment Agency. JJE received financial support through the project METhods for COmpositional analysis of DAta (CODAMET), Ministerio de Ciencia, Innovación y Universidades, Spain (Ref: RTI2018-095518-B-C21/C22, 2019-2021). We would like to thank our two reviewers Antonella Buccianti, Ricardo Olea and Editor Peter Filzmoser for their helpful comments which have improved the research.

References

1 World Health Organisation (1996) Trace elements in human nutrition and health WHO Library Cataloguing in Publication Data Geneva ISBN 92 4 156173 4

2 Lindeman R, Tobin J, Shock NW (1985) Longitudinal studies on the rate of decline in renal function with age. J. Am. Geriatr. Soc. 33:278–285

3 Couser WG, Remuzzi G, Mendis S, Tonelli M (2011) The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. Kidney International 80(12):1258-70. doi: 10.1038/ki.2011.368.

4 Musso C, Oreopoulos D (2011) Aging and physiological changes of the kidneys including changes in glomerular filtration rate. *Nephron Physiol*. 119, p1–5.

5 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.

6 Lewis R (2012) Understanding Chronic Kidney Disease: A guide for the non-specialist. England: M&K Update Ltd.157p.

7 Chiu HF, Yang CY (2005) Decreasing trend in renal disease mortality after cessation from arsenic exposure in a previous arseniasis-endemic area in southwestern Taiwan. J Toxicol Environ Health A., 68(5):319–27

8 Zheng L, Kuo CC, Fadrowski J, Agnew J, Weaver VM, Navas-Acien, A (2104) Arsenic and Chronic Kidney Disease: A Systematic Review. *Curr Environ Health Rep* 1(3):192–207

9 Jayatilake N, Mendis S, Maheepala P, Mehta FR (2013) Chronic kidney disease of uncertain aetiology: prevalence and causative factors in a developing country. *BMC Nephrol.*14:180.

10 Jayasumana, C, Gajanayake R, Siribaddana S (2014) Importance of Arsenic and pesticides in epidemic chronic kidney disease in Sri Lanka. *BMC Nephrol*.15(1):124

11 Correa-Rotter R, Wesseling C, Johnson RJ (2014) CKD of unknown origin in Central America: the case for a mesoamerican nephropathy. Am J Kidney Dis 63(3):506–520

12 Zheng LY, Umans JG, Yeh F, Francesconi KA, Goessler W, Silbergeld EK, et al. (2015) The association of urine arsenic with prevalent and incident chronic kidney disease: evidence from the strong heart study. *Epidemiology*. 26(4):601–12

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

14 Lunyera J, Mohottige D, Von Isenburg M, Jeuland M, Patel UD, Stanifer JW (2016) CKD of uncertain etiology: a systematic review. Clin J Am Soc Nephrol 11(3):379–385

15 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/ Accessed October 30th 2019.

16 Soderland P, Lovekar S, Weiner DE, Brooks DR, Kaufman JS (2010) Chronic Kidney Disease associated with environmental toxins and exposures. *Advances in Chronic Kidney Disease*, 17(3): pp 254-264

17 Orem WH, Calin CA (2001) Health effects of toxic organic compounds from coal—The case of Balkan endemic nephropathy (BEN). U.S. Geological Survey Fact Sheet 004-01, 4 pp., https://pubs.usgs.gov/fs/fs004-01/

18 Barsby A, McKinley JM, Ofterdinger U, Young M, Cave MR, Wragg J (2012) Bioaccessibility of trace elements in soils in Northern Ireland. *Science of the Total Environment* 433:398–417 http://dx.doi.org/10.1016/j.scitotenv.2012.05.099.

19 Cox S, Chelliah M, McKinley JM, Palmer S, Ofterdinger U, Young M, Cave MR, 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. Environmental Geochemistry and Health 35(5): 53-567 http://dx.doi.org/10.1007/s10653-013-953

20 McKinley JM, Ofterdinger U, Young M, Barsby A, Gavin A (2013) Investigating local relationships between trace elements in soils and cancer data. *Spatial Statistics* 5: 25-41 http://dx.doi.org/10.1016/j.spasta.2013.05.001

21 Palmer S, McIlwaine R, Ofterdinger U, Cox SF, McKinley JM, Doherty R, Wragg J, Cave M (2015) The Effects of Lead Sources on Oral Bioaccessibility in Soil and Implications for Contaminated Land Risk Management. Environmental Pollution 198:161-171. 10.1016/j.envpol.2015.01.004

22 Jackson CE, McKinley JM, Ofterdinger U, Fogarty D, Atkinson PM, Palmer S (2016) Investigating relations between environmental toxins in northern Irish soils and streams and Chronic Kidney Disease prevalence. Applied Geochemistry 75: 236-246. DOI: 10.1016/j.apgeochem.2016.10.016

23 McKinley JM, Mueller U, Atkinson PM, Ofterdinger U, Jackson, C., Cox, S F., Doherty, R., Fogarty, D., Egozcue, J.J., Pawlowsky-Glahn, V. (2020) Investigating the influence of environmental factors on the incidence of renal disease with compositional data analysis using balances. Applied Computing and Geosciences, vol 6, 100024, https://doi.org/10.1016/j.acags.2020.100024

24 McKinley JM, Mueller U, Atkinson PM, Ofterdinger U, Jackson C, Cox SF, Doherty R, Fogarty D, Egozcue JJ, Pawlowsky-Glahn V (2020) Chronic kidney disease of unknown origin is associated with environmental urbanisation in Belfast, UK. Environmental Geochemistry and Health, https://link.springer.com/article/10.1007/s10653-020-00618-y

25 Cooper M, Mitchell I (2013) Geology. In Young ME, Donald AW (eds). A guide to the Tellus data. Geological Survey of Northern Ireland (GSNI), Belfast, pp 5-20

26 Mitchell W (2004) The Geology of Northern Ireland (2nd edition), Geological Survey of Northern Ireland.

27 Young ME, Donald AW (2013) A guide to the Tellus data. Geological Survey of Northern Ireland (GSNI), Belfast, http://nora.nerc.ac.uk/509171/.

28 Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) Isometric logratio transformations for compositional data analysis. Mathematical Geology 35 (3):279-300

29 Egozcue JJ, Pawlowsky-Glahn, V (2005) Groups of parts and their balances in compositional data analysis. Mathematical Geology 37(7):795-828.

30 Aitchison J, Greenacre M (2002) Biplots for compositional data. Journal of the Royal Statistical Society. Series C (Applied Statistics) 51(4): 375-392

31 Buccianti A, Pawlowsky-Glahn V (2005) New perspectives on water chemistry and compositional data analysis. Mat Geol 37(7): 703-727. DOI: 10.1007/s11004-005-7376.

32 McKinley J, Tolosana Delgado R, Hron K, de Caritat P, Grunsky E, Reimann C, Filzmoser P, van den Boogaart KG (2016) Single Component map: Fact or Fiction? Journal of Geochemical Exploration 162: 16-28. http://dx.doi.org/doi:10.1016/j.gexplo.2015.12.005

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

34 Gozzi C, Sauro Graziano R, Frondini F, Buccianti A (2018) Innovative monitoring tools for the complex spatial dynamic of river chemistry: case study for the Alpine Region. Environ Earth Sci 77(16): 579, 1-11. DOI: 10.1007/s12665-018-7756-0. 319

35 Sauro Graziano R, Gozzi C, Buccianti A (2020) Is Compositional Data Analysis a theory able to discover complex dynamics in aqueous geochemical systems? J Geochem Explor 211: 106465. DOI:10.1016/j.gexplo.2020.106465.

36 Engle MA, Rowan EL (2013) Interpretation of Na-Cl-Br systematics in sedimentary basin brines: Comparison concentration, element ratio, and isometric log-ratio approaches. Mathematical Geosciences 45(1):87−101.

37 Shelton JL, Engle MA, Buccianti A, Blondes MS (2018) The isometric log-ratio (ilr)ion plot: A proposed alternative to the Piper diagram. Journal of Geochemical Exploration 190:130−141.

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

39 Rivera-Pinto J (2018) Statistical methods for the analysis of microbiome compositional data in HIV infection, Unpublished PhD Thesis in Experimental Sciences and Technology, Universitat Central De Catalunya, Spain.

40 Hastie T, Tibshirani R, Friedman J, (2009) The Elements of Statistical Learning; Data Mining, Inference and Prediction, Springer. Second edition.

41 Van den Boogaart KG, Tolosana-Delgado R (2008) compositions: a unified R package to analyze Compositional Data, *Computers & Geosciences* 34(4): 320-338, doi:10.1016/j.cageo.2006.11.017.

42. Pardo-Iguzquiza E, Dowd PA (2004) Normality test for spatially correlated data. Math Geol 36(6):659–681.

43. Olea RA, Pawlowsky-Glahn V (2009) Kolmogorov–Smirnov test for spatially correlated data Stoch Environ Res Risk Assess 23:749–757, doi: 10.1007/s00477-008-0255-1

44 Northern Ireland Environment Agency (NIEA) (2019) Drinking Water Quality in Northern Ireland, 2018 A Report by the Drinking Water Inspectorate for Northern Ireland, ISBN 978-1-83887-020-1

45 Smyth D, Johnston C (2013) Geochemistry: Methodlogy. In Young ME, Donald AW (eds). A guide to the Tel-lus data. Geological Survey of Northern Ireland (GSNI), Belfast, pp 33-44.

46 Brewer N, Chenery N (2013) Geochemistry. In Young ME, Donald AW (eds). A guide to the Tellus data. Geological Survey of Northern Ireland (GSNI), Belfast, pp 47-54.

47 Hylander LD, Meili M, Oliveira LJ, Silva EDE, Guimaraes JRD, Araujo DM, Neves RP, Stachiw R, Barros AJP, Silva GD (2000) Relationship of mercury with aluminum, iron and manganese oxy-hydroxides in sediments from the Alto Pantanal, Brazil. Science of the Total Environment, 260, 97–107.

48 Williams TM, (1991) A sedimentary record of the deposition of heavy metals and magnetic oxides in the Loch Dee basin, Galloway, Scotland since c. AD 1500. Holocene, 1, 142–150.

49 Gensemer RW, Playle RC (1999) The Bioavailability and Toxicity of Aluminum in Aquatic Environments, Critical Reviews in Environmental Science and Technology 29(4): 315-450, DOI: 10.1080/10643389991259245