Geoscience meets Data science: Researcher Links Workshop

Data, data everywhere!

Data driven challenges in Geoscience

Prof Jenny McKinley

Geography
Director: Centre for GIS and Geomatics, School of Natural and Built Environment, Queen's University Belfast, UK

Councilor IUGS

Vice President, Governing Council DDE

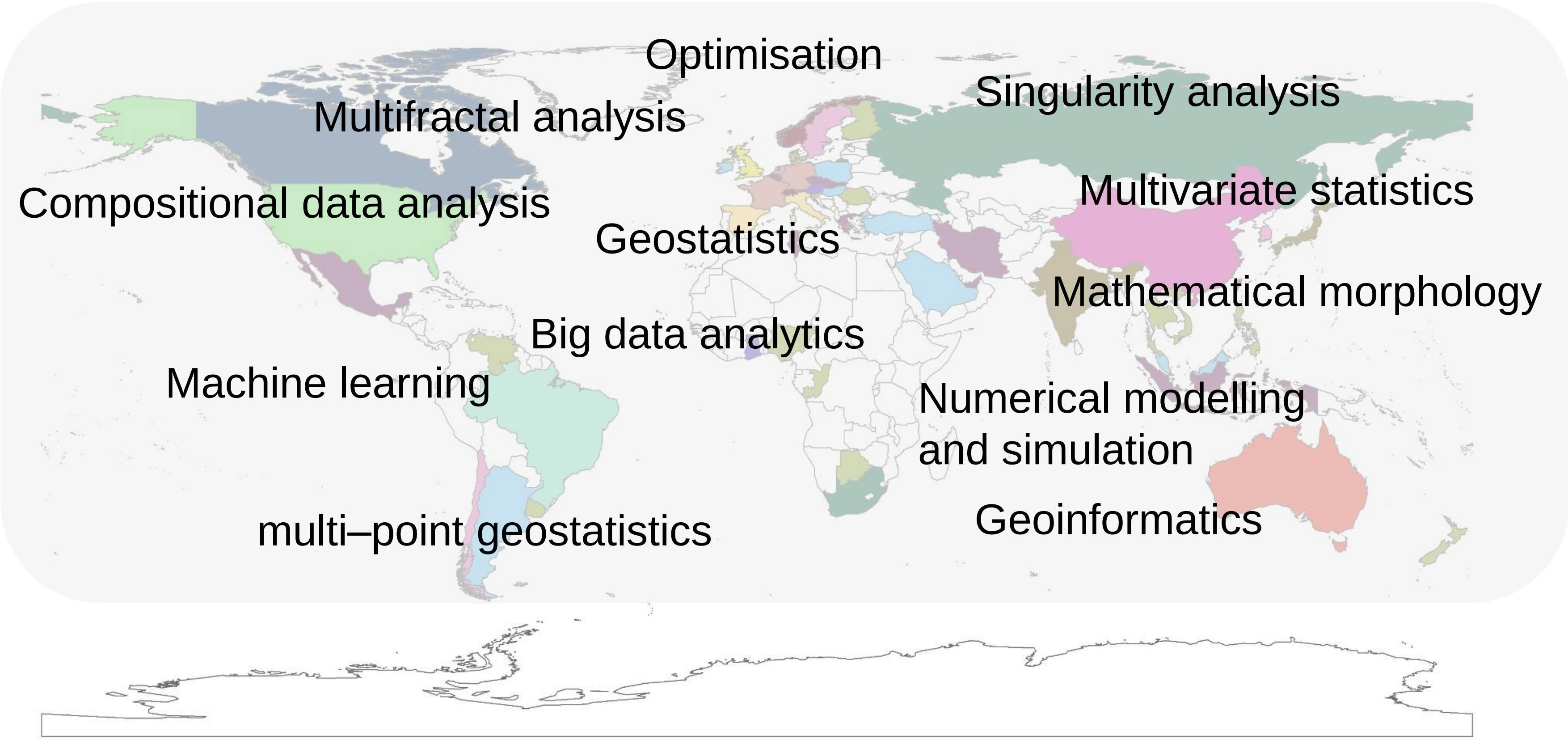
Past President of the International Association of Mathematical Geoscientists (IAMG)

QUEEN'S UNIVERSITY BELFAST

The IAMG - A Global Network

Critical problem solving with mathematical geoscience

https://iamg.org/



The mission of the IAMG is to promote worldwide the advancement of mathematics, statistics and informatics in the geosciences.

IUGS

IUGS - International Union of Geological Sciences

Home What is IUGS? Organization Activities Documents Publications Calendar Contact IUGS

Deep-time Digital Earth

- Earth Science machine learning and AI moving ahead but with big, structured data
- Access to long tail geoscience data more difficult

REVIEW SUMMARY

GEOPHYSICS

Machine learning for data-driven discovery in solid Earth geoscience

Karianne J. Bergen, Paul A. Johnson, Maarten V. de Hoop, Gregory C. Beroza*

PERSPECTIVE

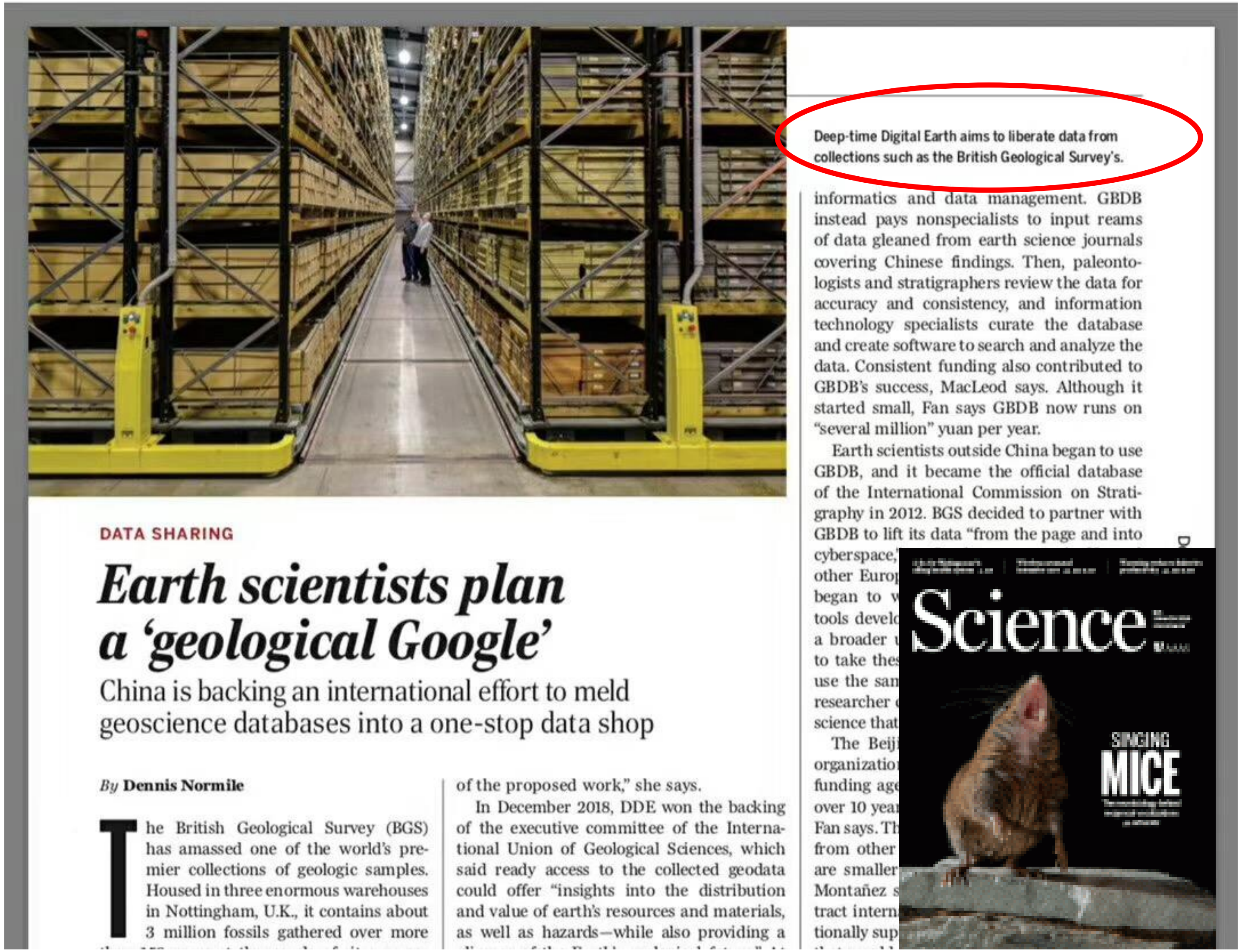
https://doi.org/10.1038/s41586-019-0912-1

Deep learning and process understanding for data-driven Earth system science

Markus Reichstein[1,2]*, Gustau Camps-Valls[3], Bjorn Stevens[4], Martin Jung[1], Joachim Denzler[2,5], Nuno Carvalhais[1,6] & Prabhat[7]

Mission and vision

- Mission: harmonise global Deep-time Digital Earth data and share global geoscience knowledge
- Vision: Transform Earth science



Deep-time Digital Earth aims to liberate data from collections such as the British Geological Survey's.

DATA SHARING

Earth scientists plan a 'geological Google'

China is backing an international effort to meld geoscience databases into a one-stop data shop

By **Dennis Normile**

The British Geological Survey (BGS) has amassed one of the world's premier collections of geologic samples. Housed in three enormous warehouses in Nottingham, U.K., it contains about 3 million fossils gathered over more

of the proposed work," she says.

In December 2018, DDE won the backing of the executive committee of the International Union of Geological Sciences, which said ready access to the collected geodata could offer "insights into the distribution and value of earth's resources and materials, as well as hazards—while also providing a

informatics and data management. GBDB instead pays nonspecialists to input reams of data gleaned from earth science journals covering Chinese findings. Then, paleontologists and stratigraphers review the data for accuracy and consistency, and information technology specialists curate the database and create software to search and analyze the data. Consistent funding also contributed to GBDB's success, MacLeod says. Although it started small, Fan says GBDB now runs on "several million" yuan per year.

Earth scientists outside China began to use GBDB, and it became the official database of the International Commission on Stratigraphy in 2012. BGS decided to partner with GBDB to lift its data "from the page and into cyberspace,"

Science, Feb. 26, 2019

Talk Outline

- ***Effective decision-making*** to support resource assessment, environmental management and understand the implications for human health practices, in sustainable and cost-effective ways, requires ***spatial data analysis and robust informed insight***.
- We will discuss the different types of geoscience data available – from remotely sensed data to ground sampled soils and waters.
- What are the ***data analytical challenges*** we face using ‘big’ geoscience data?
- What are the ***opportunities for greater use of machine learning techniques*** to complement current spatial data analysis approaches to develop the geoscientist’s toolkit towards a ***more integrated, insightful interpretation.***

Prof Mark Cooper, Geological Survey Northern Ireland states

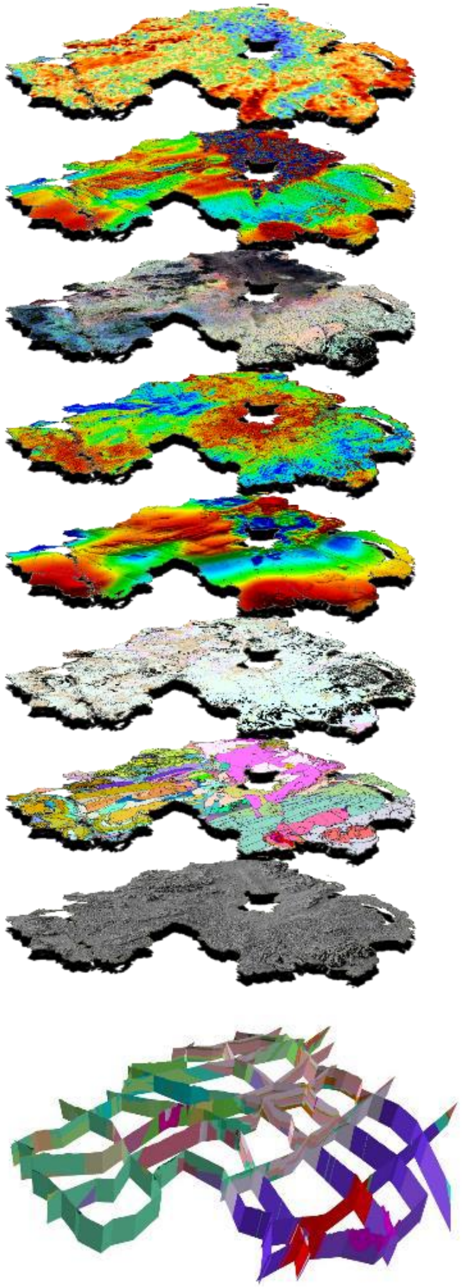
(geoENV 2018, Belfast, UK 2018)

Geological Survey in the form of mapping, research and modelling, underpin and is required to make informed decisions on for example, resource exploration, environmental assessment, urban and infrastructure development.

We face growing, global-scale impacts of climate change, through sea-level rise, changes in seasonal patterns, increased frequency and severity of extreme weather events, combined with the continuing issue of population growth.

In the future how will countries manage loss of land, resources, major cities and infrastructure in coastal and increased geohazard areas?

What resources will be needed to rehome displaced populations?



Rialtas na hÉireann
Government of Ireland

Geological Survey
Suirbhéireacht Gheolaíochta
Ireland | Éireann

Dr Mairéad Fitzsimons, Geological Survey Ireland describes the long-term stakeholder product strategy

(geoENV 2018, Belfast, UK 2018)

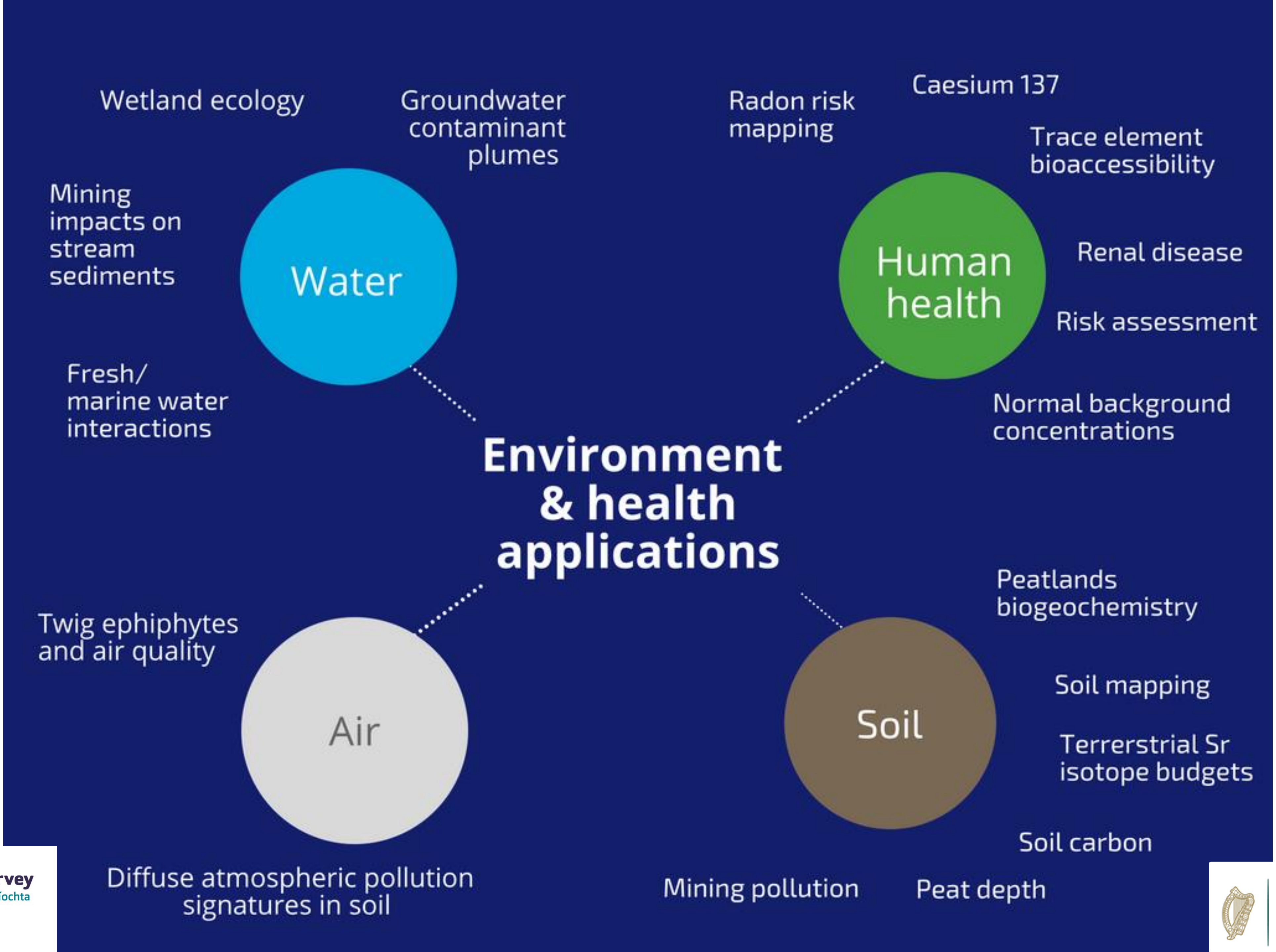
New workstream in Tellus with dedicated resources to produce highly applied, user-focused products
Geostatistics, spatial data analysis and AI will play a key role in product development; adding value to raw data by understanding underlying processes and multivariate relationships in space

tellus@gsi.ie

@TellusGSI on twitter

www.tellus.ie

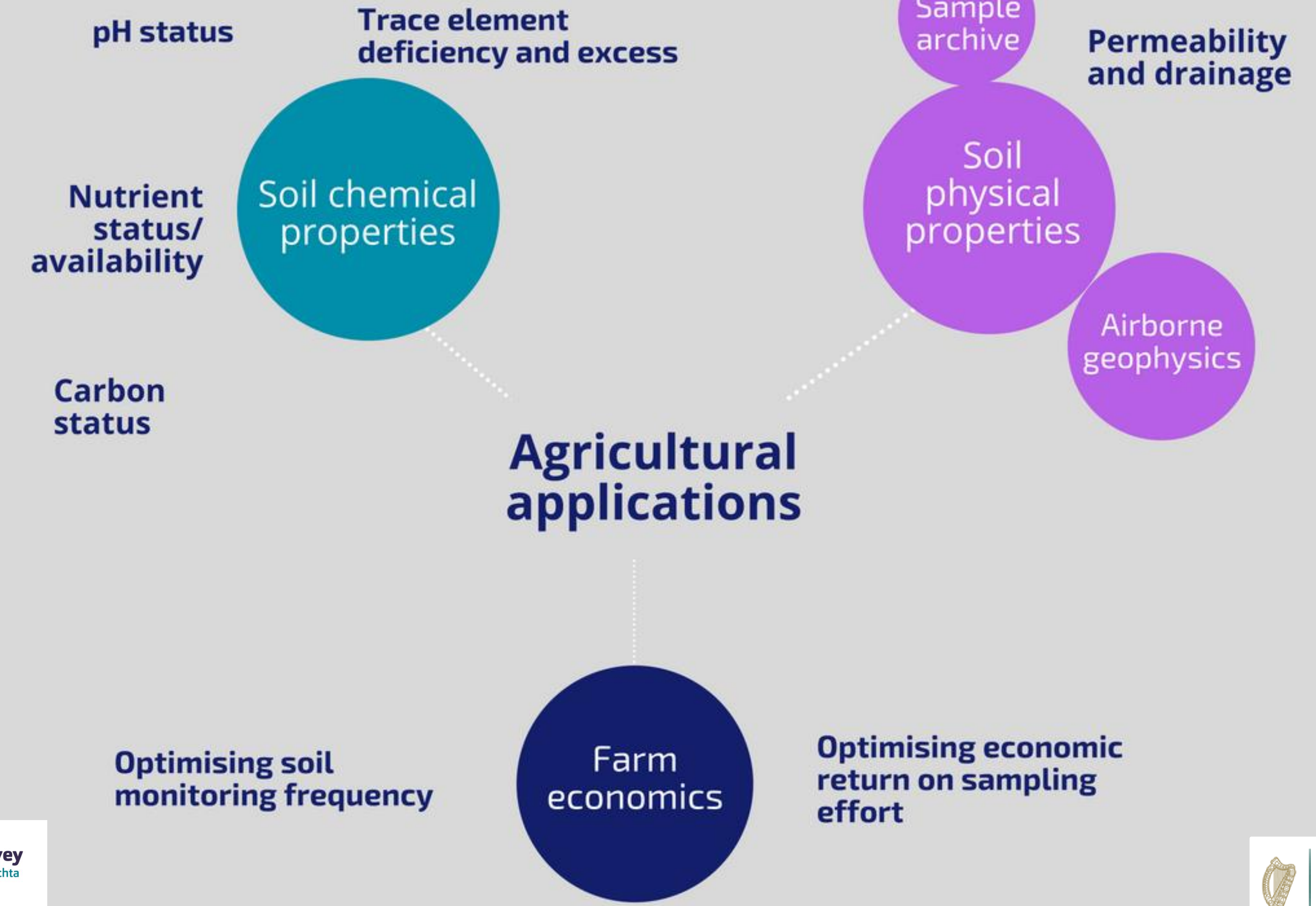
www.gsi.ie



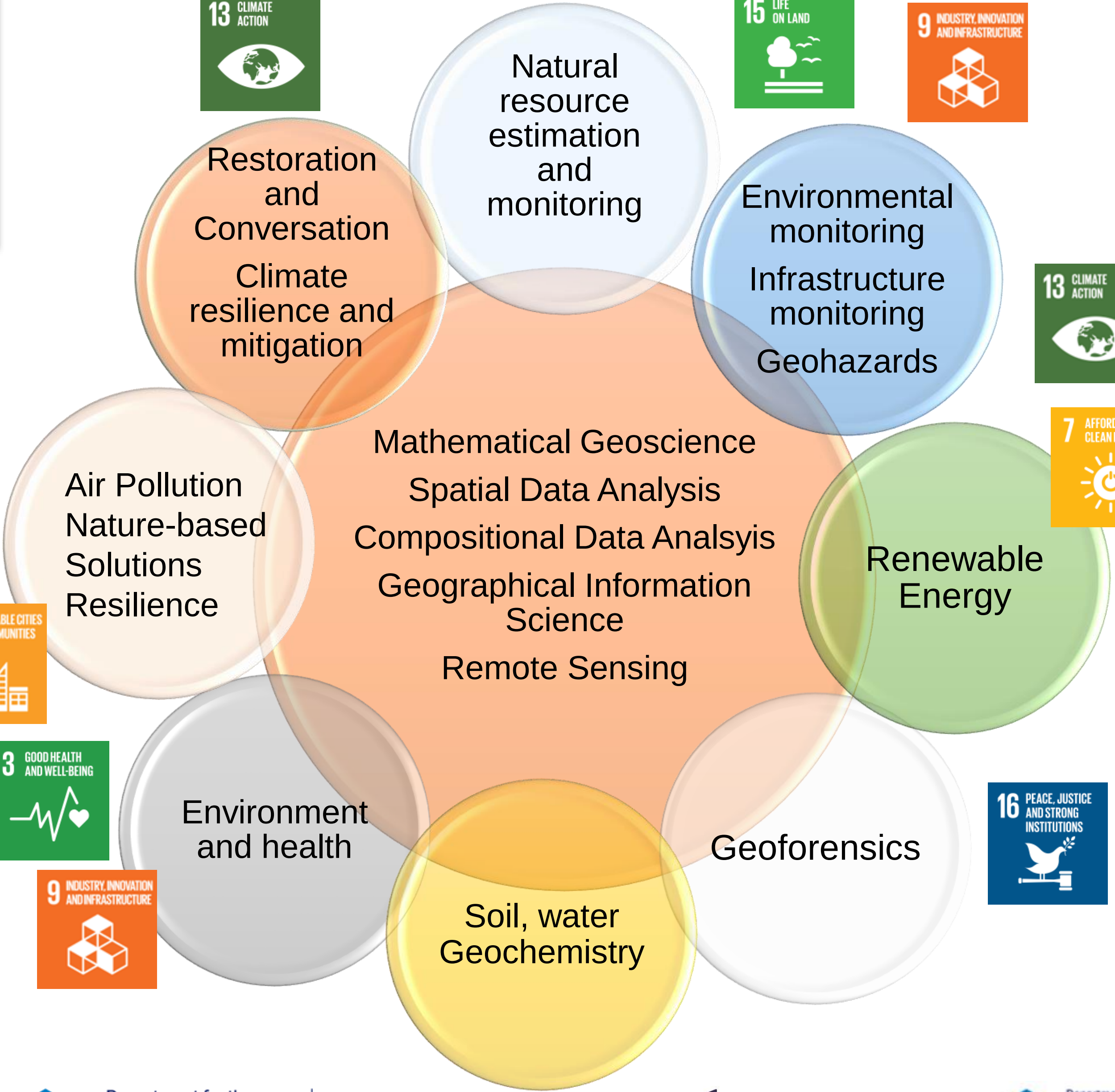
Geological Survey
Suirbhéireacht Gheolaíochta
Ireland | Éireann

Roinn Cumarsáide, Gníomhaithe ar son na hAeráide & Comhshaoil
Department of Communications, Climate Action & Environment

Agricultural applications



NERC DTP
Doctoral Training Programme



What are the different types of geoscience data?

Department for the Economy www.economy-ni.gov.uk | GSNI Geological Survey of Northern Ireland Celebrating 70 years: 1947–2017

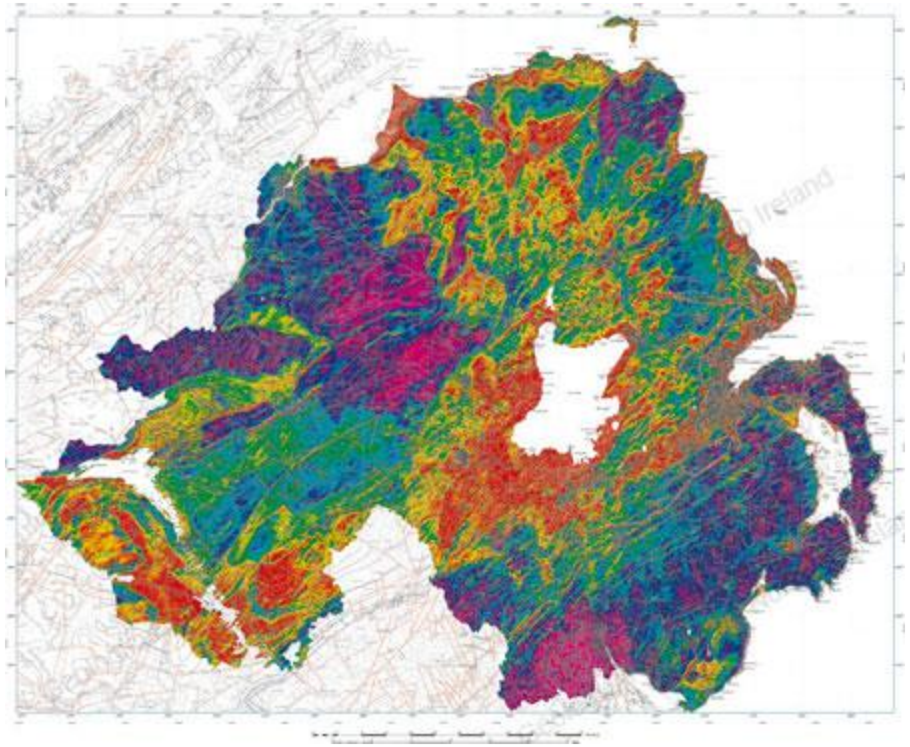
The Tellus Project

– Managed by the Geological Survey of Northern Ireland (GSNI) and funded by the Department of Enterprise Trade and Development

Magnetics

Natural radioactivity

Electrical conductivity



- National scale injection of geoscience data 2004-2011
- The data comprise multi-source airborne geophysics collected by a specialist survey aircraft
  – Magnetics
  – Natural radioactivity
  – Electrical Conductivity

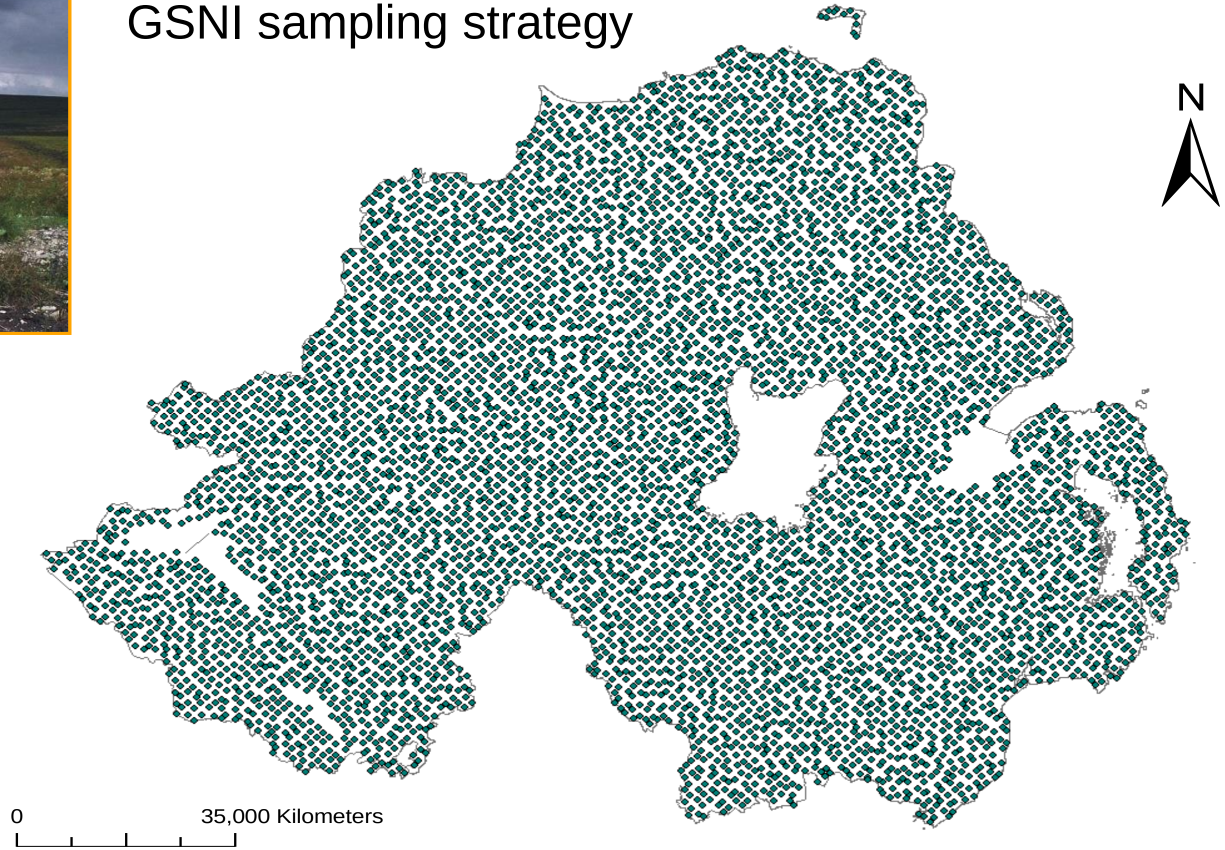
http://www.bgs.ac.uk/gsni/tellus/map_viewer/application/magnetics_tmi.html

#GSNI70

Tellus Project - soil geochemistry data

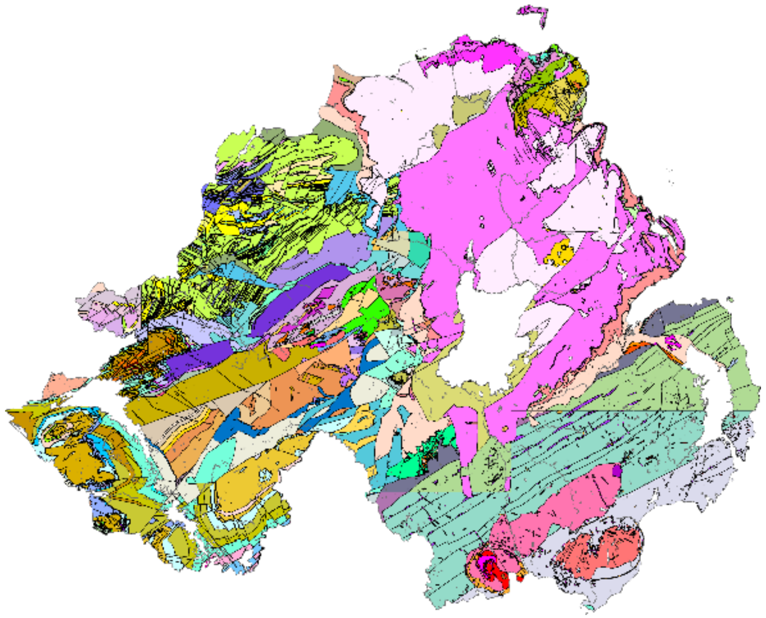
- Collection of soil, stream-sediment and stream water samples from 5 – 20 cm depth in rural and urban areas.
- 1 collated sample collected at 2km^2 intervals)

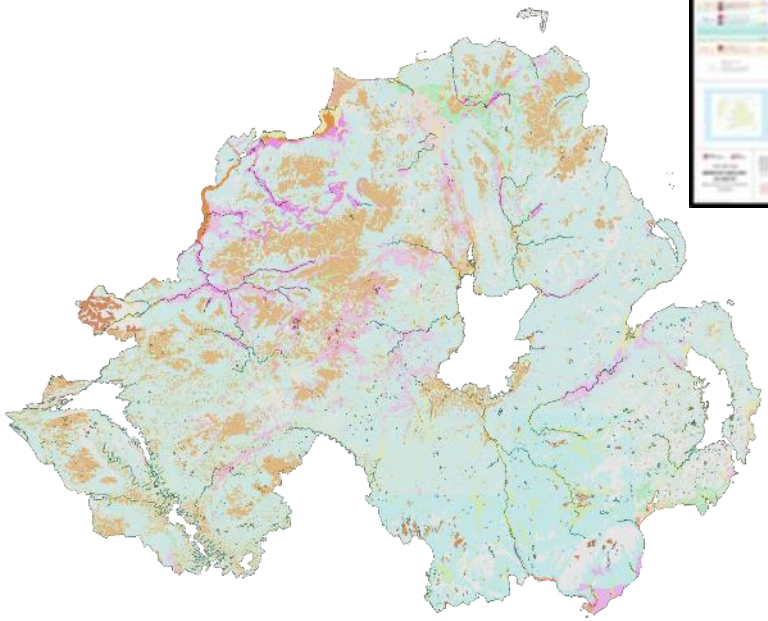
Tellus Survey
GSNI sampling strategy



Geological mapping

- Bedrock
- Superficial deposits
- Various scales, UK & Ireland 1:1,250,000 to 10K NIDigMap



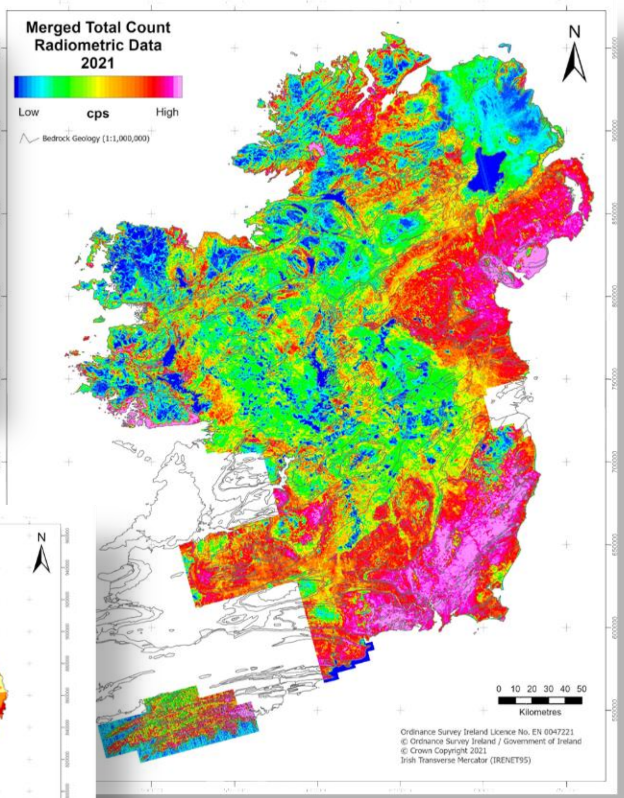


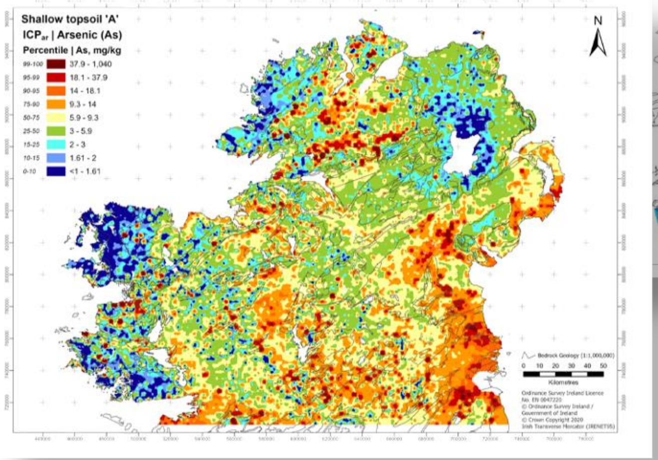
Tellus

Department for the Economy
www.economy-ni.gov.uk
GSNI Geological Survey of Northern Ireland
Celebrating 70 years: 1947–2017

Geological Survey
Suirbhéireacht Gheolaíochta
Ireland | Éireann

Roinn Cumarsáide, Gníomhaithe ar son na hAeráide & Comhshaoil
Department of Communications, Climate Action & Environment



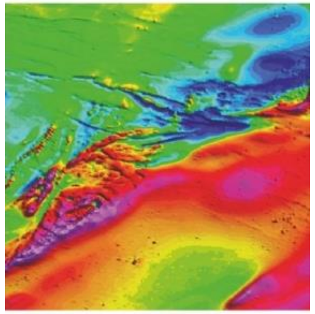


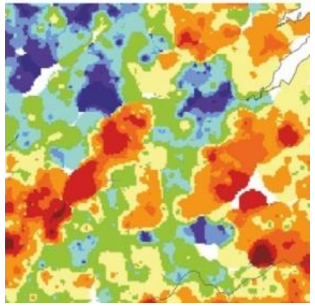
Geological Survey, Ireland (GSI) and Geological Survey of Northern Ireland (GSNI)

> *"Tellus is a ground and airborne geoscience mapping programme, collecting chemical and geophysical data that will inform the management of Ireland's environment and natural resources"* (**data available at www.tellus.ie**).

Geochemical surveys of soil: topsoil, steam water and stream sediment

#GSNI70



Geophysical and geochemical mapping for the island of Ireland

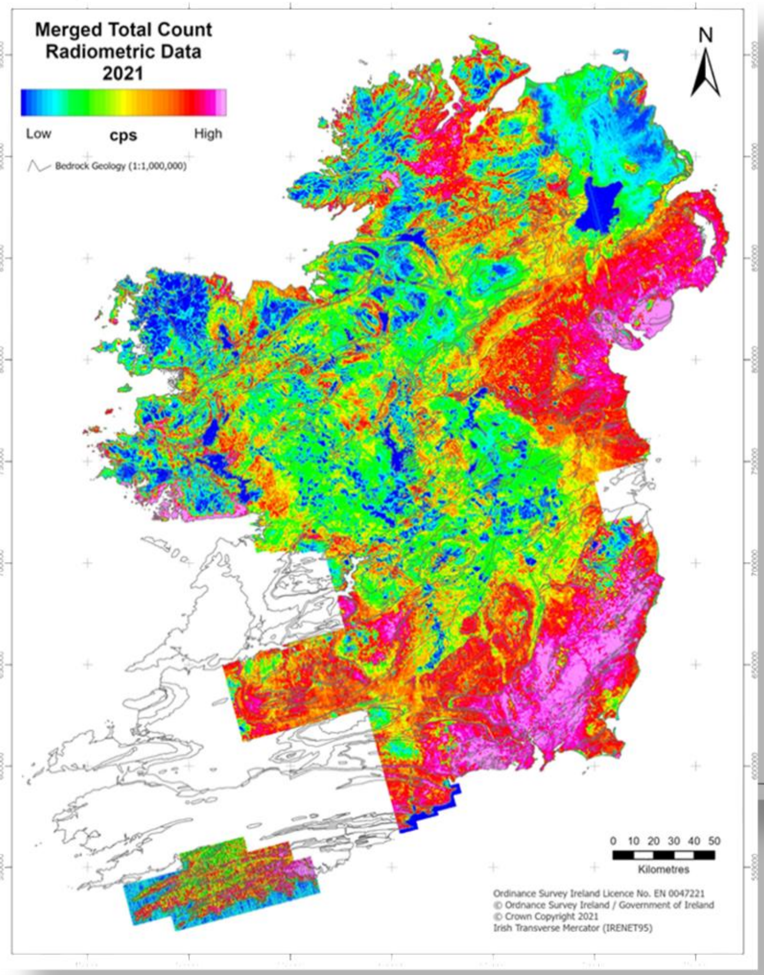
- Current Republic of Ireland phases managed by Geological Survey Ireland
- Funded by the Irish Government through the Department of Communications, Climate Action and Environment (€3-4M pa)
- Key part of government Statement of Strategy under Natural Resources 2016-2019:

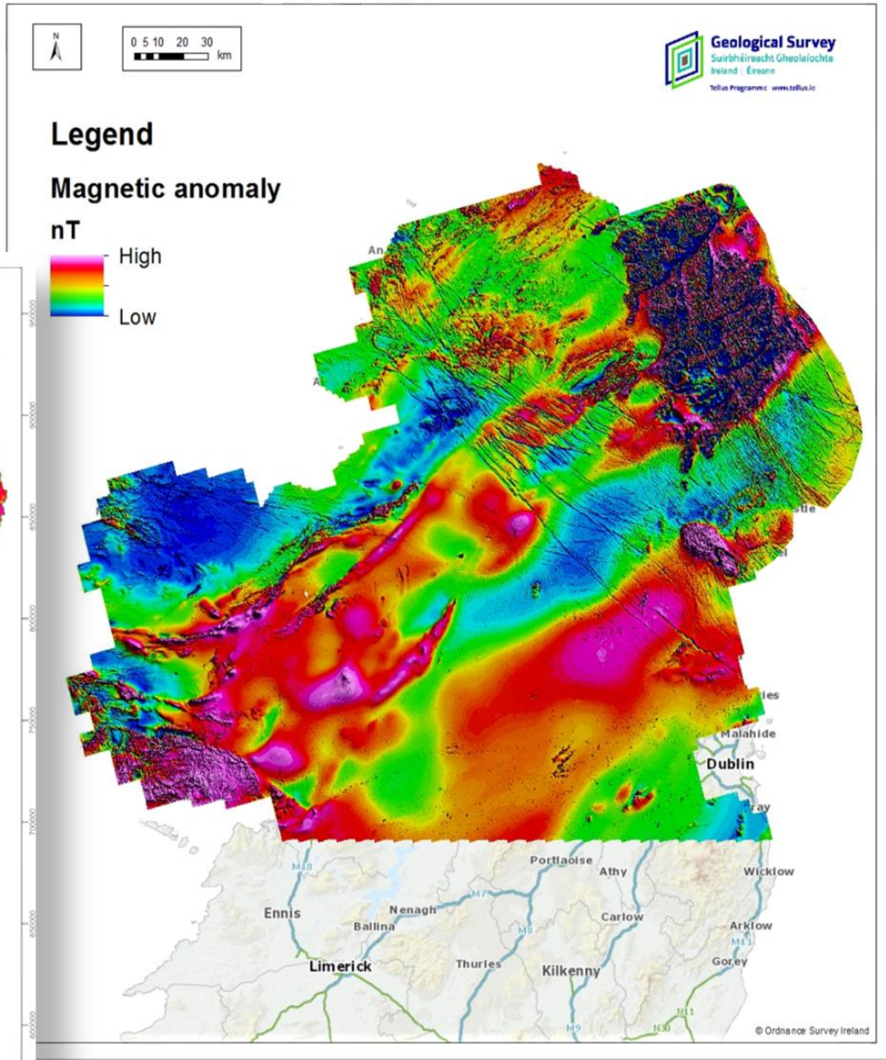
> To exploit and manage our inland fisheries, mineral, hydrocarbon and other geological resources in a sustainable and productive manner

Sample type collected	Analytical method	Chemical determinands
Stream sediment (fine fraction, <150 µm grain size)	X-ray fluorescence spectrometry (XRFS)	K, Ca, Ti, Mn, Fe, S, Cl, Sc, V, Cr, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Br, Rb, Sr, Zr, Nb, Mo, Nd, Sm, Yb, Hf, Ta, W, Tl, Pb, Bi, Th, U, Ag, Cd, In, Sn, Sb, Te, I, Cs, Ba, La, Ce, Na, Mg, Al, Si, P, Ba, Y
	Lead collection fire assay for precious group elements	Au, Pd, Pt
Stream water (filtered to <0.45 µm)	ICP-MS	Li, Be, B, Na, Mg, Al, Si, P, S, K, Ca, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, As, Se, Rb, Sr, Y, Zr, Nb, Ag, Cd, Sn, Sb, Cs, Ba, La, Ce, Pr, Nd, Sm, Eu, Tb, Tm, Yb, Lu, Hf, Ta, W, Tl, Pb, Bi, Th, U, Ti, Mo, Gd, Dy, Ho, Er
	Ion chromatography	Cl^-, SO_4^{2-}, NO_3^-, Br^-, NO_2^-, HPO_4^{2-}, F^-
	Organic carbon (NPOC) analyser	C
	Stream water pH, specific electrical conductivity, total alkalinity (bicarbonate)	
Topsoil (c.5–20 cm deep)	ICP(-OES/-MS) following aqua regia digestion	Al, B, Ba, Ca, Cr, Cu, Fe, K, Li, Mg, Mn, Na, Ni, P, S, Sr, Ti, V, Zn, Zr, Ag, As, Be, Bi, Cd, Ce, Co, Cs, Ga, Ge, Hf, Hg, In, La, Lu, Mo, Nb, Pb, Rb, Sb, Sc, Se, Sn, Ta, Tb, Te, Th, Tl, U, W, Y, Yb
	X-ray fluorescence spectrometry (XRFS)	K, Ca, Ti, Mn, Fe, S, Cl, Sc, V, Cr, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Br, Rb, Sr, Zr, Nb, Mo, Nd, Sm, Yb, Hf, Ta, W, Tl, Pb, Bi, Th, U, Ag, Cd, In, Sn, Sb, Te, I, Cs, Ba, La, Ce, Na, Mg, Al, Si, P, Ba, Y
	Soil pH (by $CaCl_2$) and soil loss-on-ignition at 450°C	

Tellus data available at www.tellus.ie

Merging Geophyics





Merged Geophysics Data

- Over 300,000 line-km of data (inc NI).
- Over 50,000 km^2
- Over 50 Million data points

Data issues

- Dealing with ‘big’ geodata
- Integrating data collected over different sampling supports
- Understanding and acknowledging nature of the data
- How do you define and indicate a baseline?
- Producing robust and meaningful results
- Providing insight – working with partners

What are the opportunities to develop the use of machine learning?

Handling ‘big’ geodata

Low order streams, small catchments scale

Fine fraction **stream sediment** 150 µm

Tellus data available at www.tellus.ie

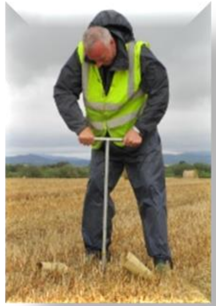
Quality controlled field information, observations, geo-locations, contamination *etc.*

Suite of **stream water** samples:

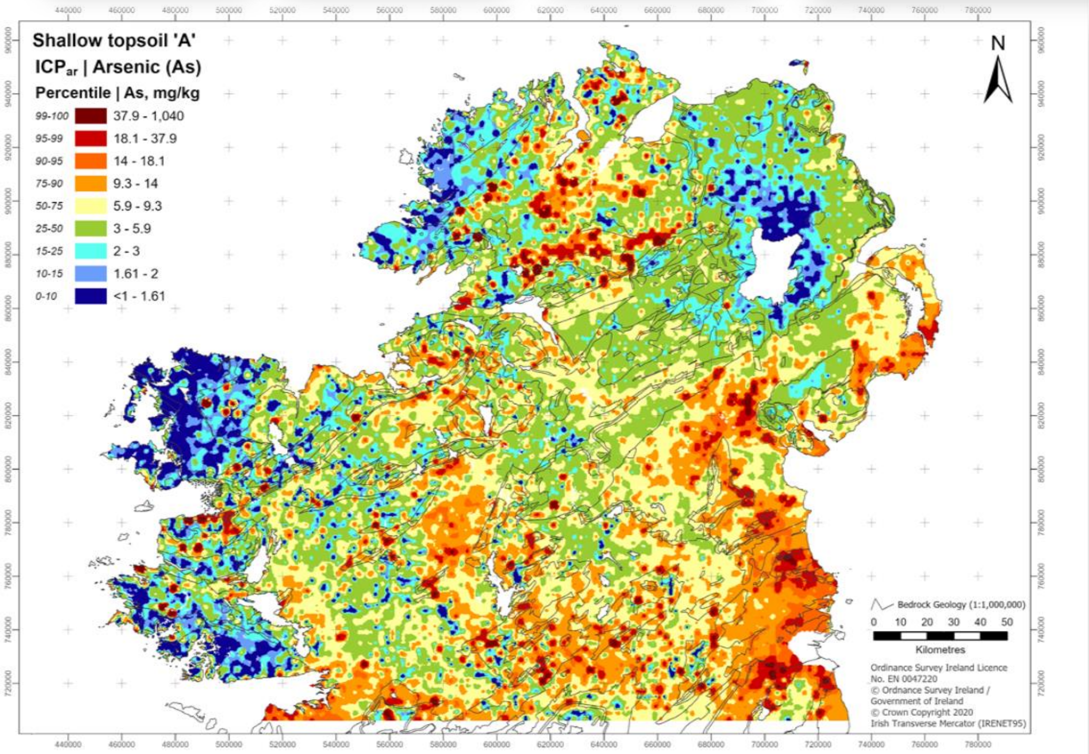
1. 2x filtered waters to 0.45 µm → lab
2. 2x unfiltered samples: pH, HCO_3^- SEC (~TDS)

Panning for **heavy mineral concentrate** <2mm fraction

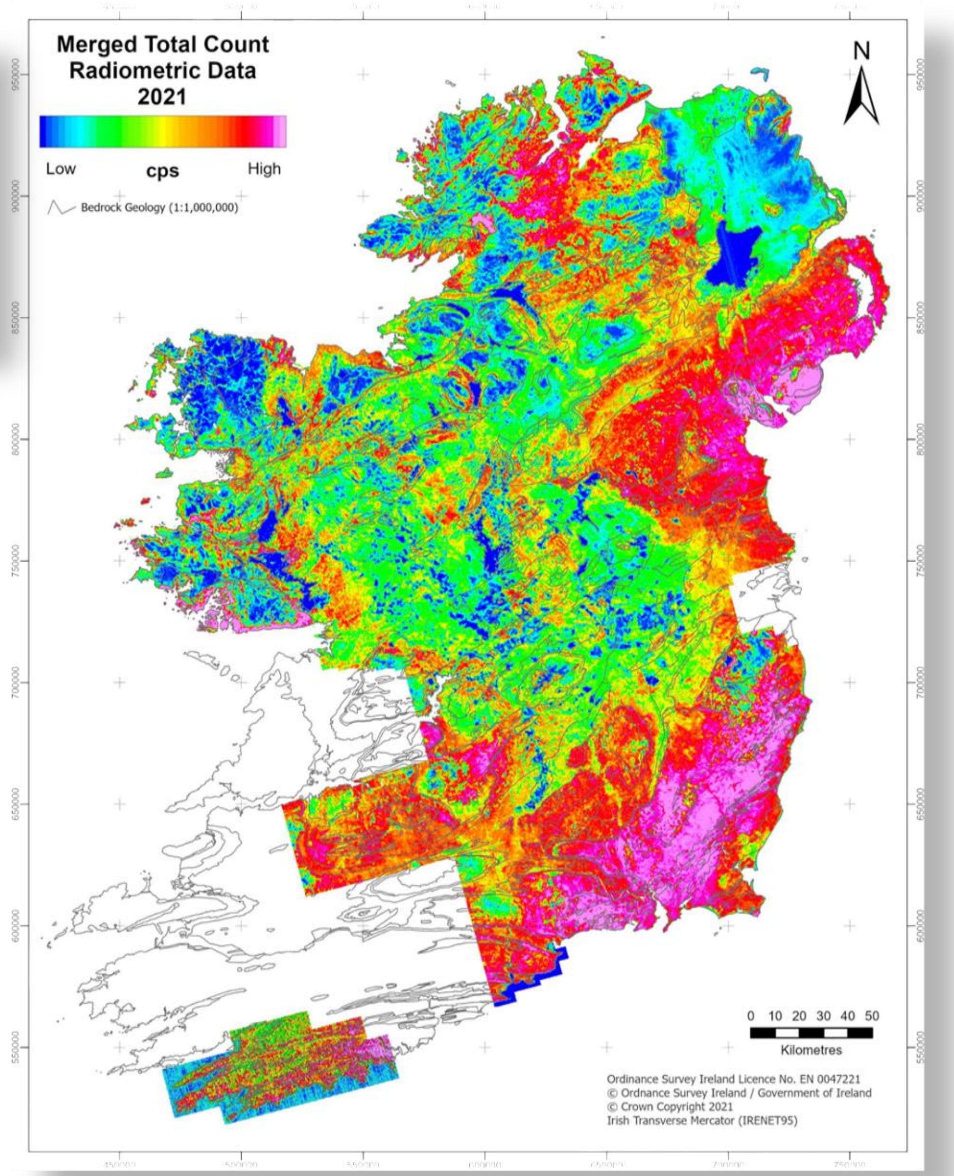
Shallow and deeper **soil** samples, average 1 per 4 km^2



Electromagnetic coil
- Ground conductivity
Gamma-ray detector
- Natural gamma radiation
Magnetometer
- Magnetic field



Shallow topsoil 'A'
ICPar | Arsenic (As)
Percentile | As, mg/kg
99-100 37.9 - 1,040
95-99 18.1 - 37.9
90-95 14 - 18.1
75-90 9.3 - 14
50-75 5.9 - 9.3
25-50 3 - 5.9
15-25 2 - 3
10-15 1.61 - 2
0-10 <1 - 1.61
N
Bedrock Geology (1:1,000,000)
0 10 20 30 40 50
Kilometres
Ordinance Survey Ireland Licence No. EN 0047220
© Ordinance Survey Ireland / Government of Ireland
© Crown Copyright 2020
Irish Transverse Mercator (IRENET95)

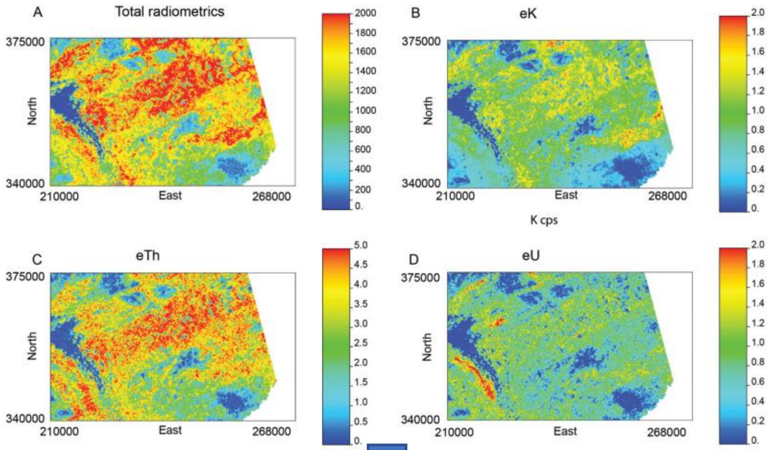


Merged Total Count
Radiometric Data
2021
Low
cps
High
Bedrock Geology (1:1,000,000)
N
0 10 20 30 40 50
Kilometres
Ordinance Survey Ireland Licence No. EN 0047221
© Ordnance Survey Ireland / Government of Ireland
© Crown Copyright 2021
Irish Transverse Mercator (IRENET95)

Integrating data collected over different sampling supports

Improving estimation for mineral exploration

Secondary data: Natural radioactivity

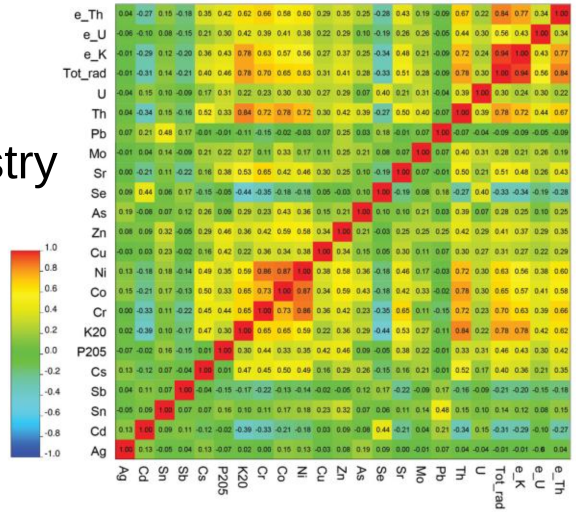


Use of geostatistical Bayesian updating to integrate airborne radiometrics and soil geochemistry to improve mapping for mineral exploration

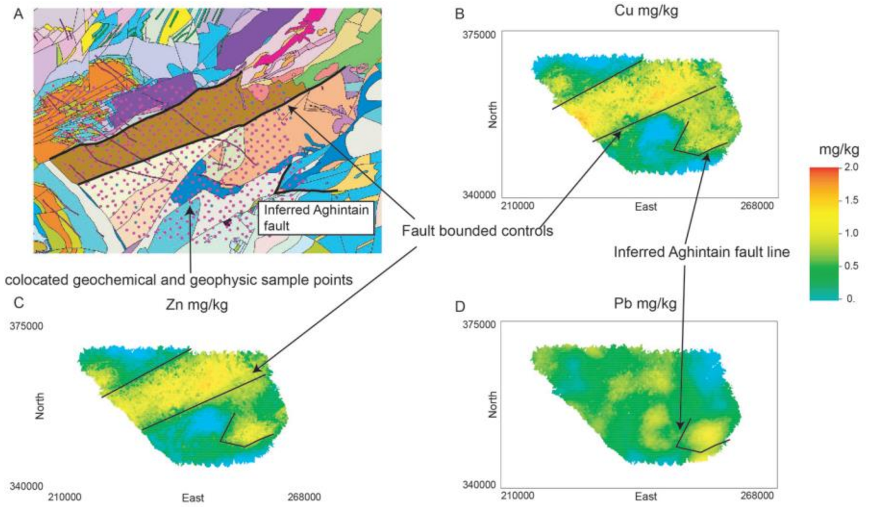
by J.M. McKinley*, C.V. Deutsch†, C. Neufeld†, M. Patton‡, M. Cooper‡, and M.E. Young‡

Danie Krige Commemorative Edition of The Saimm Journal, 2015 vol 114, pp. 575-586.

Prior data soil geochemistry

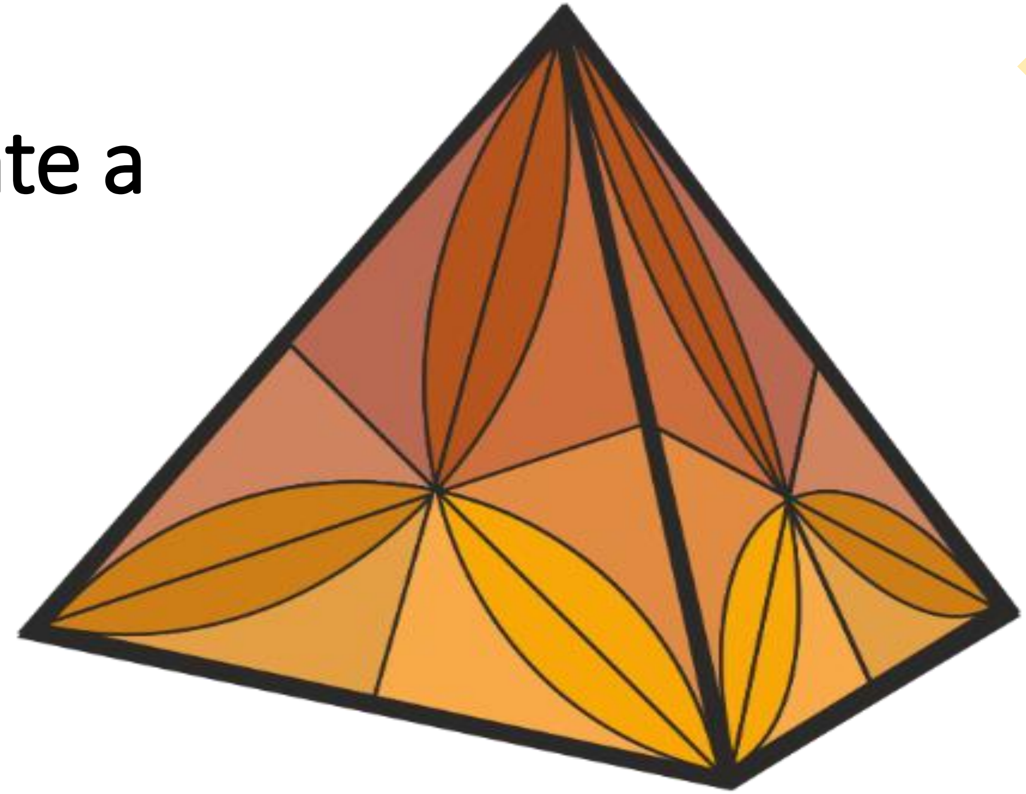


Updated post probabilities



Understanding and acknowledging nature of the data

How do you define and indicate a baseline?



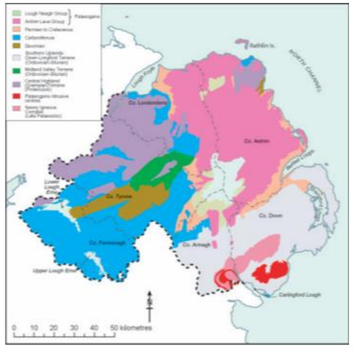
Geochemical Baseline

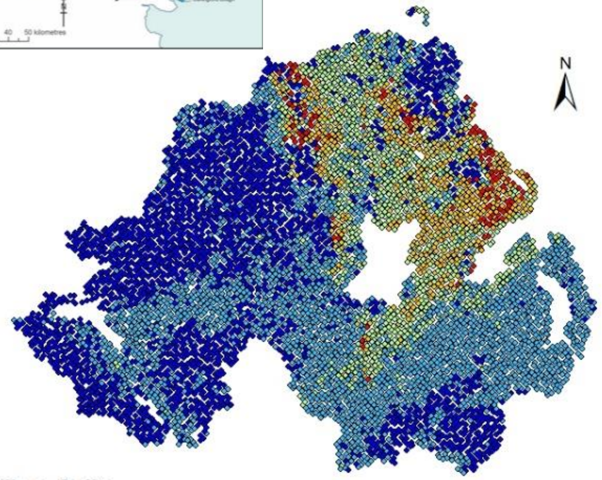
‘The natural variation in the concentration of an element in the media of the superficial environment’

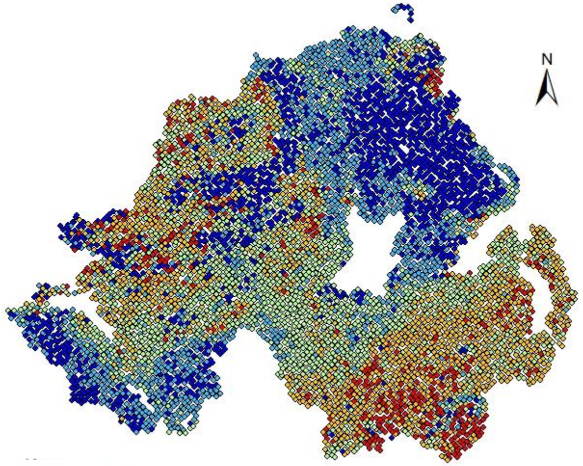
- As cited in Buccianti, Nisi and Raco (2016)
- Officially introduced in 1993 in the context of the International Geological Correlation Program (IGCP project 360)

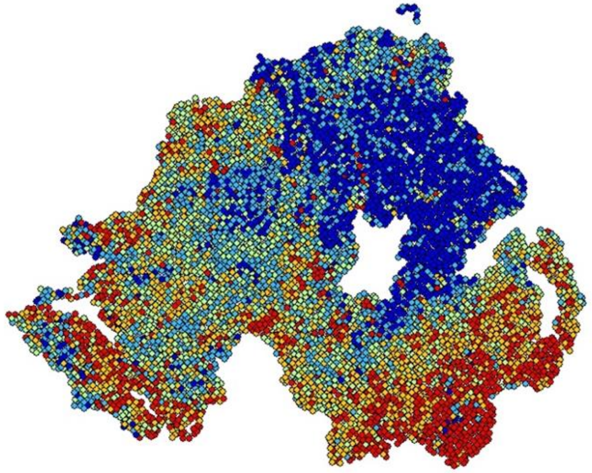
Point maps – the objective ground truth?

It is often thought that raw one-component maps report “what is there”, that they report a sort of “objective ground truth”.









How do we model and test the relationship between different types of data ?

Understanding the different types of data

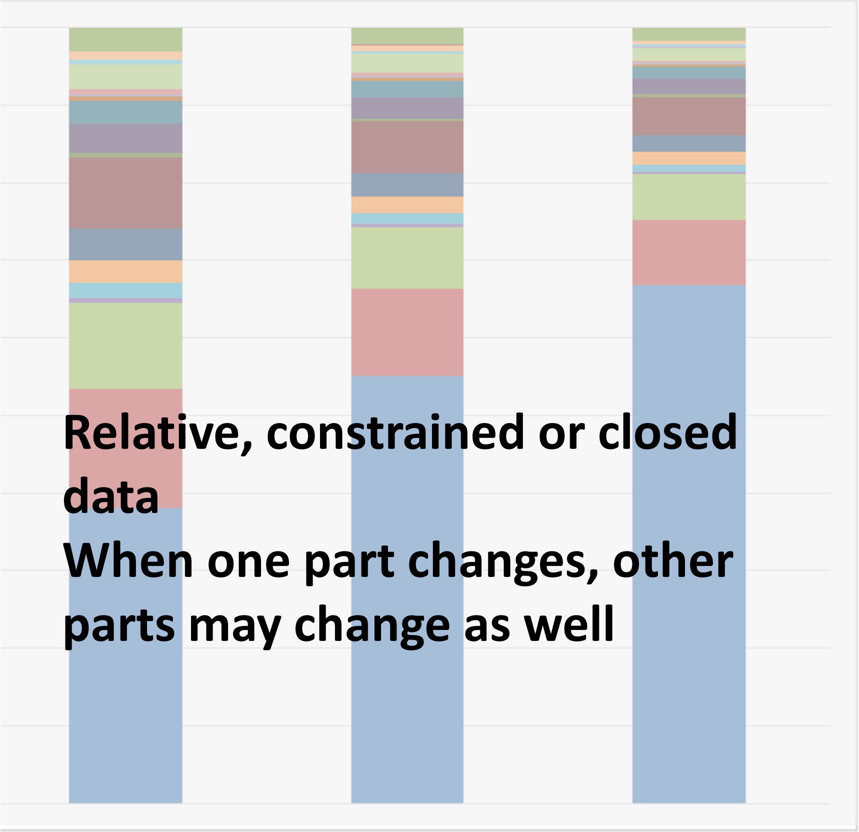
Is 3% or 3ppm large or small?

Compared with 80% or 80ppm?
Compared with 0.3% or 0.3ppm?

Data reported in different physical units (ppm, mg/kg or as percentages) are relative and only have an importance as part of a whole

So proceed with statistical analysis with caution!

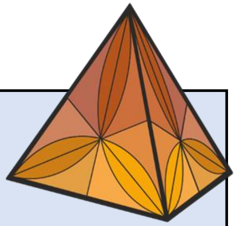
- ➢ The components may be reported in different physical units (ppm, mg/kg or as percentages) and all the components may not be reported or measured
- ➢ The components do not need to add up to 100%



No units on purpose to make a point

QUEEN'S UNIVERSITY BELFAST

Compositional Data Analysis CoDA



The **pairwise log-ratio transformation (pwlr)** takes all possible pairs of elements and computes their log-ratios

$$\mathrm{pwlr}(\mathbf{x}) = \begin{bmatrix} 0 & \ln\frac{x_1}{x_2} & \cdots & \ln\frac{x_1}{x_D} \\ \ln\frac{x_2}{x_1} & 0 & & \ln\frac{x_2}{x_D} \\ \vdots & & \ddots & \vdots \\ \ln\frac{x_D}{x_1} & \ln\frac{x_D}{x_2} & \cdots & 0 \end{bmatrix} = \left[\ln\frac{x_i}{x_j}\right] = [\xi_{ij}].$$

The **additive log-ratio transformation (alr)** takes just one of the rows or columns of the pwlr (and removes the constant zero), typically the last column

$$\mathrm{alr}(\mathbf{x}) = \left[\ln\frac{x_1}{x_D} \quad \ln\frac{x_2}{x_D} \quad \cdots \quad \ln\frac{x_{D-1}}{x_D}\right] = [\xi_{iD}],$$

The **centred log-ratio (clr)** transformation represents each element component as a ratio to a central value:

$$\mathrm{clr}(\mathbf{x}) = \left[\ln\frac{x_1}{g(\mathbf{x})} \quad \ln\frac{x_2}{g(\mathbf{x})} \quad \cdots \quad \ln\frac{x_D}{g(\mathbf{x})}\right],$$

corresponding to the geometric mean of *all considered* components,

$$g(\mathbf{x}) = \sqrt[D]{\prod_{i=1}^{D} x_i} = \exp\left(\frac{1}{D}\sum_{i=1}^{D} \ln x_i\right).$$

The (family of) isometric log-ratio (ilr) transformations are formed by D-1 log-contrasts which are computed using vectors of coefficients ω_i=[ω_i1,ω_i2,…,ω_iD] which are orthogonal to each other. Balances can be used to obtain the orthogonal log-contrasts. The four log-ratios necessary to describe the behaviour of the subcomposition (Fe2O3, V, Cr, Co, Ni) are:

$$\xi_1 \propto \ln\frac{\mathrm{Fe2O3}}{\mathrm{V}}, \qquad \xi_2 \propto \ln\frac{\mathrm{Co}}{\mathrm{Ni}}, \qquad \xi_3 \propto \ln\frac{\mathrm{Cr}}{\sqrt[2]{\mathrm{Co}\cdot\mathrm{Ni}}},$$

$$\xi_4 \propto \ln\frac{\sqrt[2]{\mathrm{Fe2O3}\cdot\mathrm{V}}}{\sqrt[3]{\mathrm{Co}\cdot\mathrm{Ni}\cdot\mathrm{Cr}}}$$

CoDa Association

https://www.coda-association.org/

Advancing mathematical geoscience - addressing the nature of data

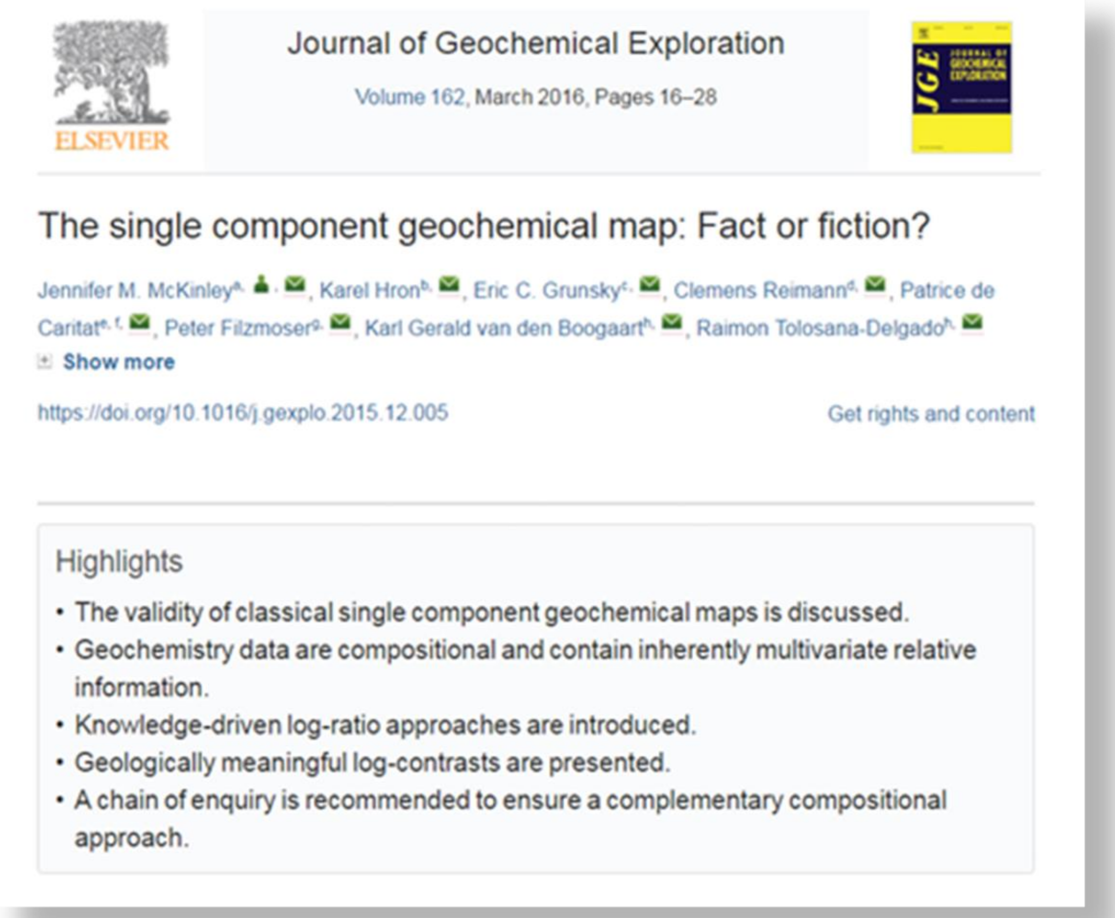
How can the compositional data of data be maintained within machine learning?

Geochemical data are proportional.

Each component has an amount which represents its importance as part of the whole composition.

Ratios between components are unaffected by constant sum closure effects caused by the relative nature of geochemical data.

Compositional log-ratio transformations allow us to explore associations between elements affected by geological/geochemical processes.


ELSEVIER

Journal of Geochemical Exploration

Volume 162, March 2016, Pages 16–28

The single component geochemical map: Fact or fiction?

Jennifer M. McKinley[a], Karel Hron[b], Eric C. Grunsky[c], Clemens Reimann[d], Patrice de Caritat[e, f], Peter Filzmoser[g], Karl Gerald van den Boogaart[h], Raimon Tolosana-Delgado[h]

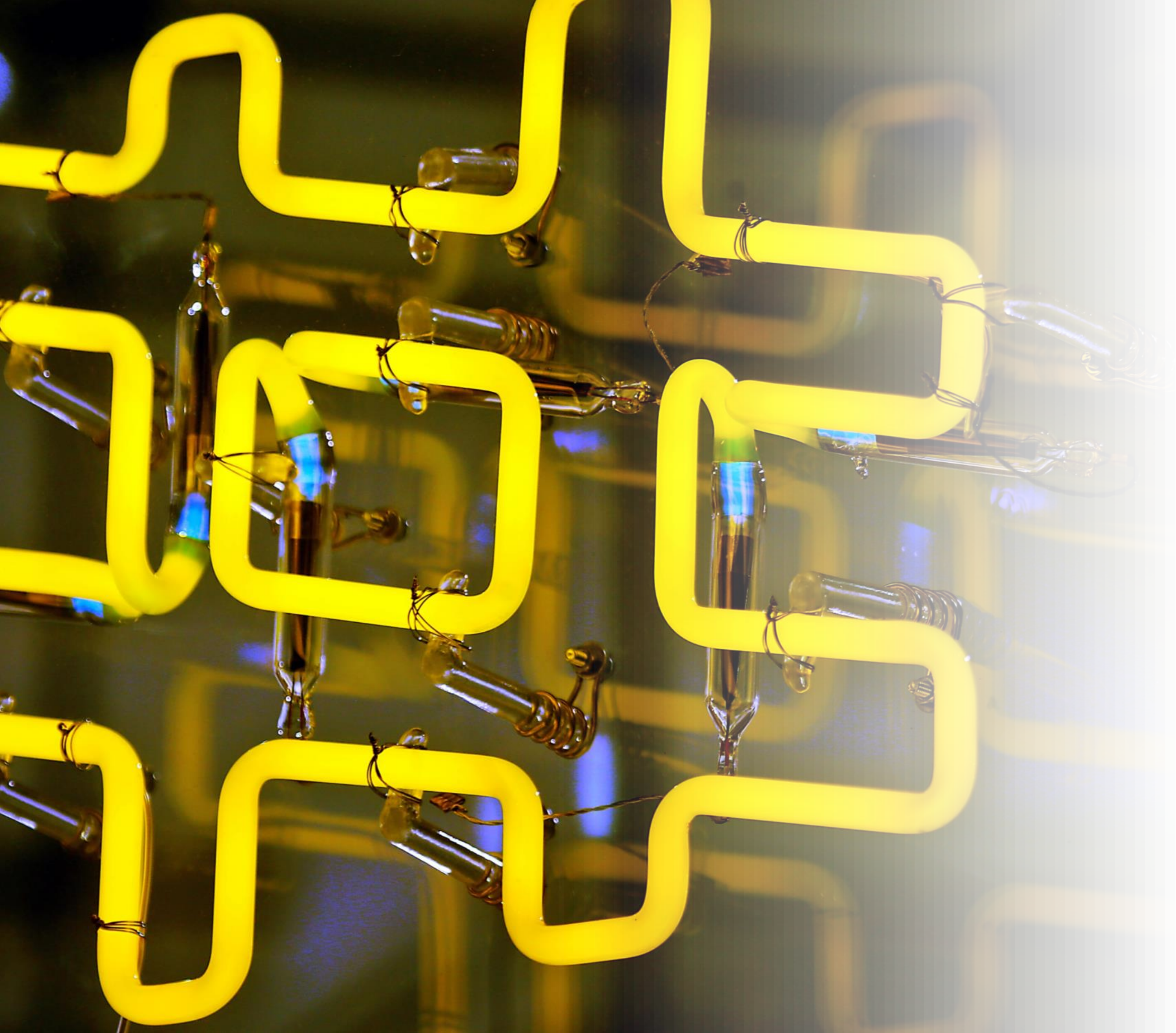
Show more

https://doi.org/10.1016/j.gexplo.2015.12.005

Get rights and content

Highlights

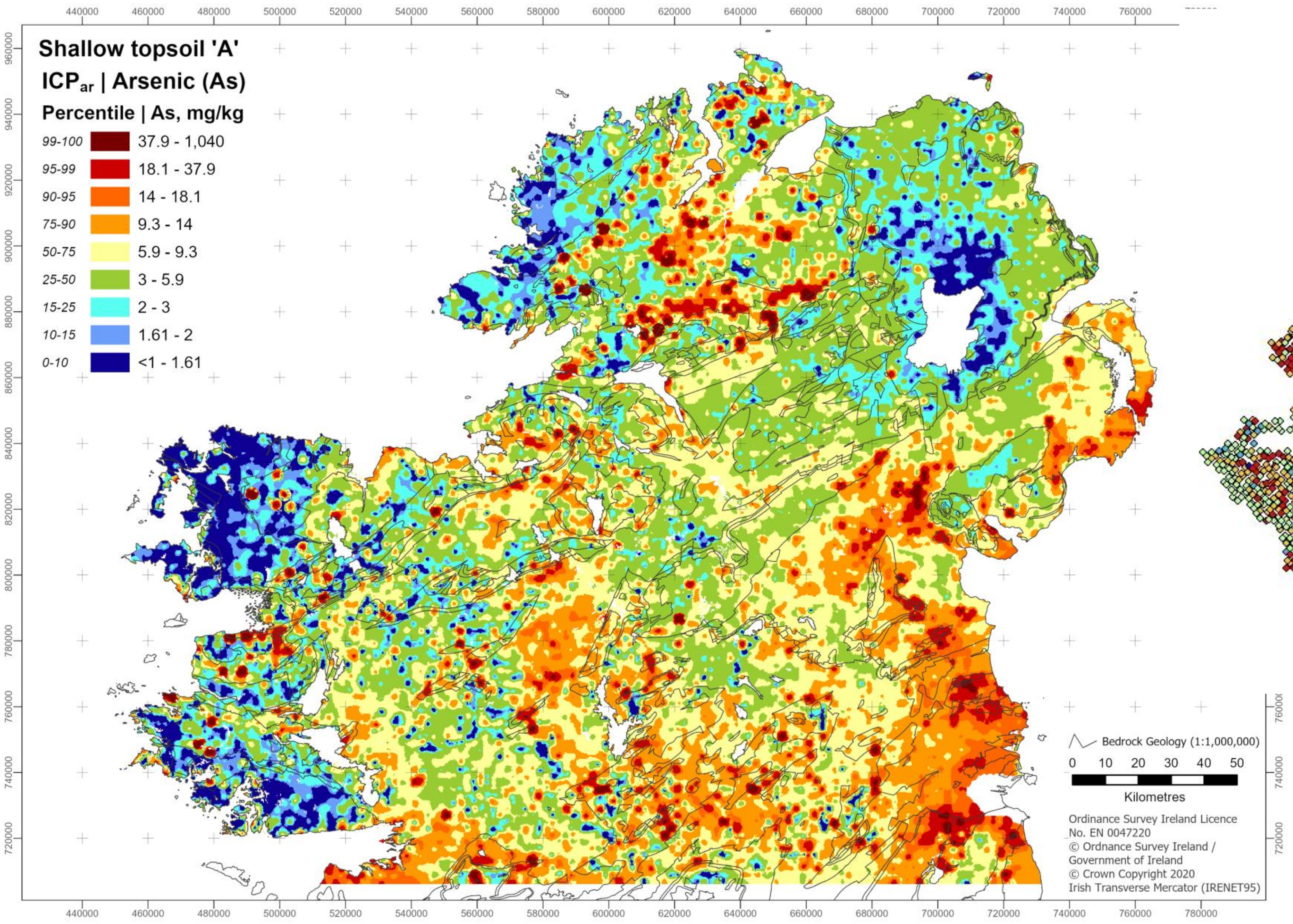
- The validity of classical single component geochemical maps is discussed.
- Geochemistry data are compositional and contain inherently multivariate relative information.
- Knowledge-driven log-ratio approaches are introduced.
- Geologically meaningful log-contrasts are presented.
- A chain of enquiry is recommended to ensure a complementary compositional approach.



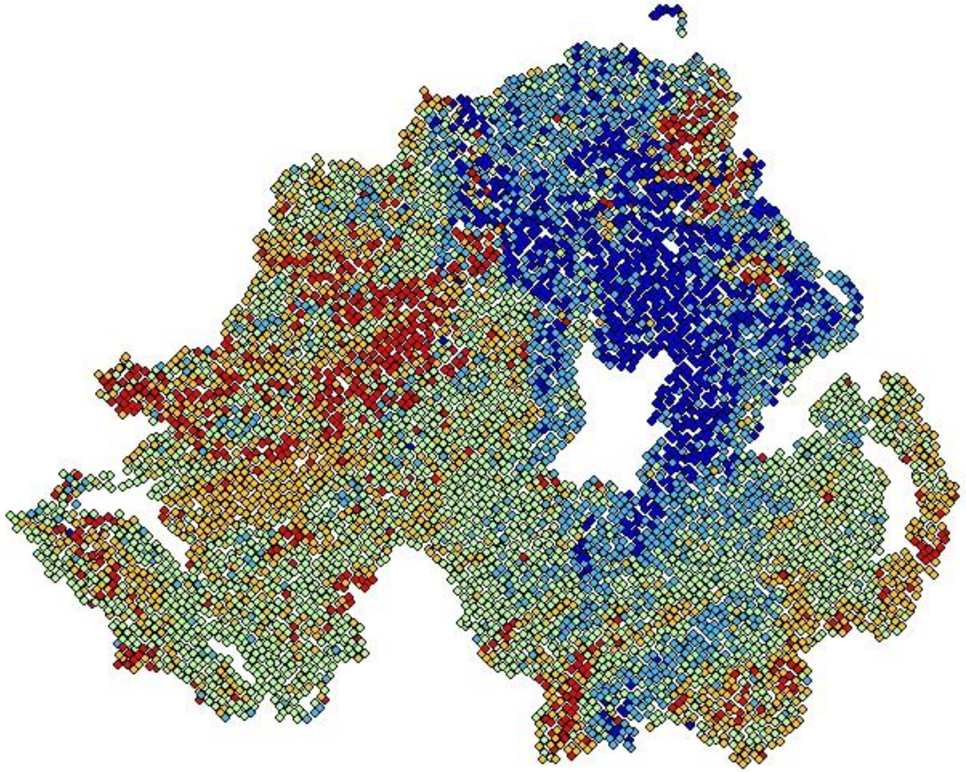
Producing robust and meaningful results

Providing insight – working with partners

Tellus reporting of As in top soils



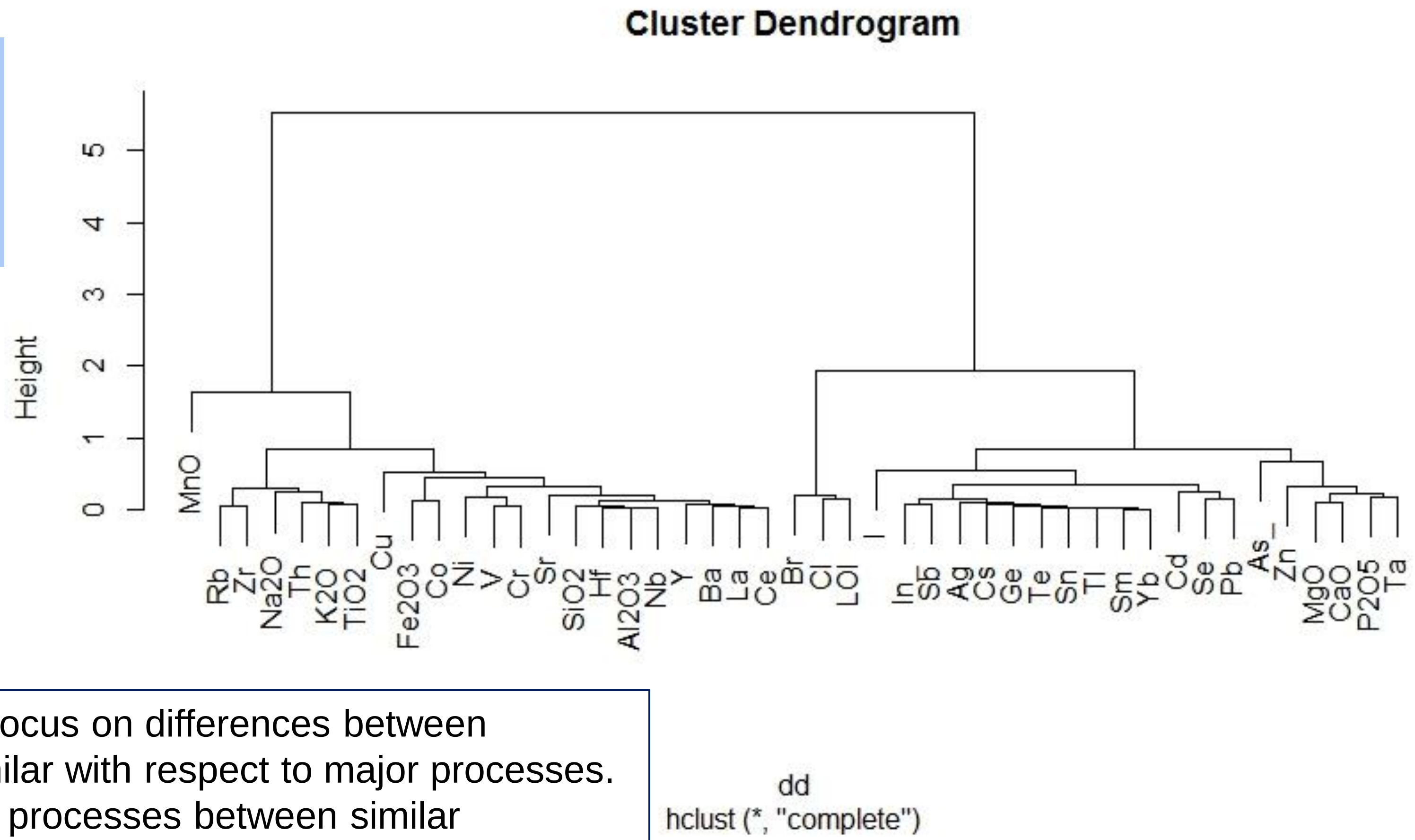
Compositional centred log ratio of As



Takes into account the proportional relationship between elements

Tellus Soil geochemistry

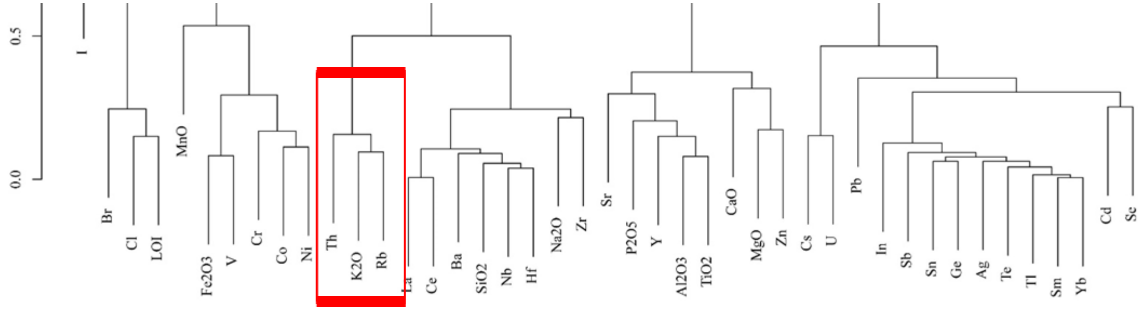
The (family of) isometric log-ratio (ilr) transformations

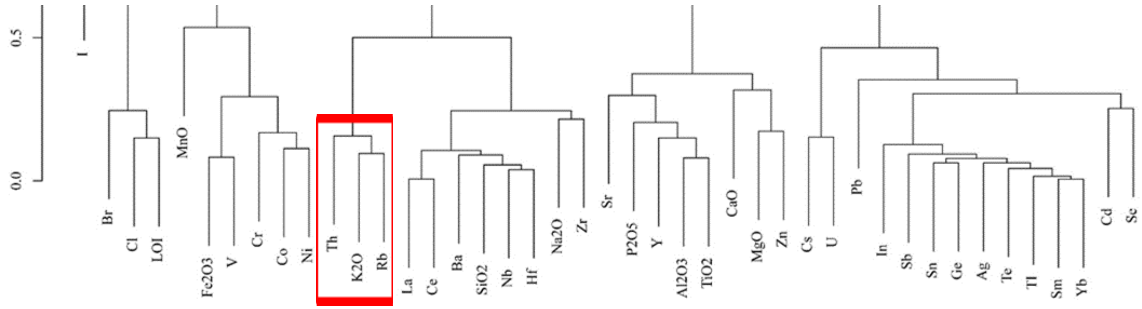


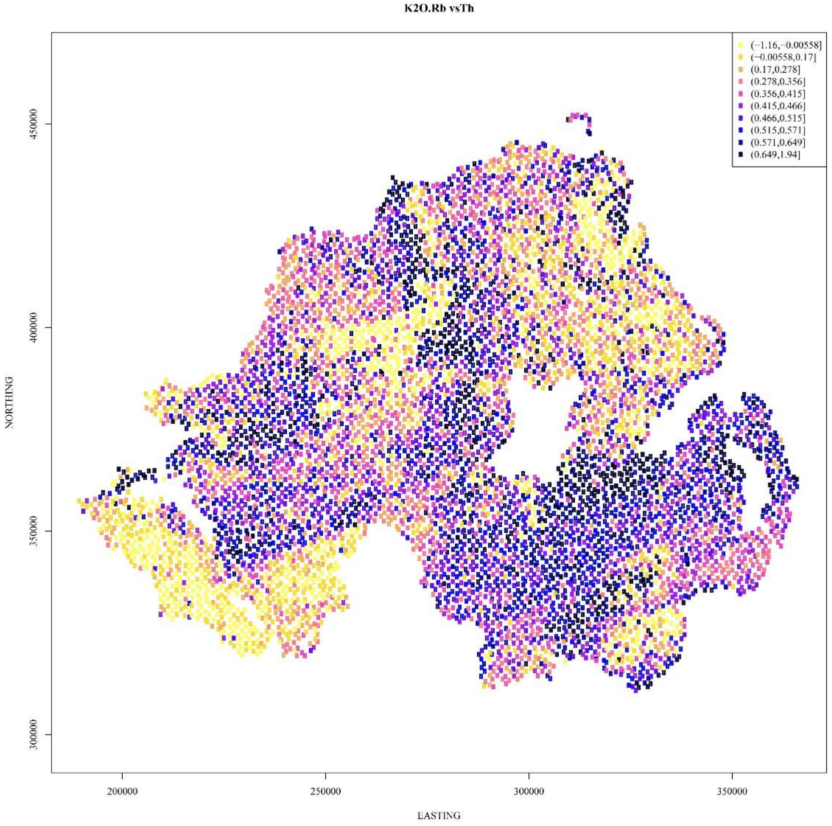
- Balances of elements focus on differences between elements behaving similar with respect to major processes.
- The balances focus on processes between similar elements.
- Identifies second-order processes that would otherwise be overshadowed by the major processes.

Subcompositions

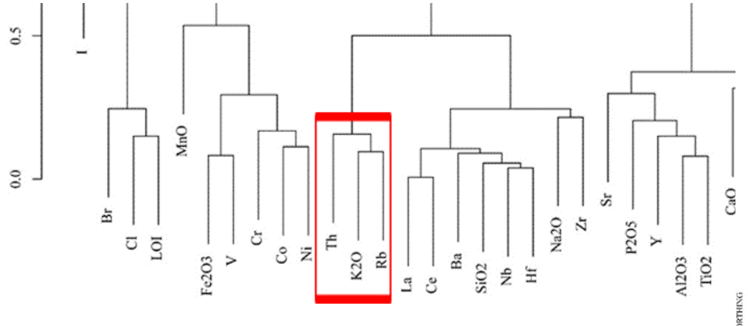
- For instance the two log-ratios necessary to describe the behaviour of the subcomposition (Th, K_2O, Rb):
- $\xi_1 \propto \ln \frac{K_2O}{Rb}$, $\quad \xi_2 \propto \ln \frac{Th}{\sqrt[2]{K_2O \cdot Rb}}$,

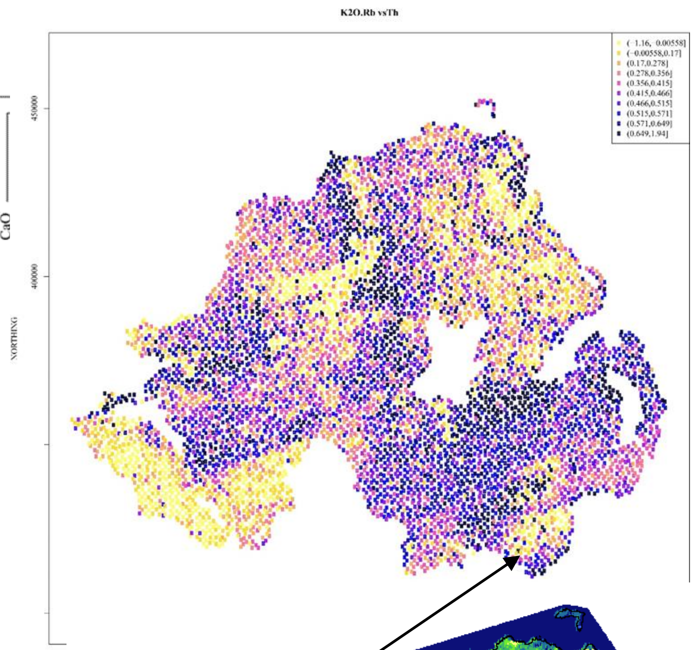




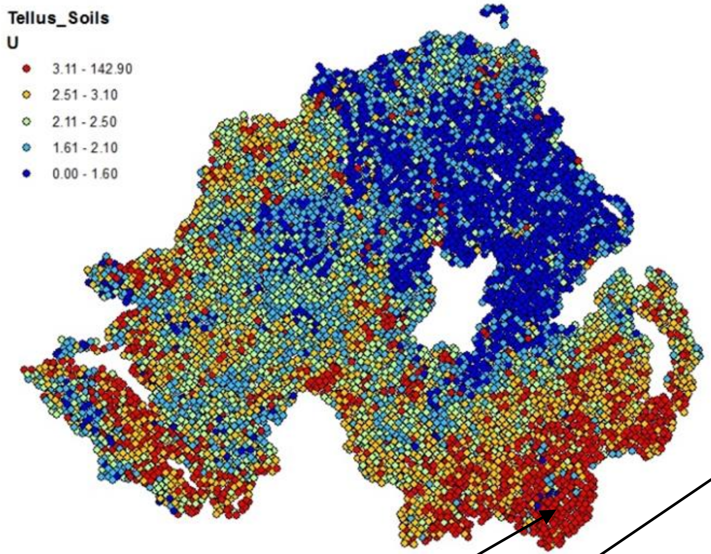


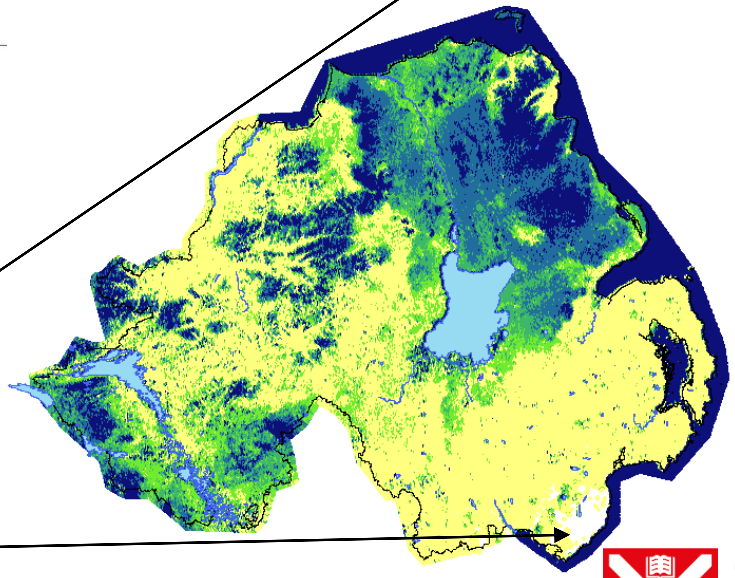
QUEEN'S UNIVERSITY BELFAST





Uranium in soils





Elevated areas – Mourne Mountains

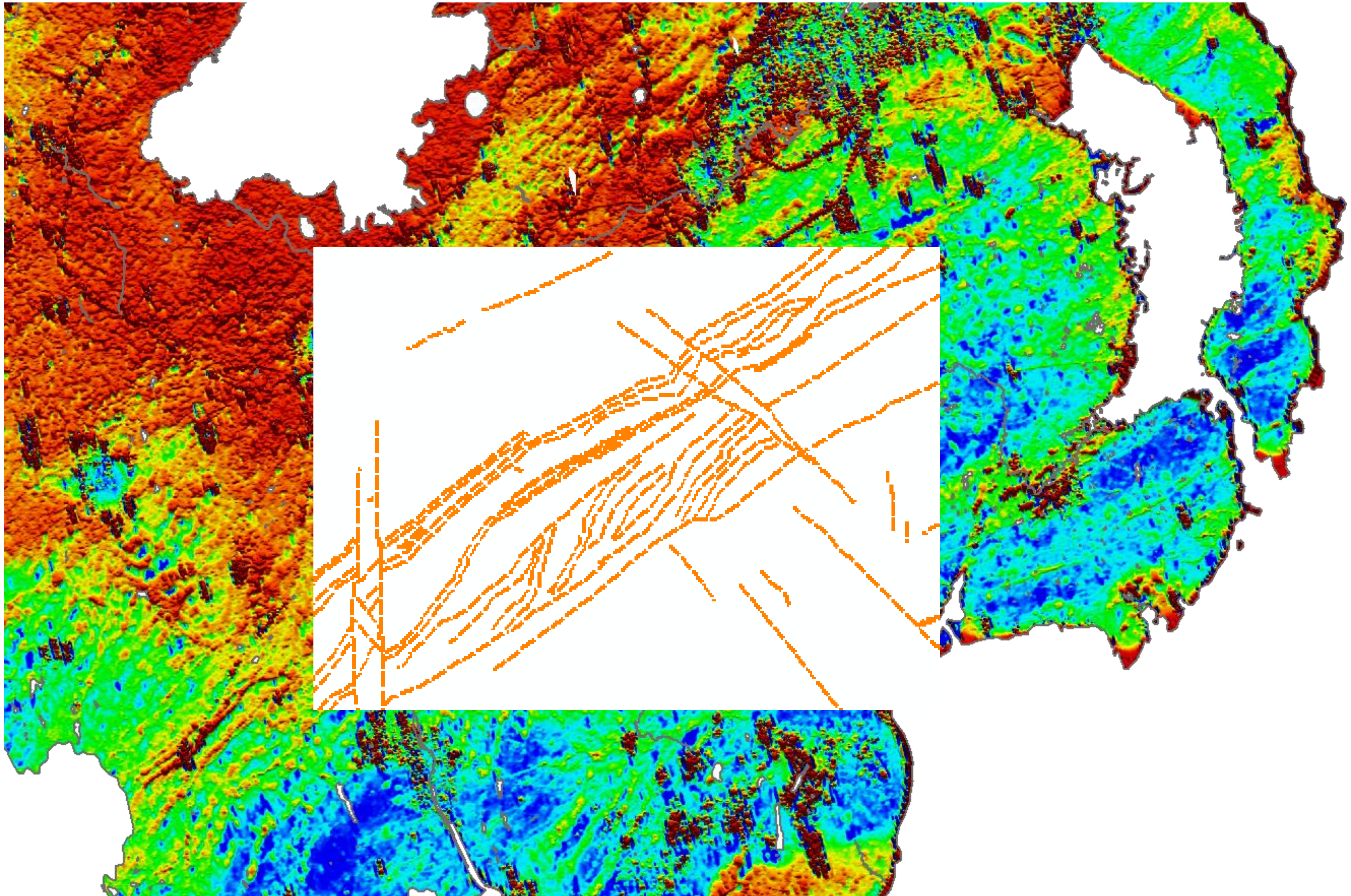
Gamma Radiation Dose rate

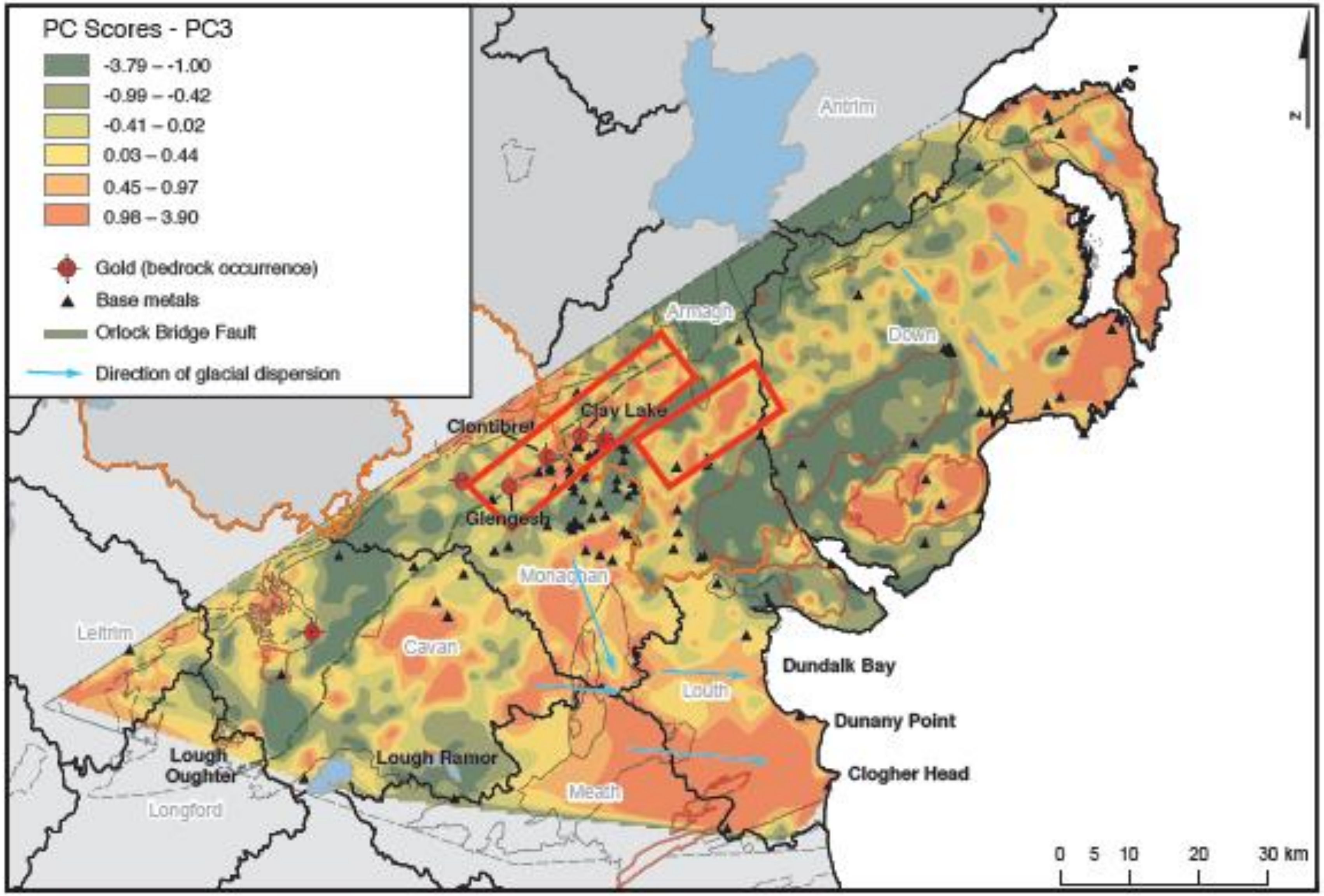
Case studies

- Natural resource estimation and Mineral prospectivity
- Climate action and Environmental monitoring
- Environment and Health

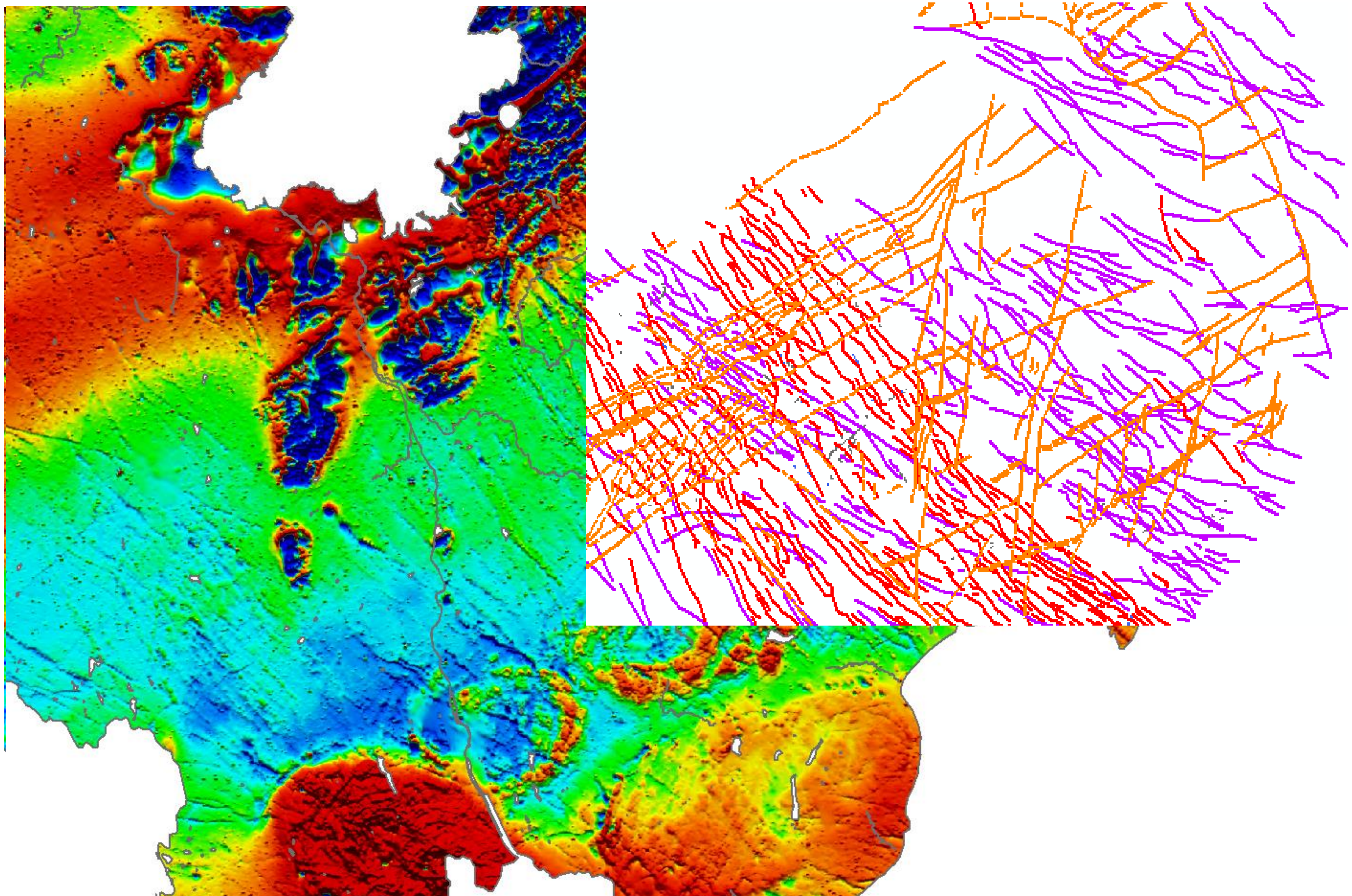
Natural resource estimation

- Use in exploration for gold, and arsenic in groundwater



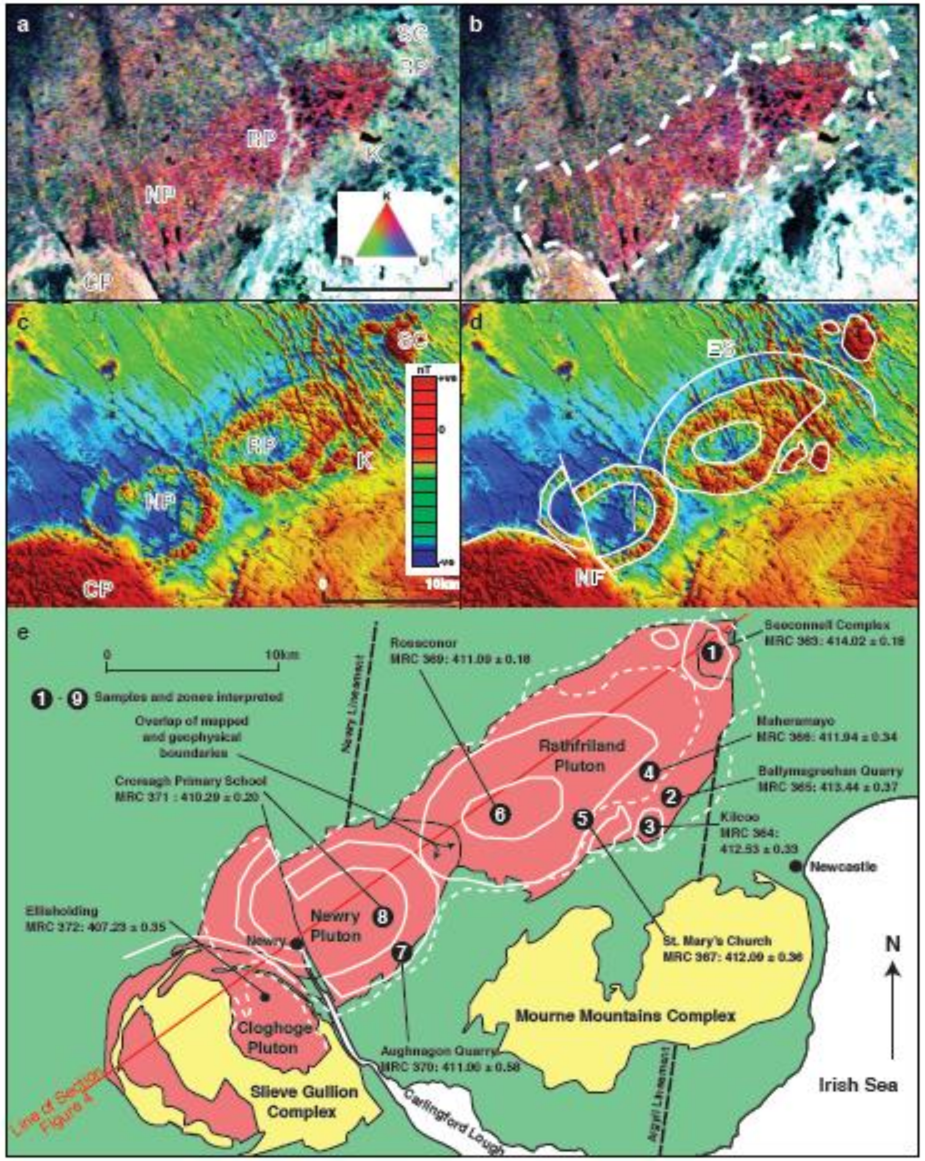

PC Scores - PC3
-3.79 – -1.00
-0.99 – -0.42
-0.41 – 0.02
0.03 – 0.44
0.45 – 0.97
0.98 – 3.90
Gold (bedrock occurrence)
Base metals
Orlock Bridge Fault
Direction of glacial dispersion
Antrim
Armagh
Down
Clay Lake
Clontibret
Glengesh
Monaghan
Leitrim
Cavan
Louth
Dundalk Bay
Dunany Point
Clogher Head
Lough Oughter
Lough Ramor
Meath
Longford
0 5 10 20 30 km
N

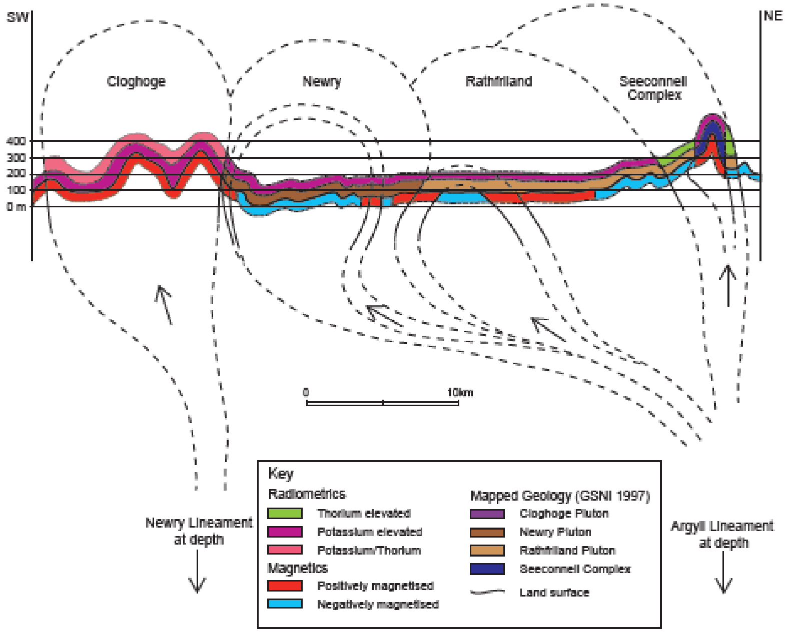
- Important to reservoirs, hydrocarbons, groundwater, geothermal



Department for the Economy
www.economy-ni.gov.uk

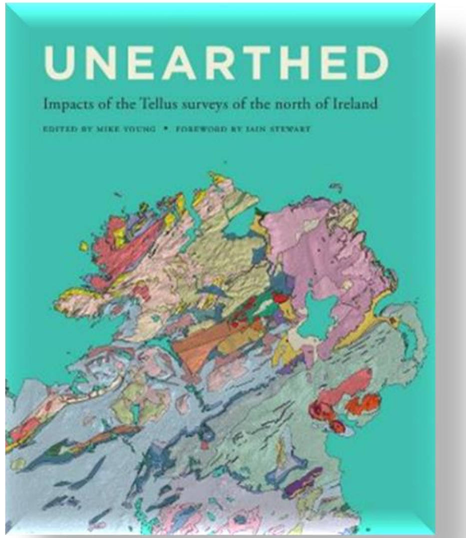
GSNI Geological Survey of Northern Ireland
Celebrating 70 years: 1947–20:





11. Shape and intrusion history of the Late Caledonian Newry Igneous Complex, Northern Ireland

Mark Cooper,[1] Paul Anderson,[2] Daniel Condon,[3] Carl Stevenson,[2] Rob Ellam,[4] Ian Meighan[1] and Quentin Crowley[5]



Current research by Z. Smillie, V. Demyanov, J McKinley, M. Cooper

Spatial feature selection to link multivariate radiometric data with geology

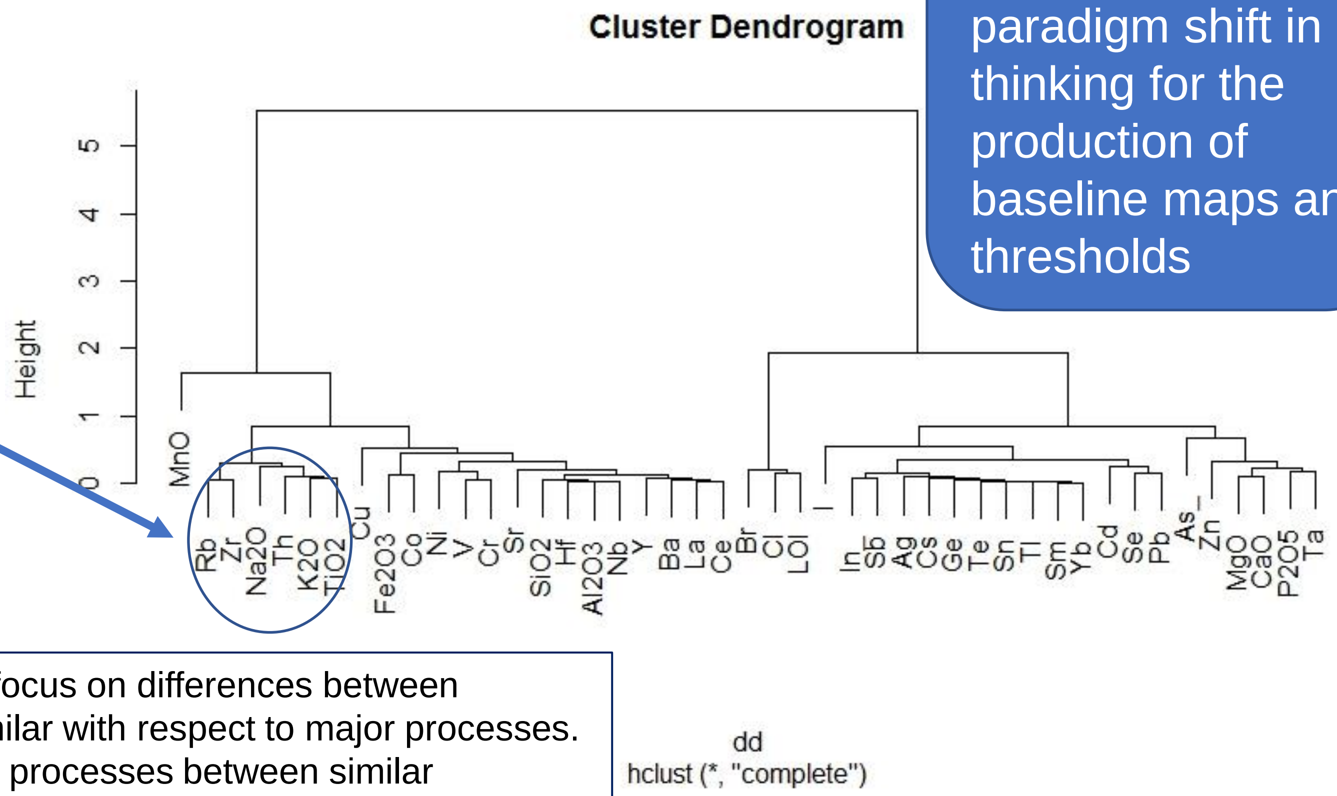
Spatial multivariate interpolation problem is commonly solved in geostatistics under assumptions imposed by cokriging estimator – stationarity and co-linearity. This makes cokriging application quite tedious with auto- and cross-covariance models to be derived for all the variables. This poses a difficulty for multivariate data that do not have good prior understanding of the correlation between the variables.

#GSNI70

Environmental monitoring

The nature of geochemical data

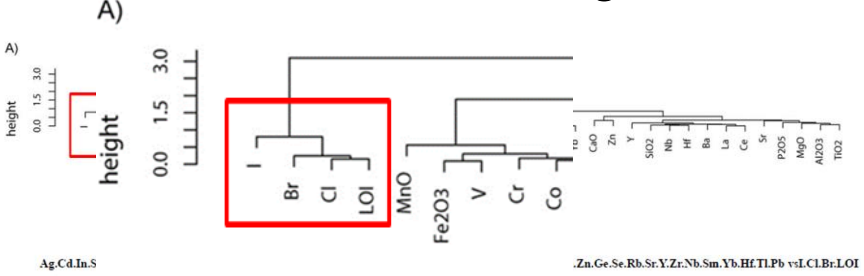
Elements behaving in a similar way with respect to major processes

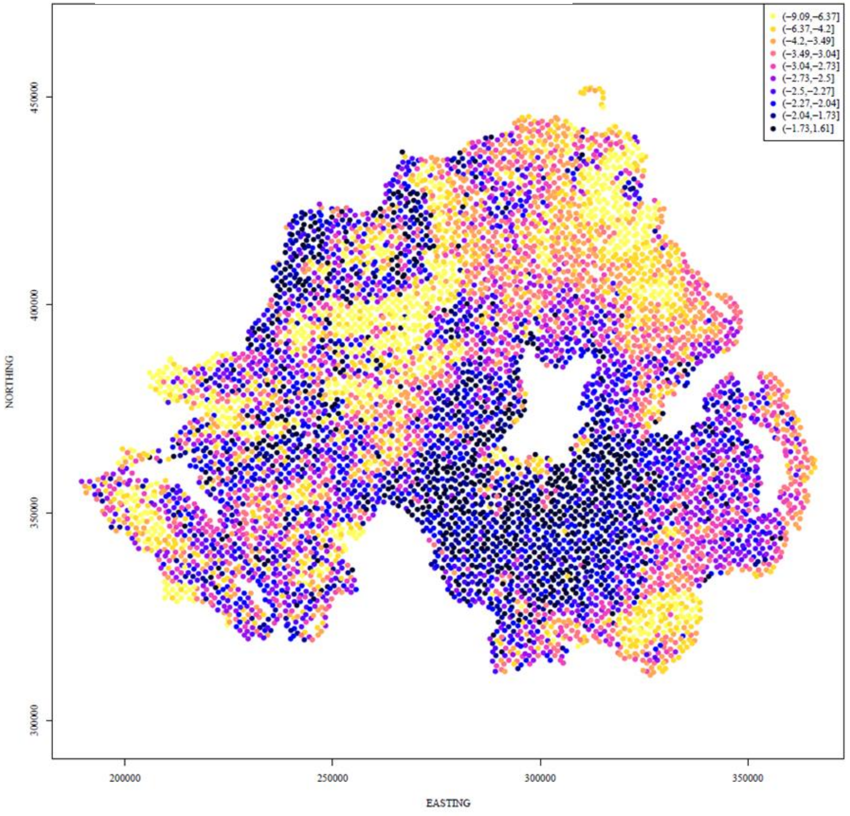


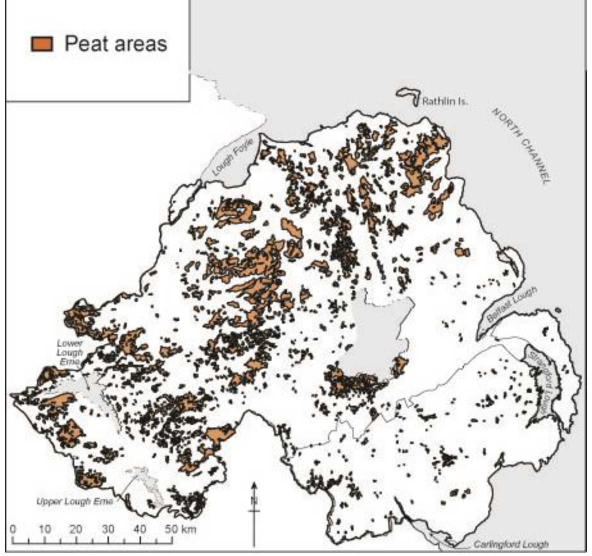
- Balances of elements focus on differences between elements behaving similar with respect to major processes.
- The balances focus on processes between similar elements.
- Identifies second-order processes that would otherwise be overshadowed by the major processes.

Key message: Requires a paradigm shift in thinking for the production of baseline maps and thresholds

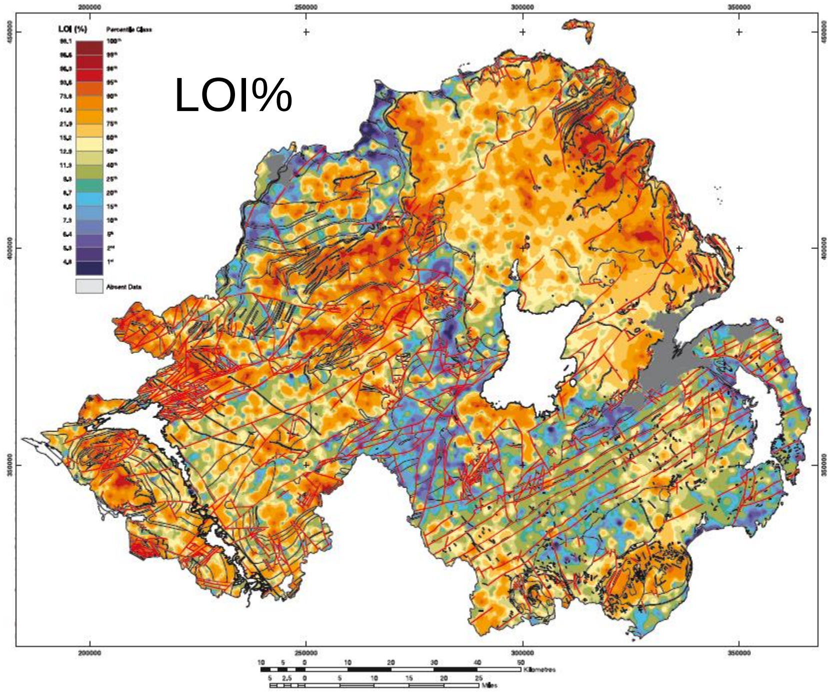
Environmental Management







GSNI
Superficial Geology

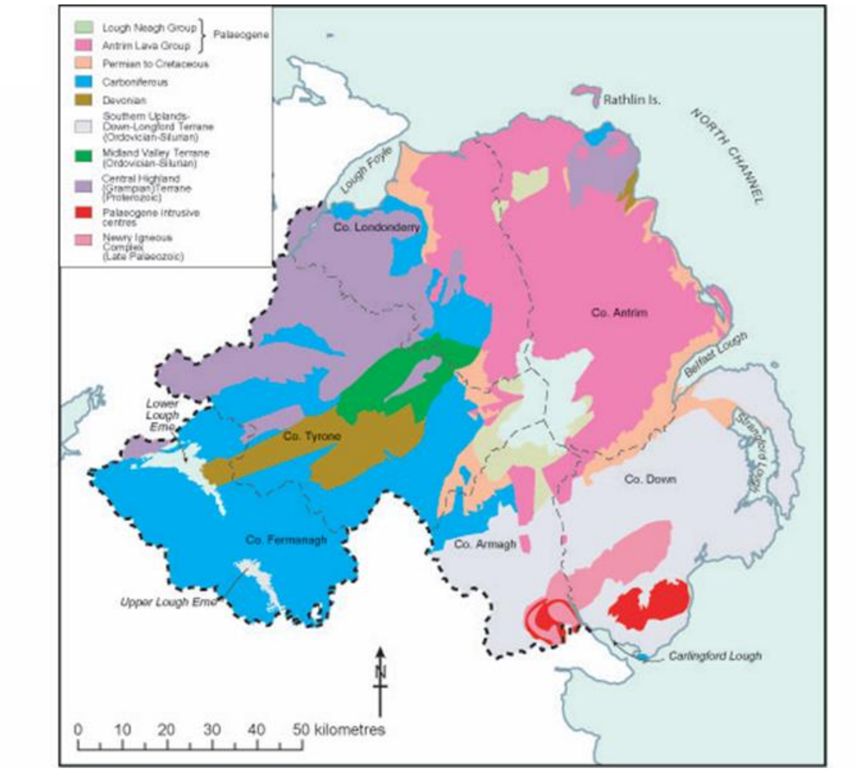


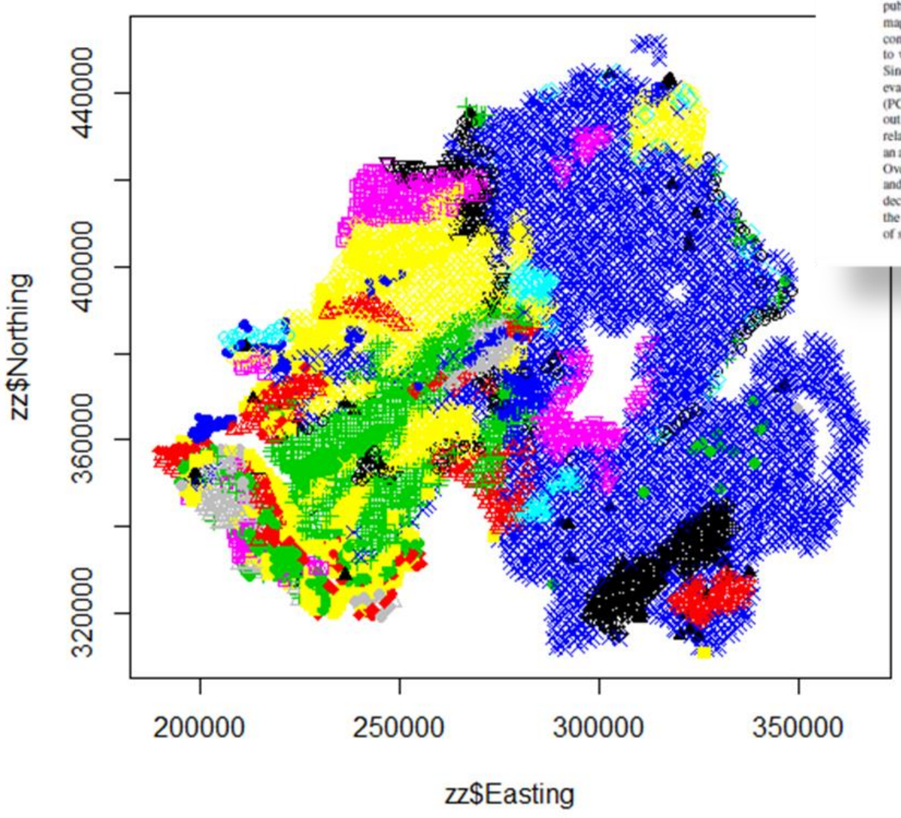
LOI%

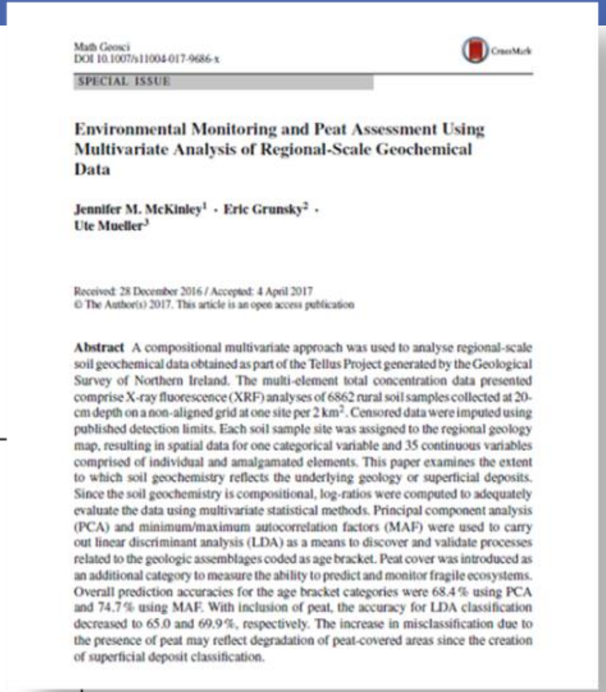
High LOI values are typical of peat substrates in both lowland and mountainous terrain.

Compositional PCA

Soil sample sites assigned to regional geology (GSNI) using the dominant lithology for the map polygon







Math Geosci
DOI 10.1007/s11004-017-9686-x
SPECIAL ISSUE

Environmental Monitoring and Peat Assessment Using Multivariate Analysis of Regional-Scale Geochemical Data

Jennifer M. McKinley[1] · Eric Grunsky[2] · Ute Mueller[3]

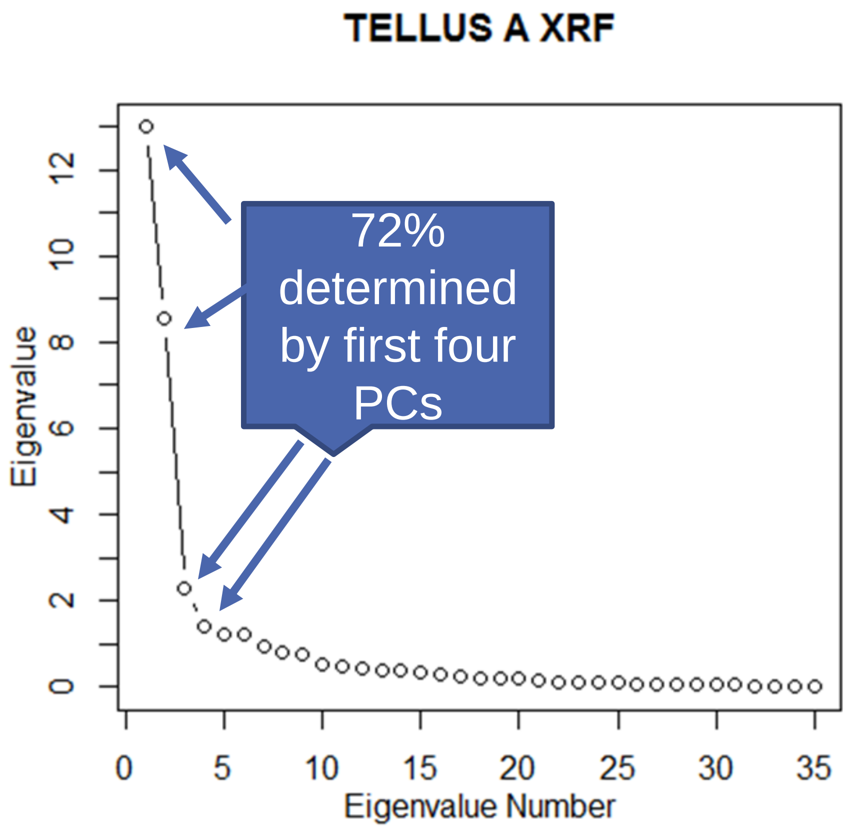
Received: 28 December 2016 / Accepted: 4 April 2017

Abstract A compositional multivariate approach was used to analyse regional-scale soil geochemical data obtained as part of the Tellus Project generated by the Geological Survey of Northern Ireland. The multi-element total concentration data presented comprise X-ray fluorescence (XRF) analyses of 6862 rural soil samples collected at 20-cm depth on a non-aligned grid at one site per 2 km^2. Censored data were imputed using published detection limits. Each soil sample site was assigned to the regional geology map, resulting in spatial data for one categorical variable and 35 continuous variables comprised of individual and amalgamated elements. This paper examines the extent to which soil geochemistry reflects the underlying geology or superficial deposits. Since the soil geochemistry is compositional, log-ratios were computed to adequately evaluate the data using multivariate statistical methods. Principal component analysis (PCA) and minimum/maximum autocorrelation factors (MAF) were used to carry out linear discriminant analysis (LDA) as a means to discover and validate processes related to the geologic assemblages coded as age bracket. Peat cover was introduced as an additional category to measure the ability to predict and monitor fragile ecosystems. Overall prediction accuracies for the age bracket categories were 68.4 % using PCA and 74.7 % using MAF. With inclusion of peat, the accuracy for LDA classification decreased to 65.0 and 69.9 %, respectively. The increase in misclassification due to the presence of peat may reflect degradation of peat-covered areas since the creation of superficial deposit classification.

PCA Data clr transformed

PCA showed that the first two eigenvalues (PCs) are the most significant and 72% of the variation was determined by the first four principal components (PCs) implying "significant" structure in the data.

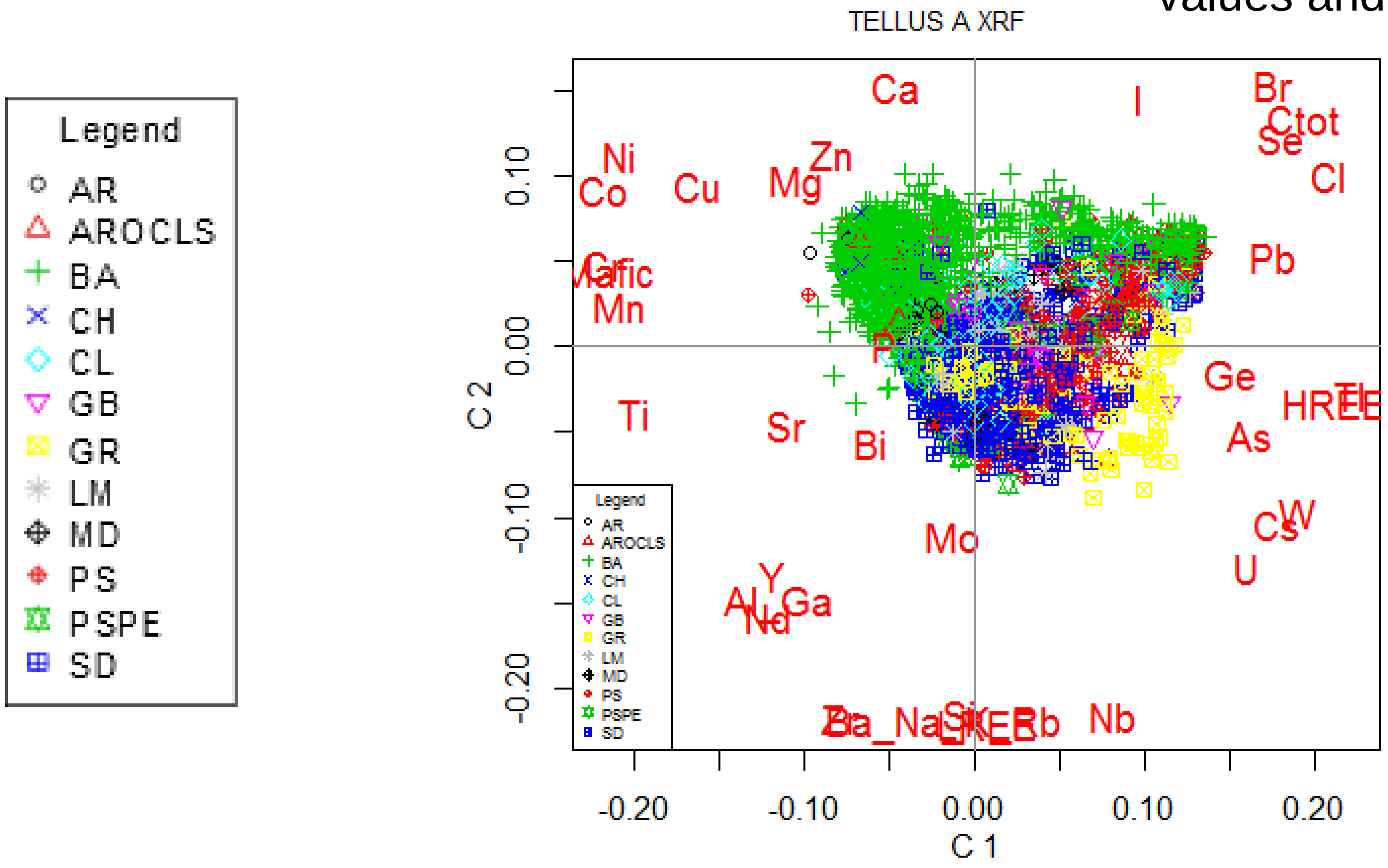
RQPCA [clr] TELLUS A (XRF)				
	PC1	PC2	PC3	PC4
λ	16.34	12.01	3.14	1.72
$\lambda\%$	35.52	26.11	6.83	3.74
$\Sigma\lambda\%$	35.52	61.63	68.46	72.20



PCA Biplot PC1-PC2

Based on 46 variables

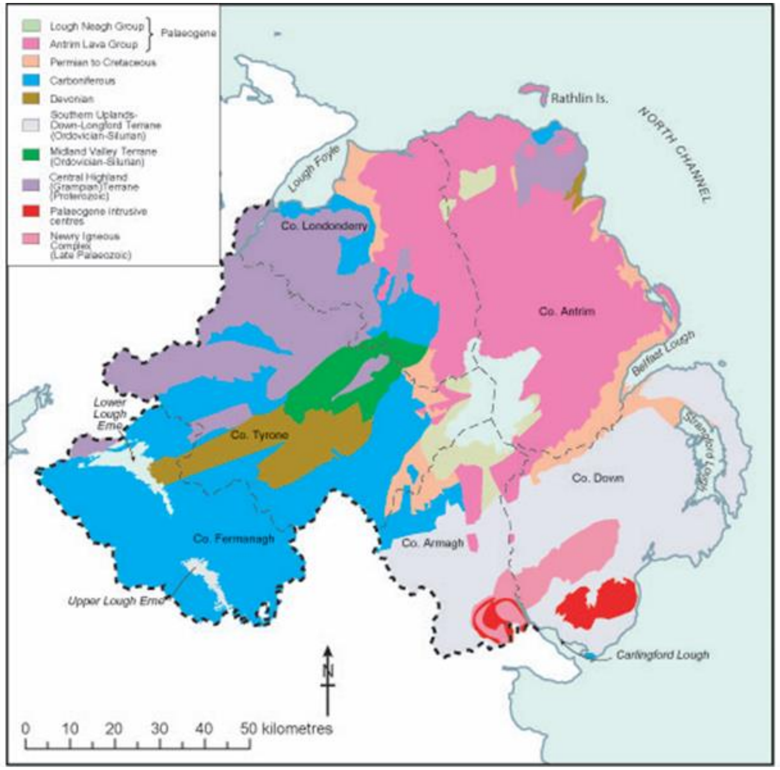
The positioning of elements in the biplots reflect their positive or negative values and relative contributions.

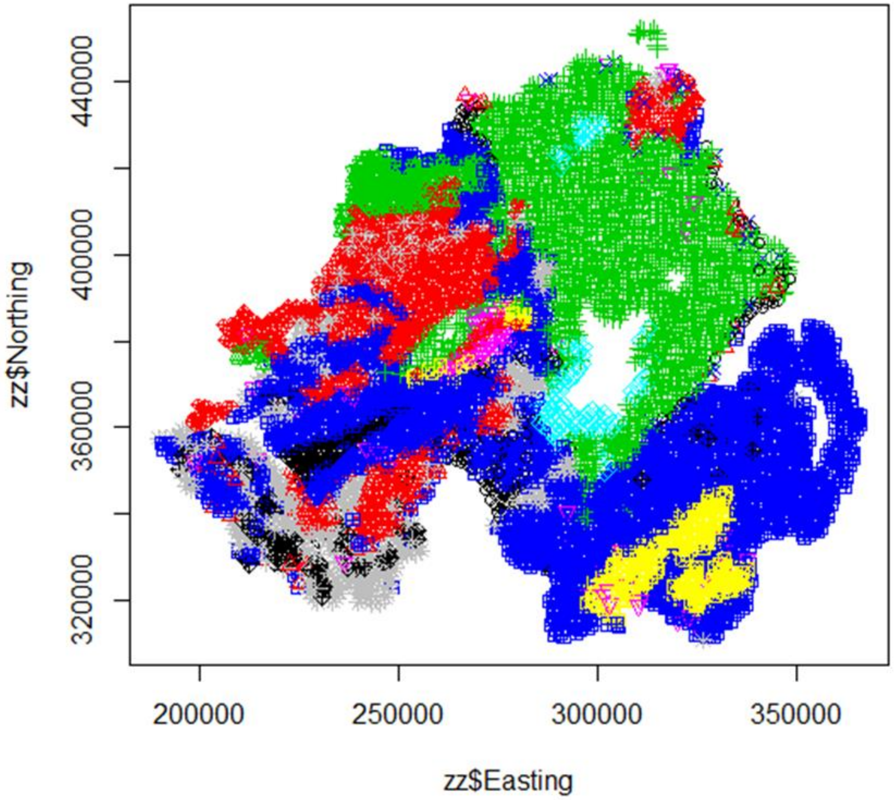


Reasonably discriminant between major lithologies

Minimum/maximum autocorrelation factor (MAF) Analysis

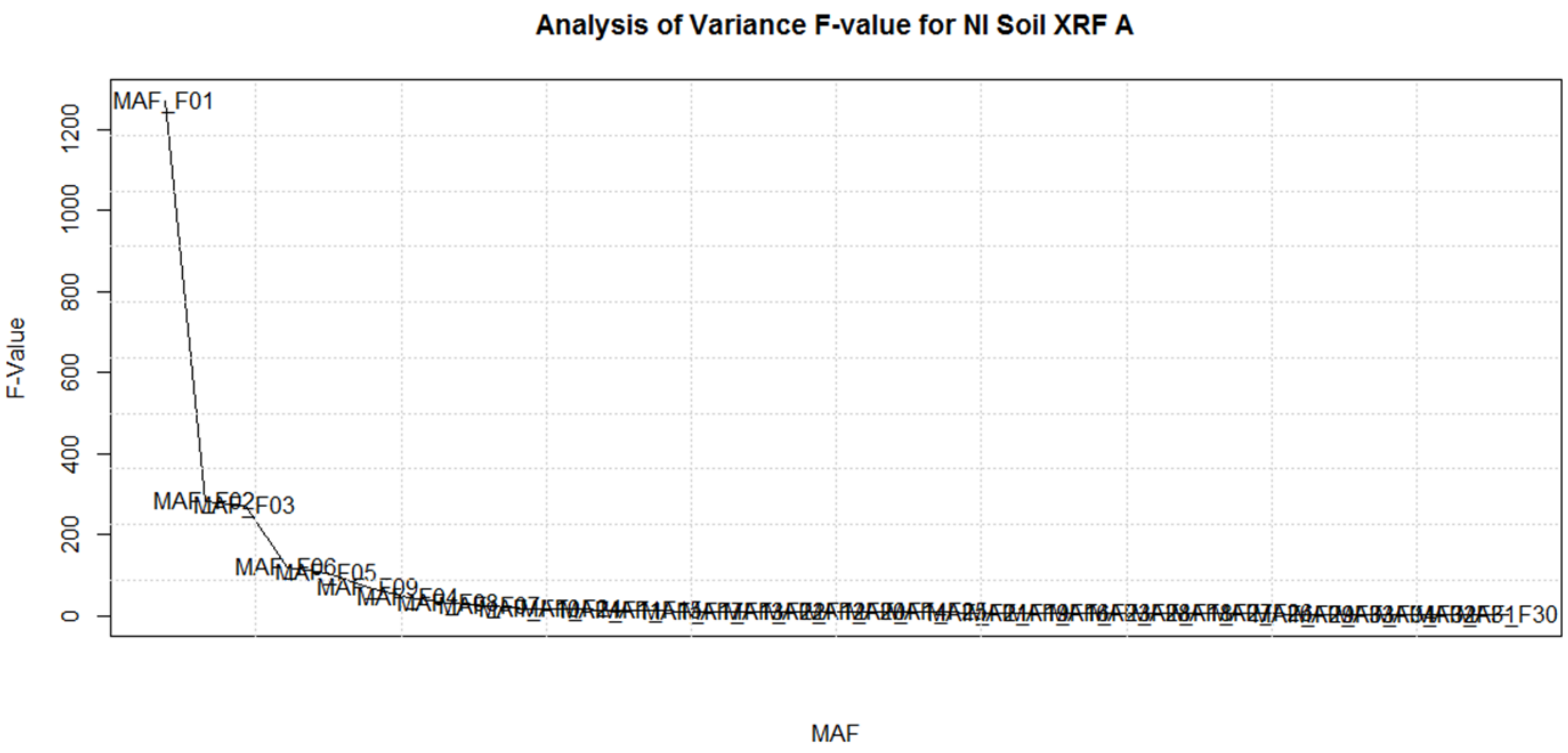
To consider an improvement over PCA that uses the spatial relationships of the data, a classification based on MAF analysis was undertaken.





MAF analysis (work of Ute Mueller)

- The usual procedure for determining the MAF transformation matrix is via two successive principal component analyses.
- The first PCA is performed on the sample correlation matrix. The linear model of coregionalisation is transformed by the product of the matrix of eigenvectors and the square root of the corresponding matrix of eigenvalues.
- The second PCA is performed on one of the transformed coregionalisation matrices.
- The MAF transformation matrix is built as the product of the matrix of eigenvectors and the square root of the corresponding matrix of eigenvalues from the first PCA and the matrix of eigenvectors from the second PCA.
- *Compositional MAF analysis conducted in this study on clr transformed data*



Analysis of variance showed that only 10 PC’s were necessary to classify the soil geochemical data. MAF analysis used the first 6 dominant factors.

Usefulness for environmental monitoring

- Understanding the relationship between soil geochemistry and superficial deposits is important for environmental monitoring of fragile ecosystems such as peat.

- To explore whether peat cover could be predicted from the classification, the lithology designation was adapted to include the presence of peat, based on GSNI superficial deposit polygons.
- Linear discriminant analysis (LDA) was undertaken.

PCA based Classification (LDA) Accuracy (60.98%)

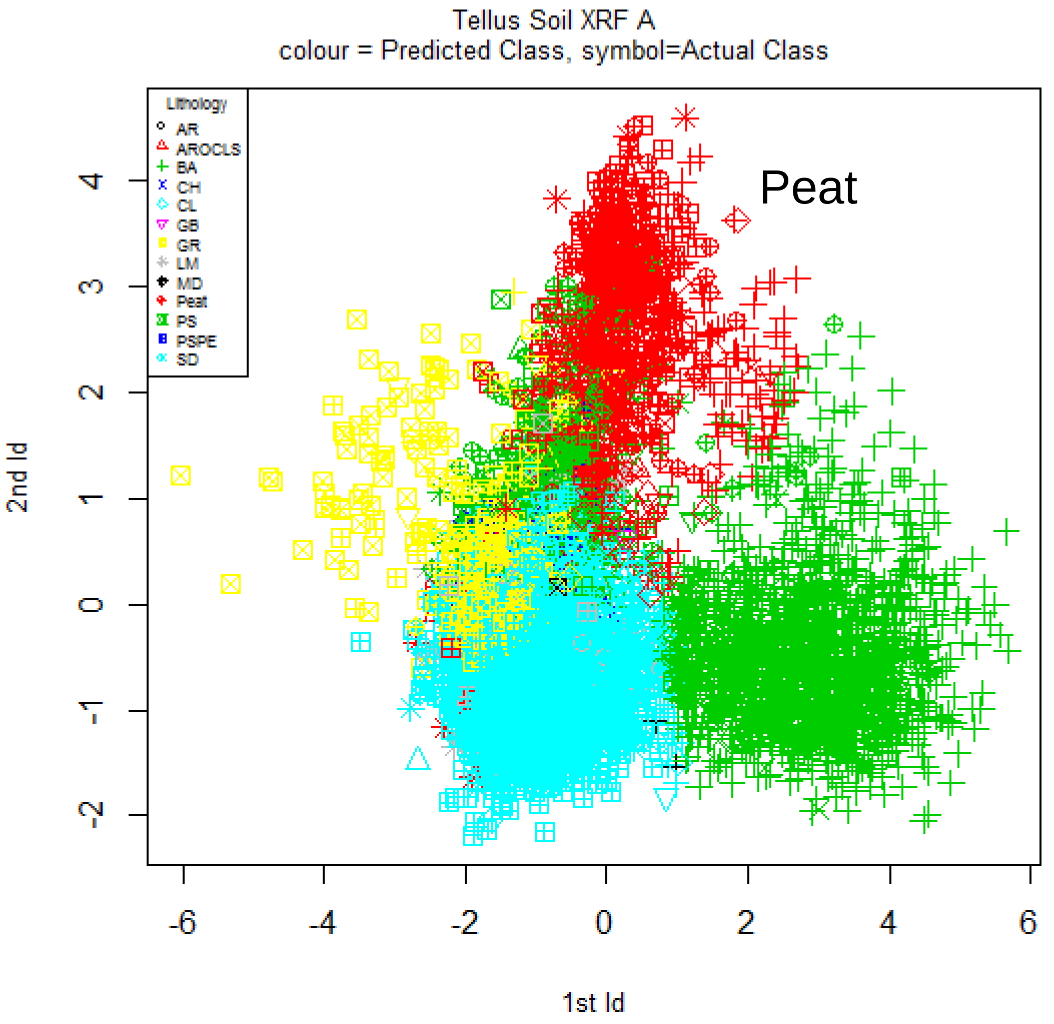
predicted

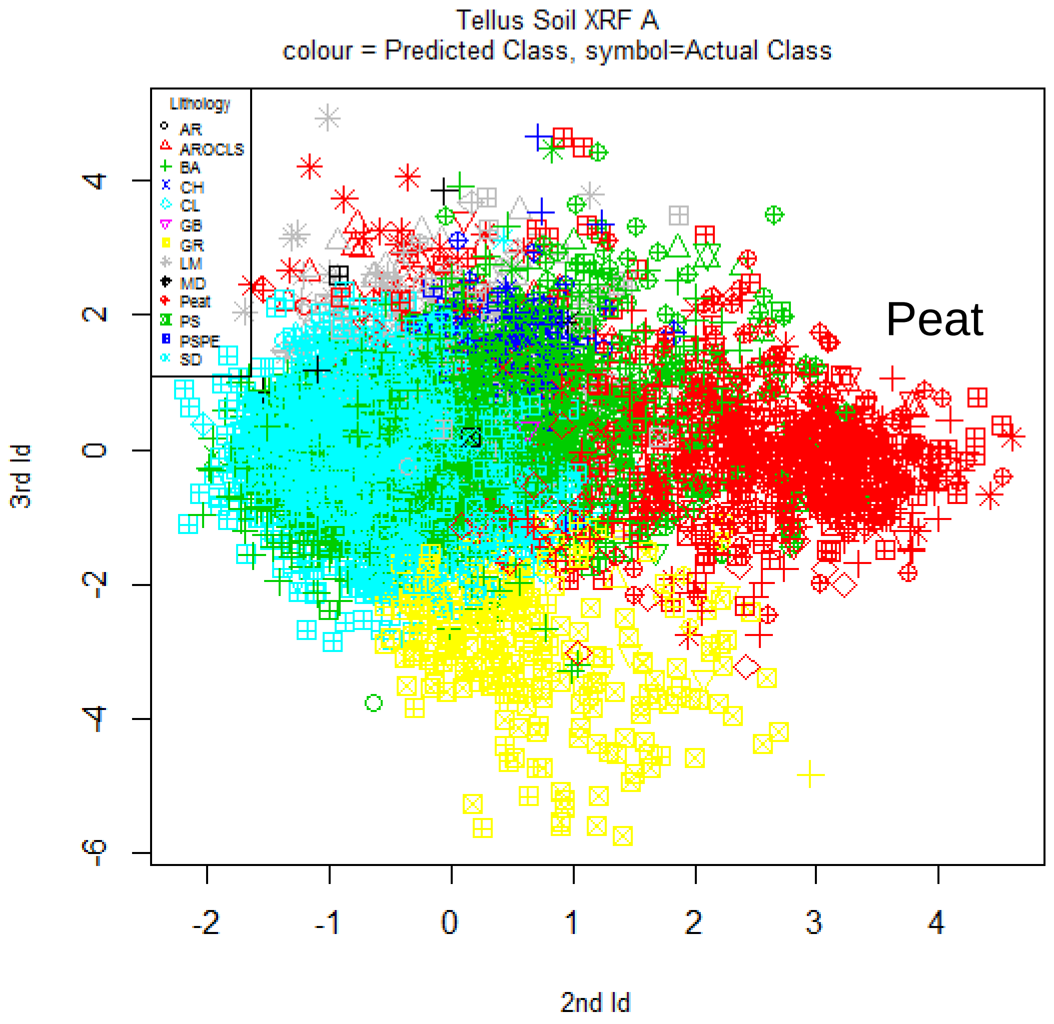
	AR	AROCLS	BA	CH	CL	GB	GR	LM	MD	**Peat**	PS	PSPE	SD
AR	0.0	1.8	14.3	0.0	0.0	0.0	0.6	10.7	0.0	0.6	1.2	0.0	70.8
AROCLS	0.0	16.6	4.9	0.0	0.0	0.0	1.5	12.2	0.0	2.4	2.4	0.0	60.0
BA	0.1	0.0	84.8	0.0	0.4	0.0	0.0	0.2	0.1	5.9	0.9	0.1	7.6
CH	0.0	2.7	67.6	0.0	0.0	0.0	0.0	0.0	0.0	10.8	2.7	0.0	16.2
CL	0.0	0.9	29.2	0.0	0.9	0.0	0.9	0.0	0.0	7.1	4.4	0.0	56.6
GB	0.0	2.2	7.5	0.0	0.0	0.0	9.7	4.3	0.0	12.9	15.1	0.0	48.4
GR	0.0	0.0	0.0	0.0	0.0	0.0	61.7	1.6	0.0	4.7	2.8	0.0	29.2
LM	0.0	6.8	0.6	0.0	0.2	0.0	0.2	30.5	0.0	3.5	7.8	0.6	49.7
MD	0.0	5.5	1.1	0.0	0.0	0.0	0.0	38.1	0.0	2.2	1.1	0.0	51.9
Peat	0.0	2.6	15.6	0.0	0.0	0.0	1.6	3.3	0.0	**48.4**	10.0	0.1	18.3
PS	0.0	0.2	1.1	0.0	0.0	0.0	2.1	1.9	0.0	18.9	41.9	3.0	31.0
PSPE	0.0	0.8	0.0	0.0	0.0	0.0	0.0	3.0	0.0	3.0	42.4	6.1	44.7
SD	0.0	1.2	1.1	0.0	0.1	0.0	2.7	2.6	0.0	2.6	4.3	0.2	85.1

error rate = 39.02 %

Linear Discriminant Plots

Prediction accuracy for LDA classification 60.98% based on PCA using 10 principal components





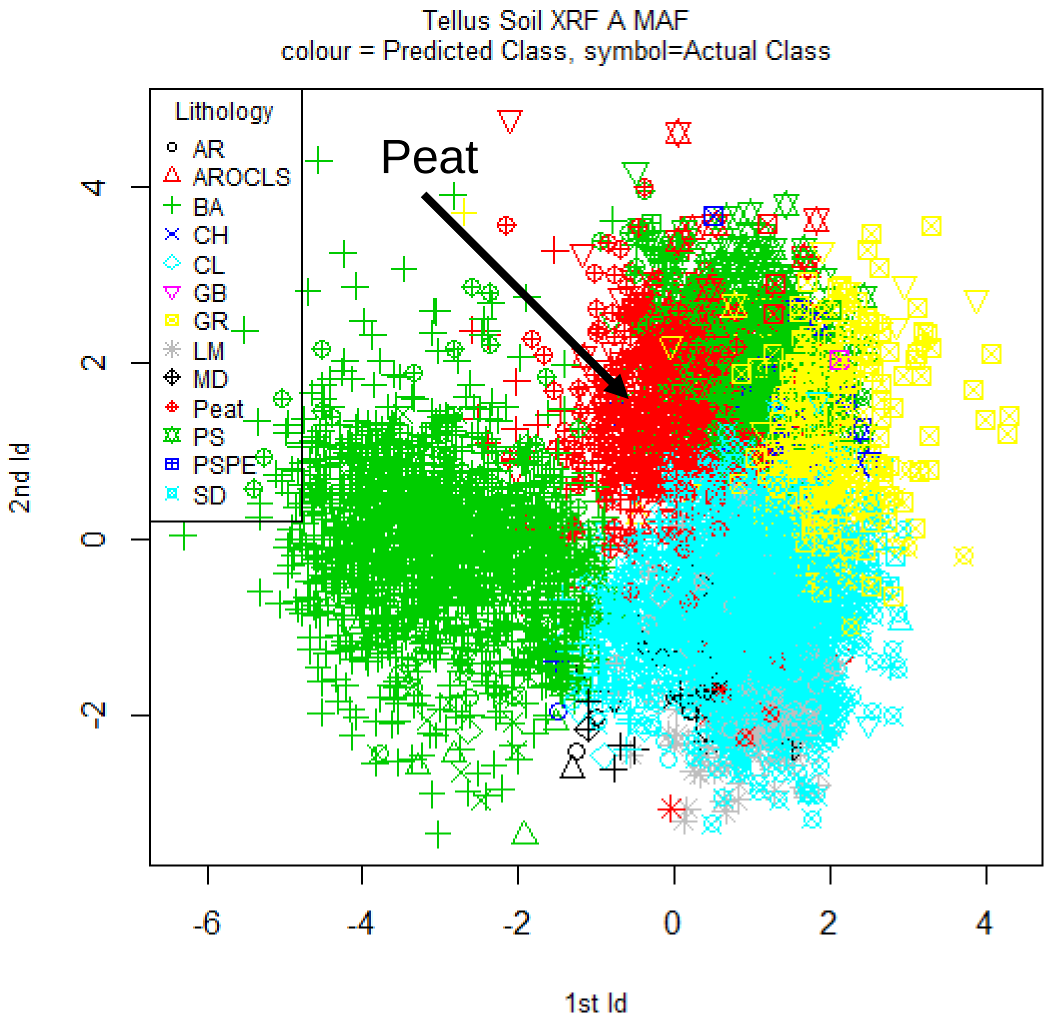
MAF based Classification (LDA) Accuracy (64.37%)

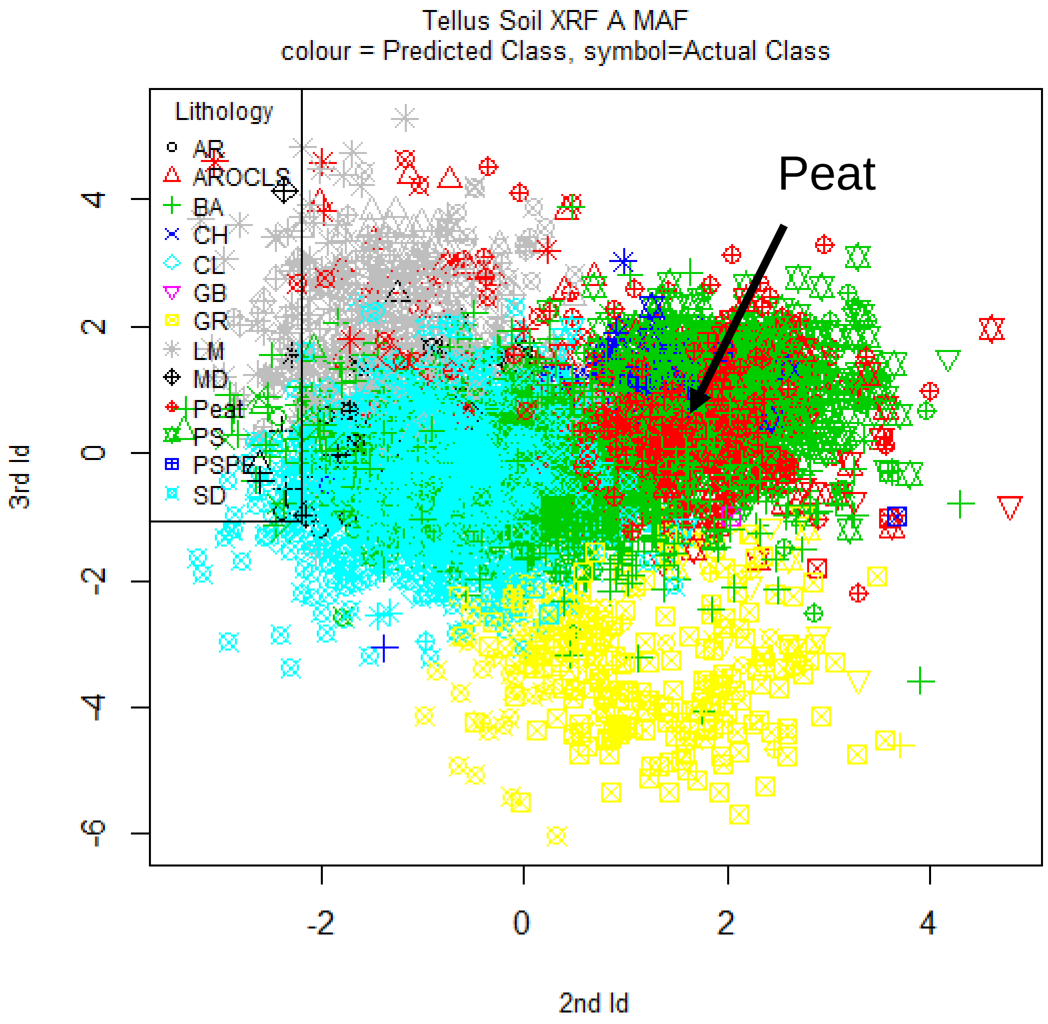
predicted

	AR	AROCLS	BA	CH	CL	GB	GR	LM	MD	**Peat**	PS	PSPE	SD
AR	3.57	4.17	7.74	3.57	2.38	0.00	0.00	19.05	0.00	0.00	0.60	0.00	58.93
AROCLS	0.49	8.78	3.90	0.49	0.98	0.00	0.98	31.71	0.49	4.88	0.49	0.00	46.83
BA	0.53	0.00	86.84	0.27	1.66	0.00	0.07	0.33	0.00	3.79	1.20	0.00	5.32
CH	5.41	0.00	59.46	0.00	8.11	0.00	0.00	0.00	0.00	10.81	0.00	0.00	16.22
CL	5.31	0.88	30.97	0.00	19.47	0.00	0.00	4.42	0.00	3.54	0.00	0.00	35.40
GB	1.08	0.00	5.38	0.00	0.00	0.00	12.90	5.38	0.00	23.66	13.98	0.00	37.63
GR	0.00	0.00	0.00	0.00	0.00	0.00	75.89	0.00	0.00	5.93	1.98	0.40	15.81
LM	1.03	2.27	0.62	0.00	0.62	0.00	0.00	46.39	0.82	2.47	10.93	2.68	32.16
MD	0.55	1.10	0.00	0.55	0.00	0.00	0.00	48.07	2.76	0.55	1.10	0.00	45.30
Peat	0.36	1.19	17.18	0.00	0.95	0.00	1.43	4.89	0.36	**46.42**	12.29	0.00	14.92
PS	0.00	0.00	0.75	0.00	0.00	0.00	1.12	0.00	0.00	21.31	62.06	4.86	9.91
PSPE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.03	0.00	6.82	62.88	15.15	12.12
SD	0.60	0.79	0.79	0.00	0.60	0.05	2.37	3.48	0.19	3.53	3.90	0.28	83.42

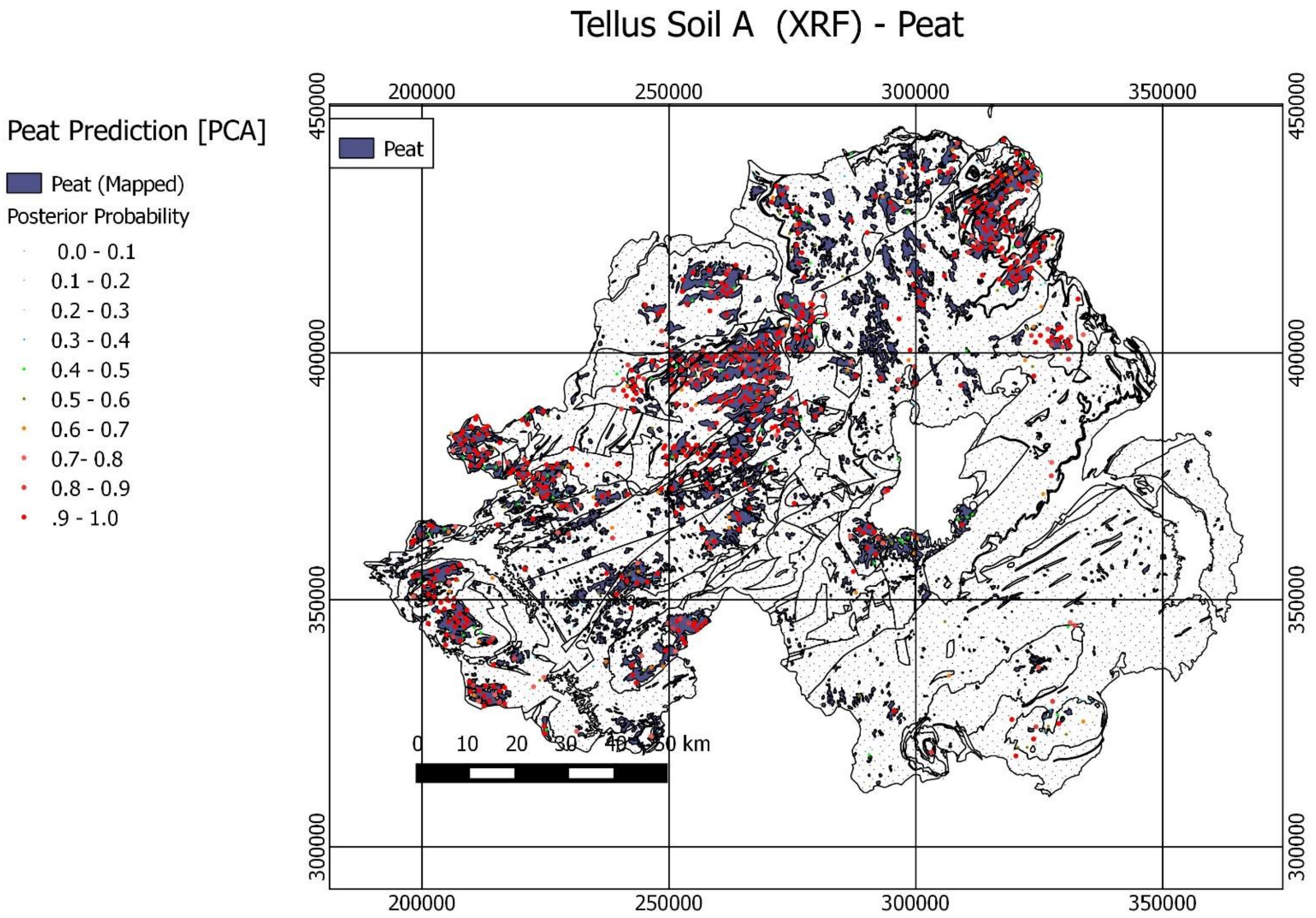
error rate = 35.63 %

Linear Discriminant Plots

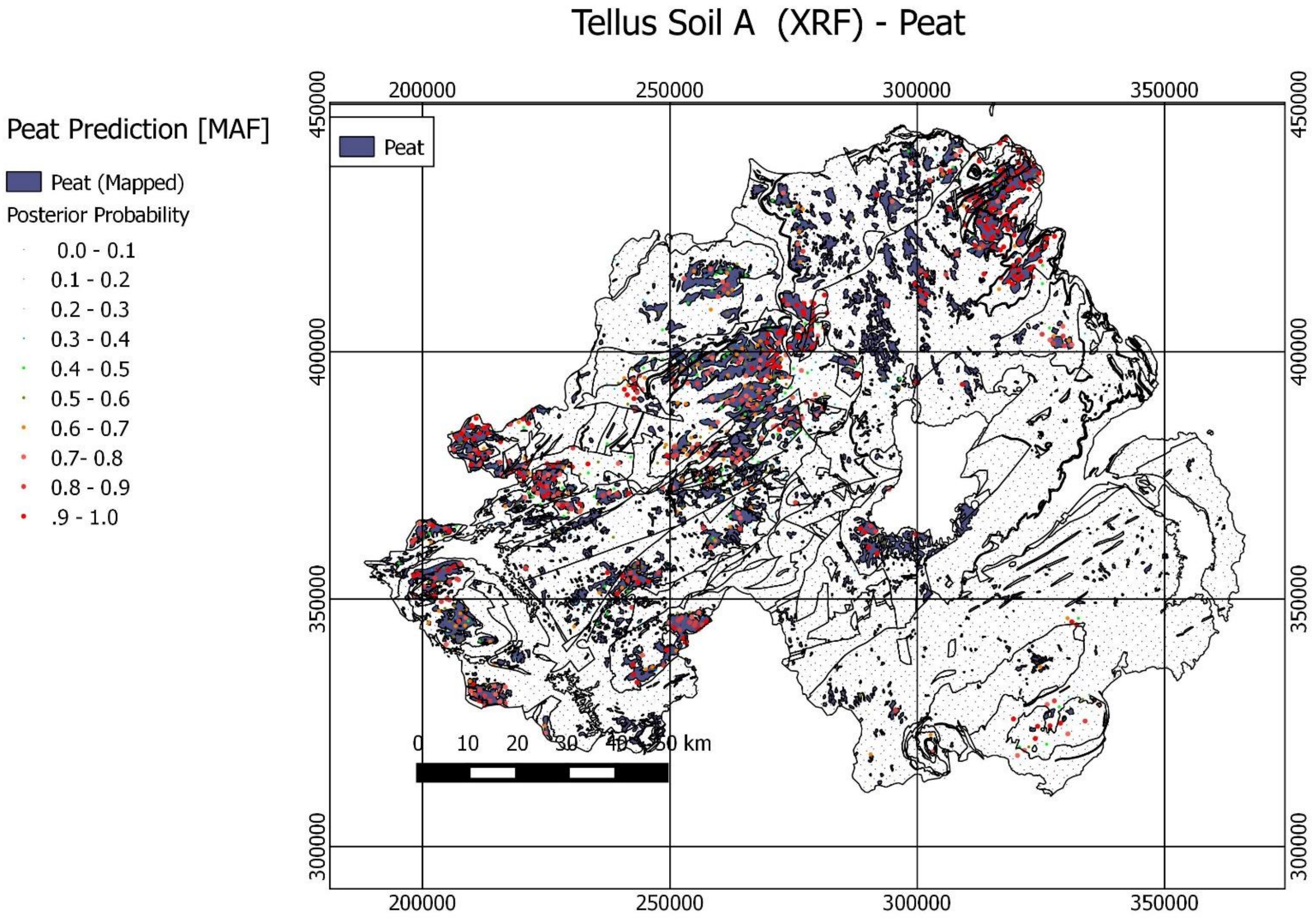




Predictive Map of Peat based on PCA



Predictive Map of Peat based on MAF



Conclusions

- The plotted PCA and MAF typicalities demonstrate a good match between the reported peat areas and the highest probability or typicality for peat.
- However, there are areas of mapped peat where the predictions indicate a low probability of peat.
- The explanation for the misclassification of peat:
  - ➢ These areas reflect degradation of peat covered areas since the creation of superficial deposit classification
  - ➢ A further refinement in the classification of peat is required.
- The prediction of peat covered areas using MAF analysis methods, which use the spatial relationships of the data, has been more successful in predicting the more extensive upland blanket bogs than lowland raised bogs.

DECOUPLING PROCESSES FROM SOIL GEOCHEMISTRY: MAPPING THE SURFICIAL/BEDROCK GEOCHEMICAL SIGNATURES IN NORTHERN IRELAND

Grunsky, E.C [12]., McKinley, J.M. [3], Mueller, U.A. [4]

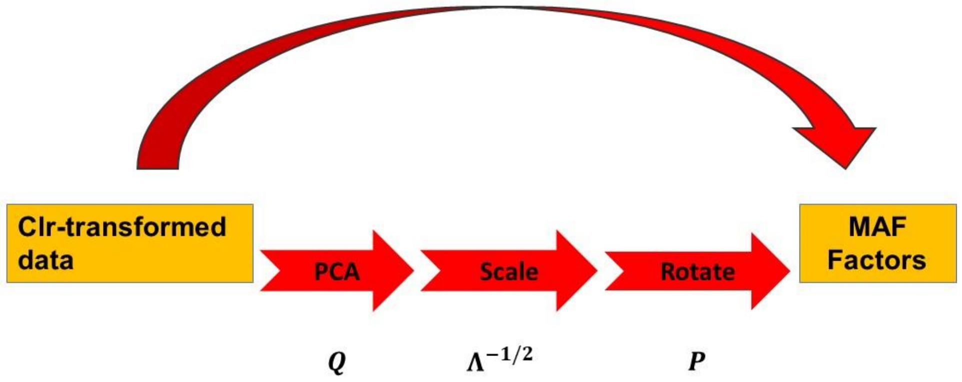
[1] China University of Geosciences, Beijing, China
[2] Earth and Environmental Sciences, University of Waterloo, Waterloo, Canada
[3] School of Natural and Built Environment, Queen's University Belfast (QUB), UK
[4] School of Science, Edith Cowan University Jonndalup, WA, Australia

Minimum/Maximum Autocorrelation Factor Analysis (MAF) & Random Forest Classification

- **MAF analysis** is a generalised eigenvalue problem equivalent to the application of a PCA followed by a further rotation derived from a covariance matrix of increments of the standardized PCA factors.
- Remove spatial cross-correlation that yield factors that are approximately spatially uncorrelated
- Similar to PCA components, the MAF factors are made up of linear combinations of the original variables.

- The **Random Forest (RF)** method is based on the construction of classification trees in which nodes (splits in classes) are based on continuous variables from which a series of branches in the tree will correctly classify (categorical variables) all of the data.
- RF “grows” many trees and each tree provides a classification and termed a “vote”.

$$\boldsymbol{Y}_{\boldsymbol{MAF}}(u) = \boldsymbol{Z}_{\boldsymbol{NS}}(u)\boldsymbol{Q}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}$$



The Process of Predictive Mapping

Process Prediction using MAF and Random Forests

The use of modelled methods for process confirmation:

- Established set of classes for prediction (Age Bracket, Lithology, Surficial)
- In this case the MAF metric is used.
- Select an appropriate lag interval(s) to test the classification based on Random Forest classification methodology. In this case 12.5km lag interval.
- Classification determines the votes and normalized votes for each class.
- Subsequent kriging or co-kriging (interpolation) produces predictive maps for each class.

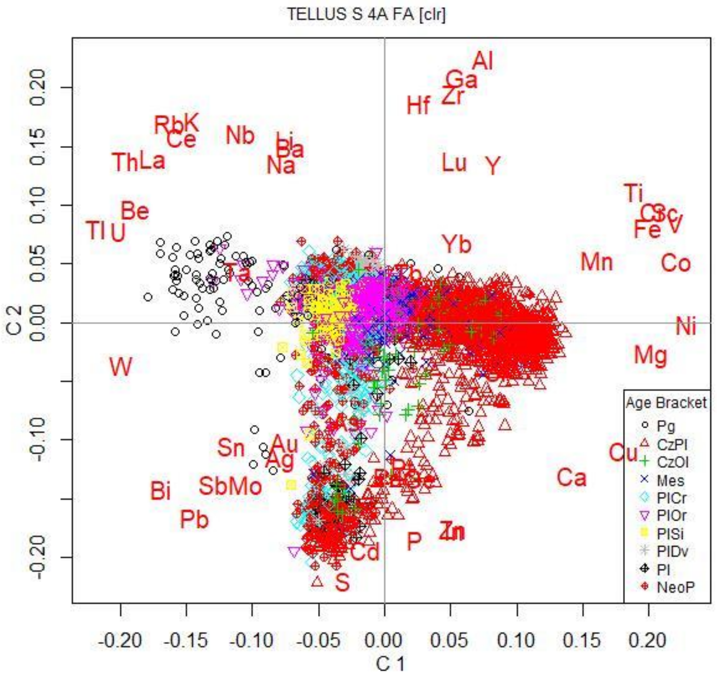
Exploring the Geochemical Signature of Peat

- The signature of Peat, as it is mapped in Northern Ireland is consistent with the “organic signature” shown in the principal component biplots of the Tellus soil geochemistry.
- An examination of the organic signature may provide a better understanding of the processes that are reflected in the soil geochemistry.
- Filtering out the organic signature should help in identifying the underlying lithologies and processes associated with sources of metals and their bioavailability (adsorbed or locked in crystal lattices).
- Studies of the organic signature may be useful in understanding environmental issues, agriculture, groundwater and population health.

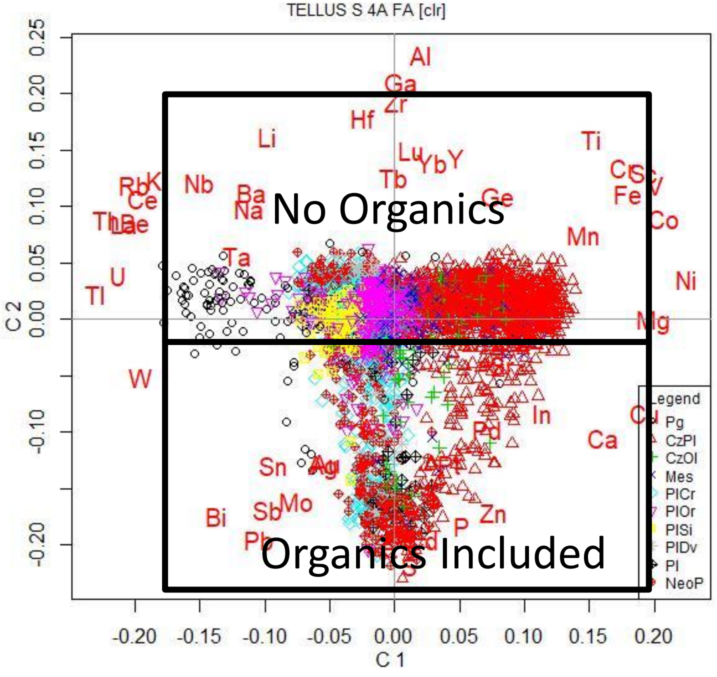
MAF based on: Organics / No Organics

- Distinction between processes may not be evident in biplots of PCs.
- An adhoc adjustment through rotation improves the identification of processes.
- No Organics were selected on the basis of Select samples where rotated PC2 > -0.025
- MAF is computed on clr-transformed elements using the separation criteria above.

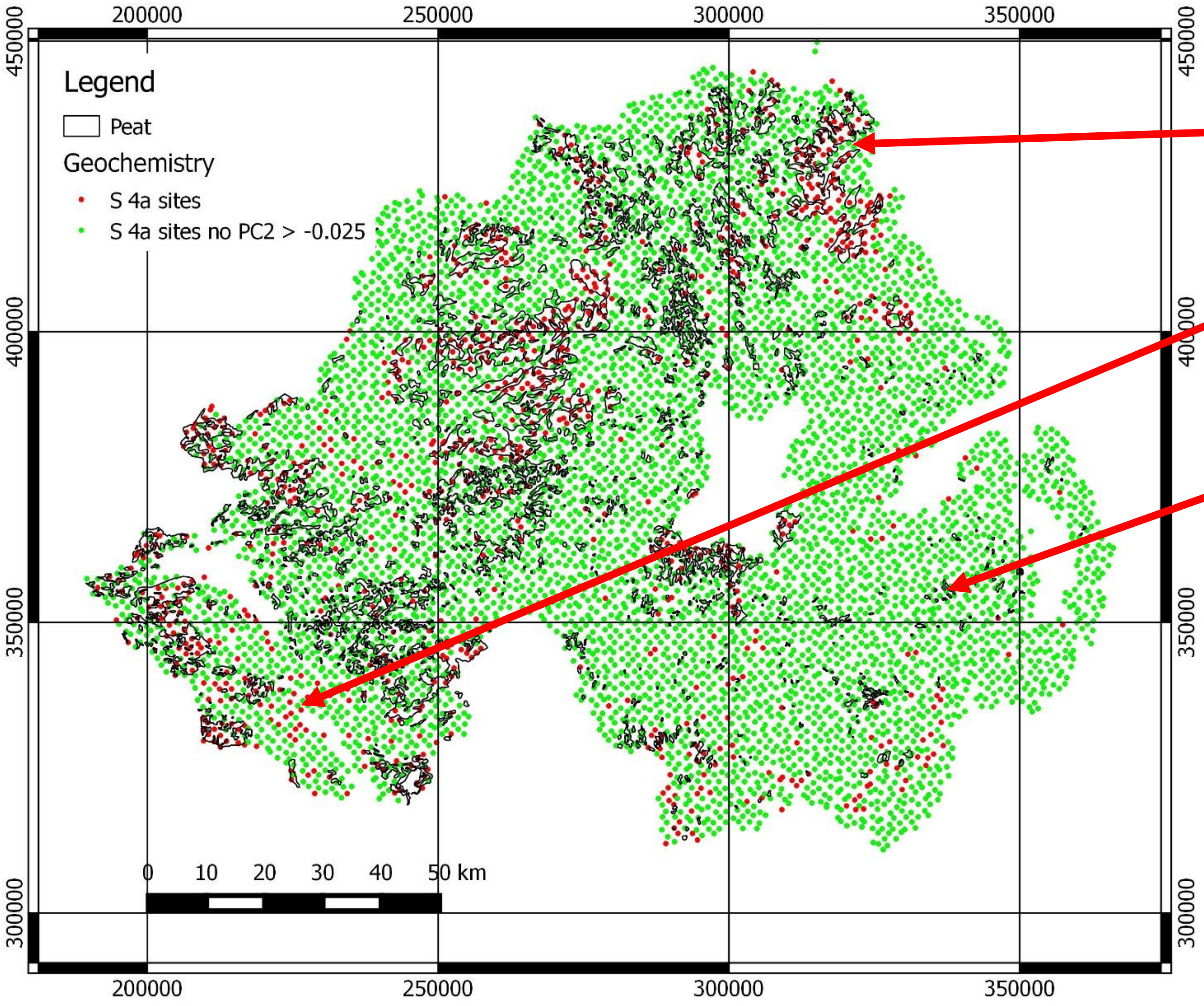
Original



Rotated



Organic Locations ≅ Chalcophile-rich Sites



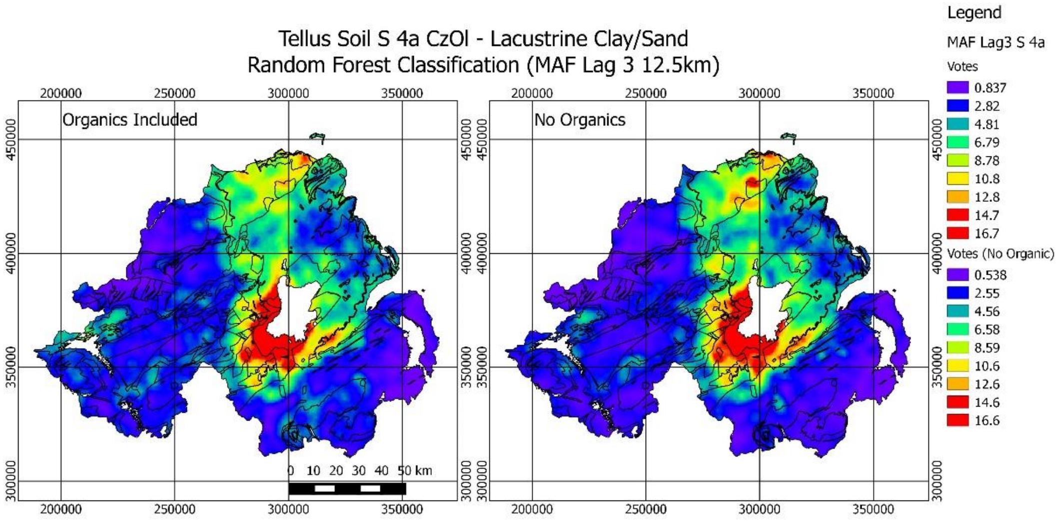
Many chalcophile-rich sites are coincident with Peat localities.

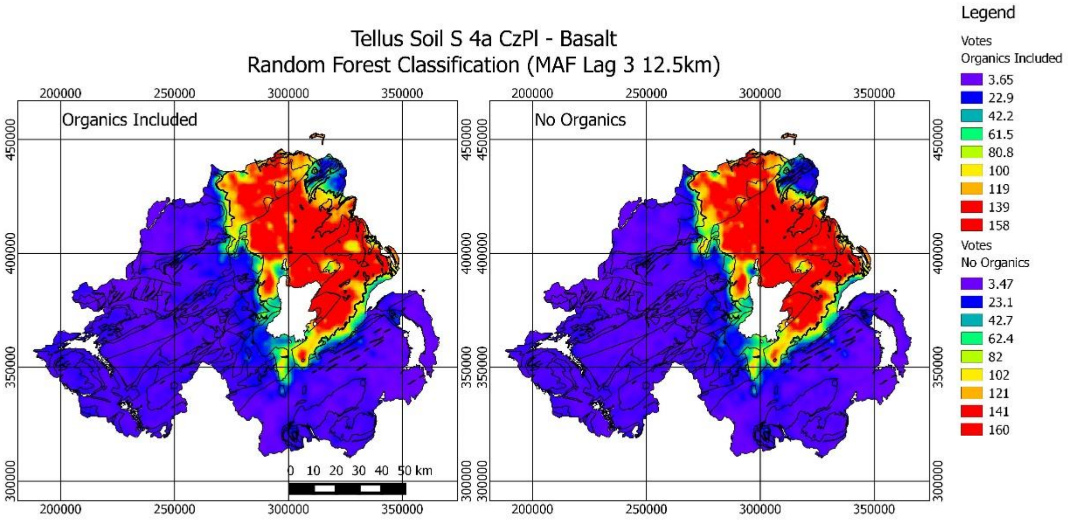
Some chalcophile-rich sites are not associated with Peat.

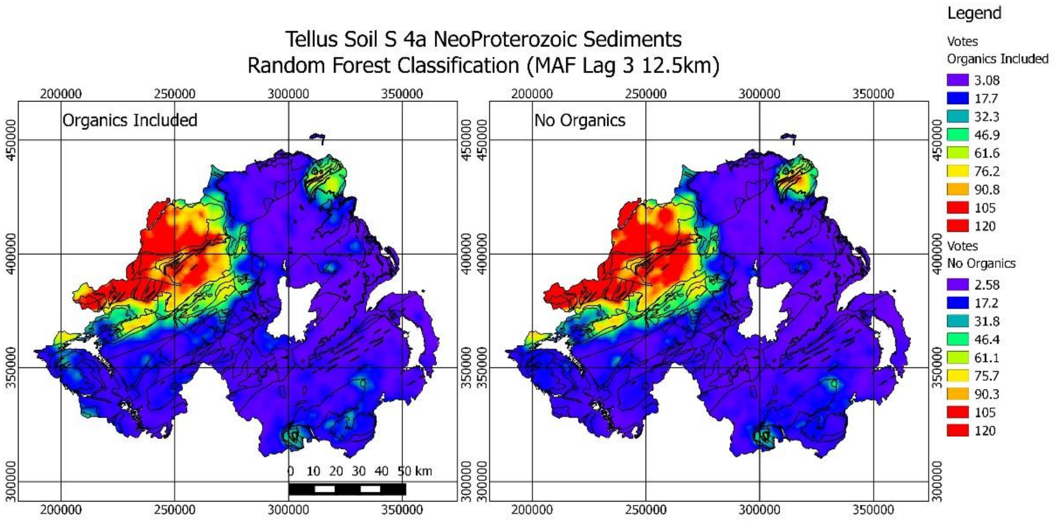
Some Peat sites are not associated with chalcophile-rich elements.

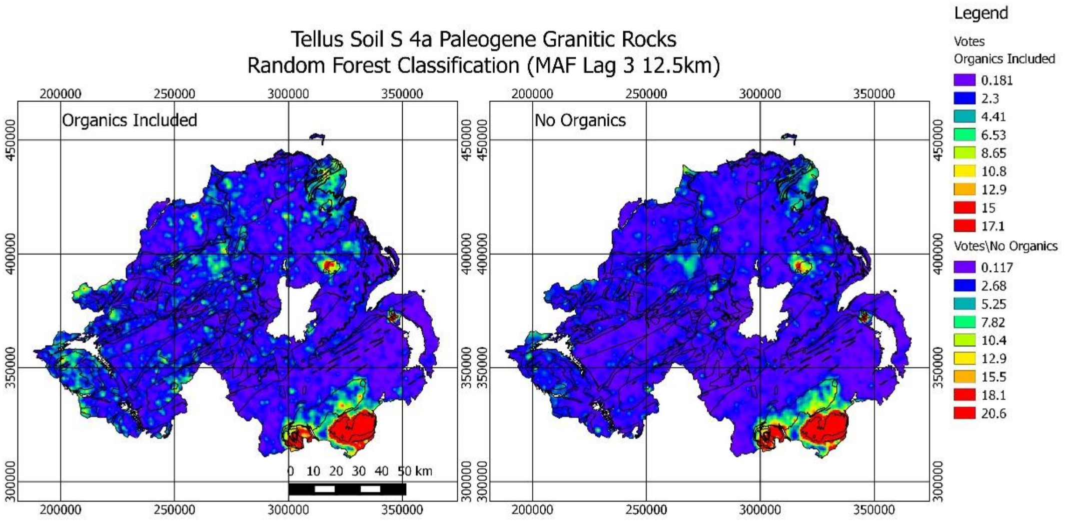
No distinction between raised bogs and blanket bogs.

Random Forest Prediction
S 4a MAF Lag 3 (12,500m)



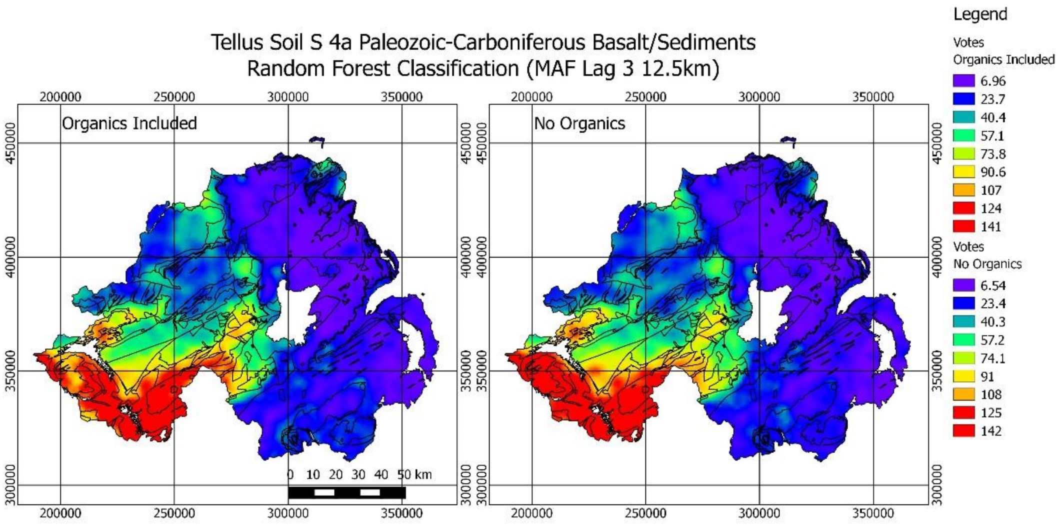


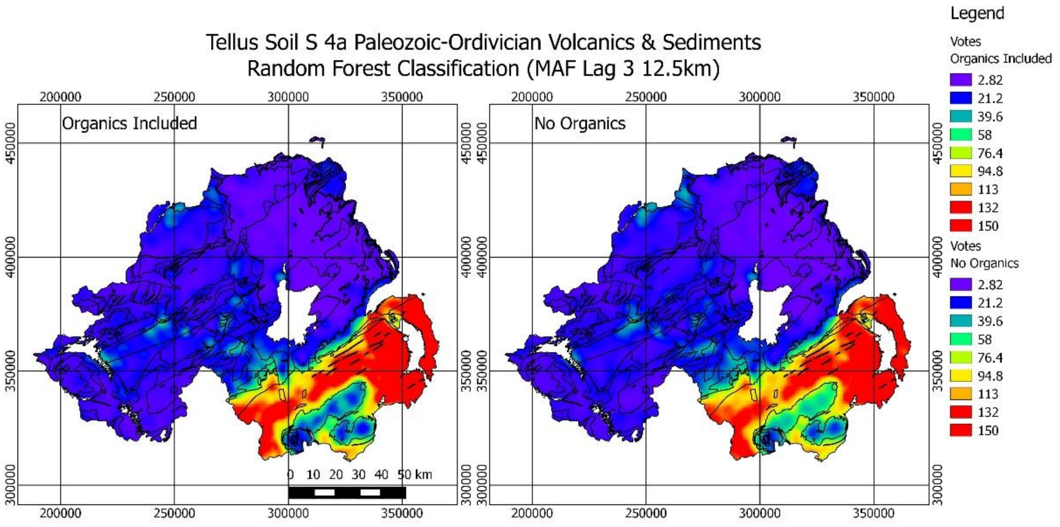


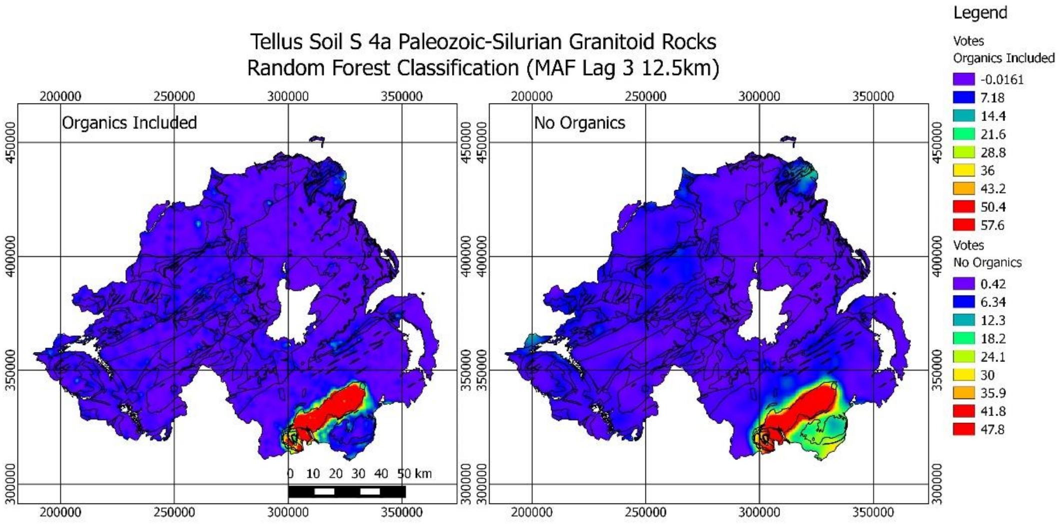


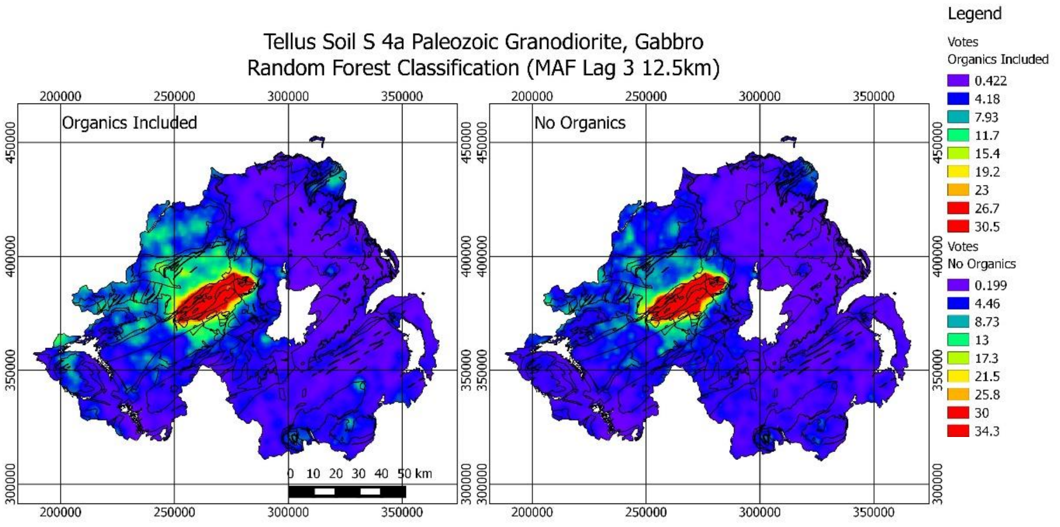
Process Validation

Random Forest Prediction
S 4a MAF Lag 3 (12,500m)









Process Validation

Interpretation and Future Work

- Soil geochemistry, regardless of the location in the soil profile, contains information on processes that formed the rock forming minerals and subsequent modifications. The surface environment has a distinct multi-element signature as shown in the PCA biplots and map images.
- Using multivariate statistical methods applied to logratio transforms, these processes can be discovered and validated
- Organic content determined from the PC1-PC2 biplot includes sites where peat occurs and also sites where shales/mudstones occur with similar chalcophile enrichment. Separation of the soil geochemistry based on perceived “organic” content does not demonstrate a distinctive difference in the classification of Age Brackets.
- Further work required in separating peat and peaty soils containing chalcophile elements and shale-like lithologies rich in chalcophile elements.

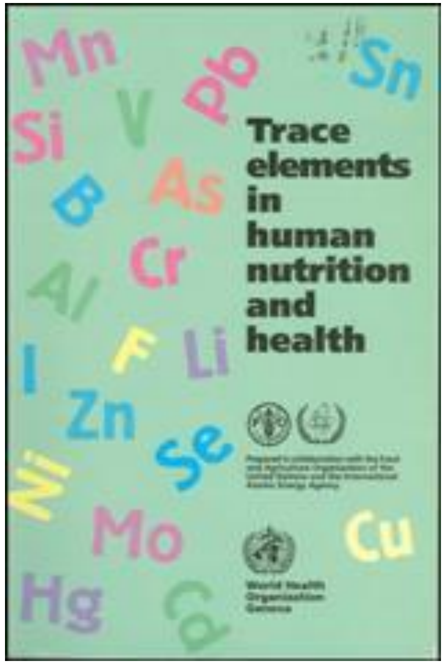
Environment and Health

Compositional analysis using balances of geochemical environmental toxins to explore potential associations with chronic kidney disease

J.M. McKinley[1], S. Cox[1], U. Mueller[2], P.M. Atkinson[1,3], U. Ofterdinger[1], Siobhan F. Cox[1], Rory Doherty[1] ,D.Fogarty[4] , J.J. Egozcue[5], V. Pawlowsky-Glahn[6]

[1]School of Natural and Built Environment, Queen's University Belfast
[2] School of Science, Edith Cowan University, Perth, Western Australia
[3]Lancaster University, UK
[4]Belfast Health Trust, Belfast, Northern Ireland
[5]Dept. Civil and Environmental Engineering, U. Politécnica de Cataluña (UPC), Barcelona, Spain
[6] Dep. Computer Sciences, Applied Mathematics, and Statistics, University of Girona, Spain

People, Place and Health



WHO 1996

The WHO divides trace elements into three groups on their nutritional significance in humans:
(1) **essential elements**: Iodine, zinc, selenium, copper, molybdenum, chromium
(2) elements which **are probably essential**: manganese, silicon, nickel, boron, vanadium
(3) **potentially toxic elements (PTEs)** (some have essential functions at low levels): Fluoride, lead, cadmium, mercury, arsenic, aluminum, lithium and tin.

Chronic Kidney Disease of unknown causes (CKDu)

- **Chronic Kidney Disease (CKD),** a collective term for many causes of progressive renal failure is increasing worldwide due to ageing, obesity & diabetes.
- **Chronic kidney disease of unknown aetiology (CKDu)** which has been **linked to environmental factors** is a major concern worldwide.
- **No definitive cause has been found.** A high level task force has been convened by the World Health Organisation (WHO) to identify the environmental causes.
- Known nephrotoxins are lead (Pb), cadmium (Cd), mercury (Hg) and arsenic (As). **The link between heavy metals and CKDu remains to be established.**
- Factors that cause CKDu may be relevant to the heterogeneity of progressive CKD in diabetes & hypertension.


BMC Nephrology BMC

BMC Nephrol. 2015; 16: 145. PMCID: PMC4539684
Published online 2015 Aug 19. doi: 10.1186/s12882-015-0105-6 PMID: 26282933

Global dimensions of chronic kidney disease of unknown etiology (CKDu): a modern era environmental and/or occupational nephropathy?

Virginia M. Weaver,[1,2,3] Jeffrey J. Fadrowski,[3,4,5] and Bernard G. Jaar[2,3,5]

High level task force convened by the WHO

Dataset 1: United Kingdom Renal Registry (UKRR)

- The UKRR regularly collects data on all patients with advanced CKD on dialysis or with a kidney transplant (Renal Replacement Therapy (RRT)) across the UK
  - The UKRR provided Standardised Incidence Rates (SIRs) for patients starting RRT between 2006-2016, by Super Output Area (SOA).
  - Data were provided in age brackets (16-39, 40-64 and 65+, all ages >16 and for uncertain aetiology (CKDu) for 2006-2016).
- *SIRs of exactly 1 indicate that a SOA's incidence for RRT is equal to that expected based on Northern Ireland's average age specific incidence rates.*

SIRs above 1 -the incidence is higher than expected

SIRs below 1 - the incidence is lower than expected


Environ Geochem Health
https://doi.org/10.1007/s10653-020-00618-y

ORIGINAL PAPER

Chronic kidney disease of unknown origin is associated with environmental urbanisation in Belfast, UK

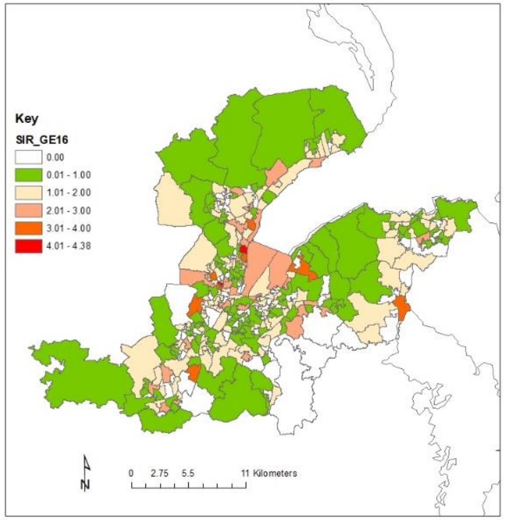
Jennifer M. McKinley · Ute Mueller · Peter M. Atkinson · Ulrich Ofterdinger · Siobhan F. Cox · Rory Doherty · Damian Fogarty · J. J. Egozcue · V. Pawlowsky-Glahn

Environ Geochem Health
https://doi.org/10.1007/s10653-020-00618-y

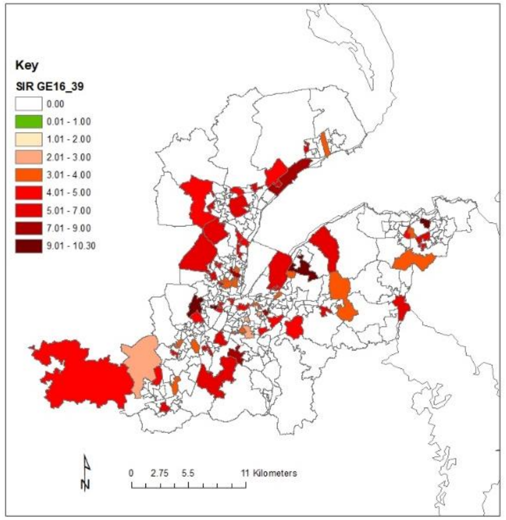
Mapping UKRR CKD Standardised Incidence Rates (SIRs) for Greater Belfast area

QUEEN'S UNIVERSITY BELFAST

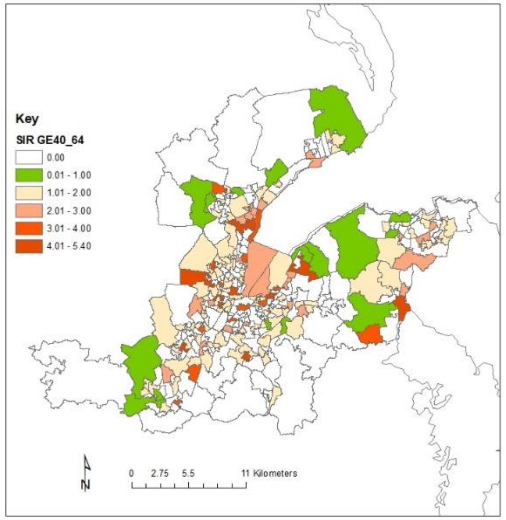
>16 yrs



16 to 39 yrs



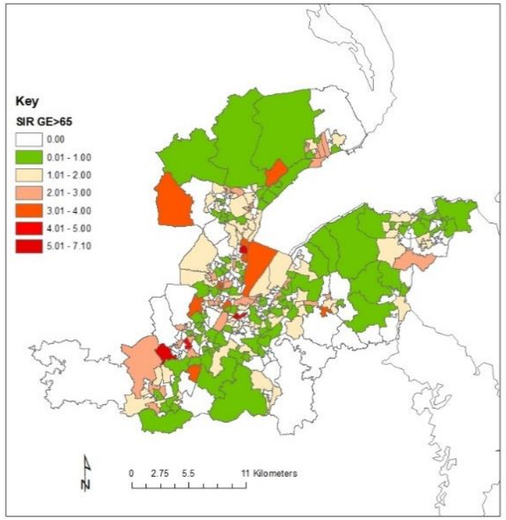
40 to 65 yrs



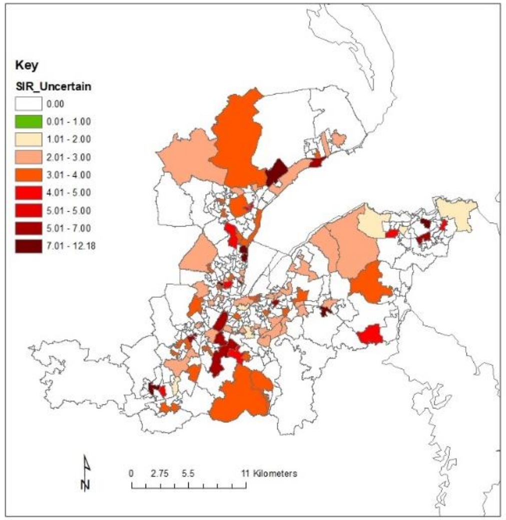
CKD with Uncertain aetiology shows SIRs up to 12 times higher than expected for NI's average incidence rates

Environ Geochem Health
https://doi.org/10.1007/s10653-020-00618-y

>65 yrs



Uncertain aetiology (CKDu)

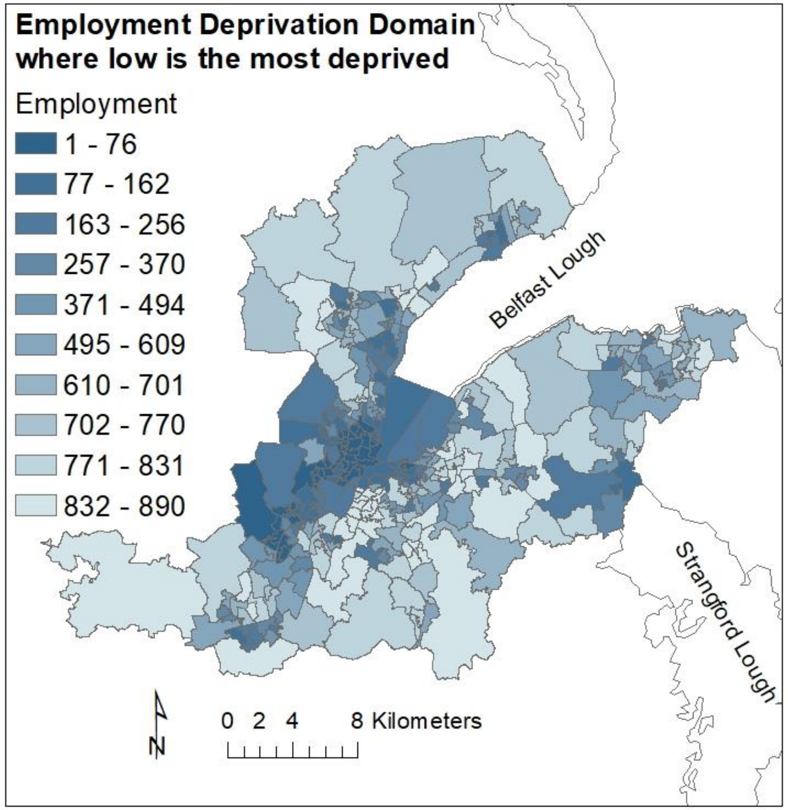


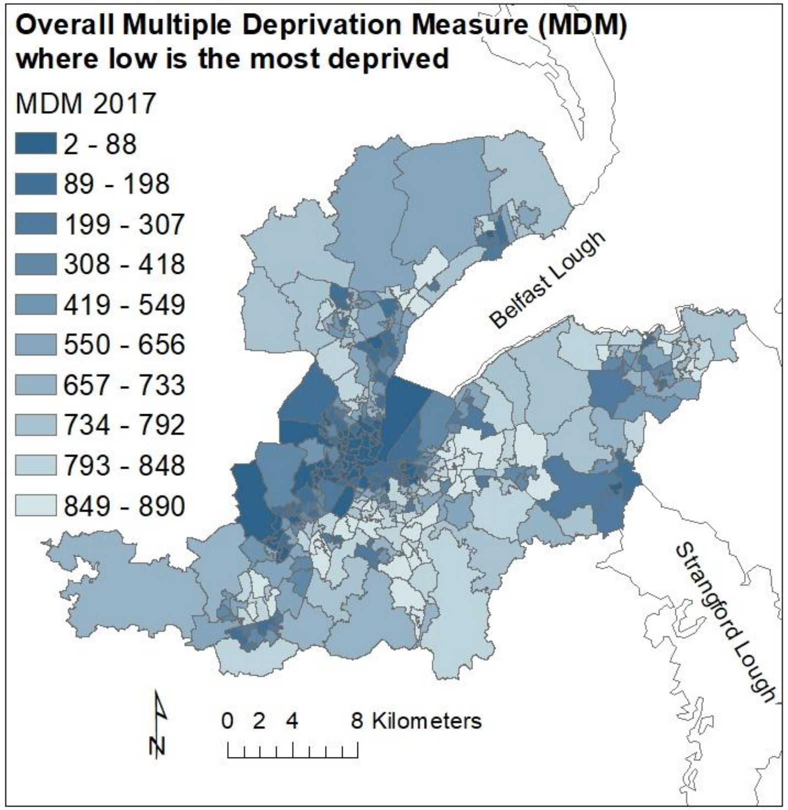
SIRs

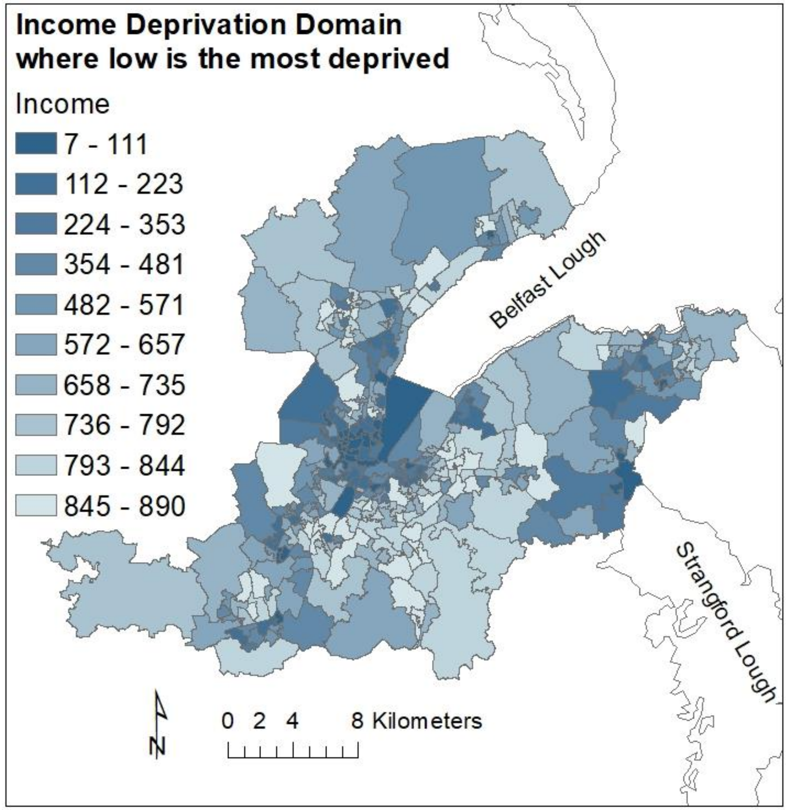
- 0.00 No CKD recorded for SOA
- 0.01 - 1.00
- 1.01 - 2.00
- 2.01 - 3.00
- 3.01 - 4.00
- 4.01 - 5.00
- 5.01 - 7.00
- 7.01 - 9.00
- 9.01 - 12.00

SIRs below 1 - the incidence is lower than expected

SIRs above 1 -the incidence is higher than expected







Dataset 2: Mapping Deprivation Measures

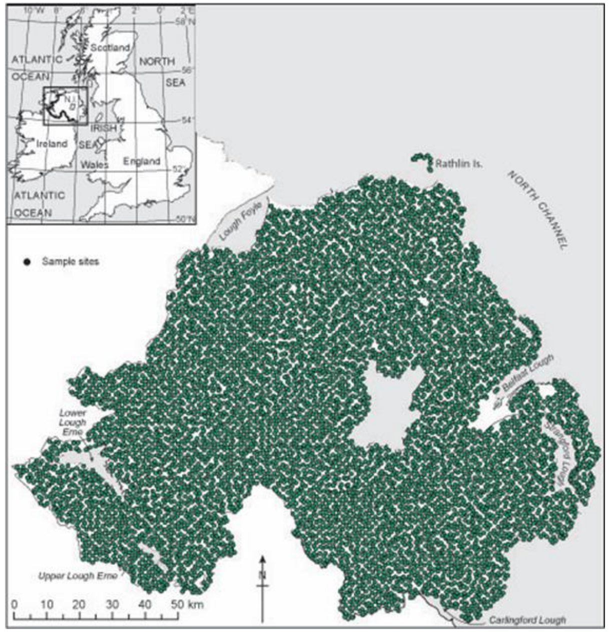
[Northern Ireland Multiple Deprivation Measure 2017 (NIMDM2017) | Northern Ireland Statistics and Research Agency (nisra.gov.uk)]

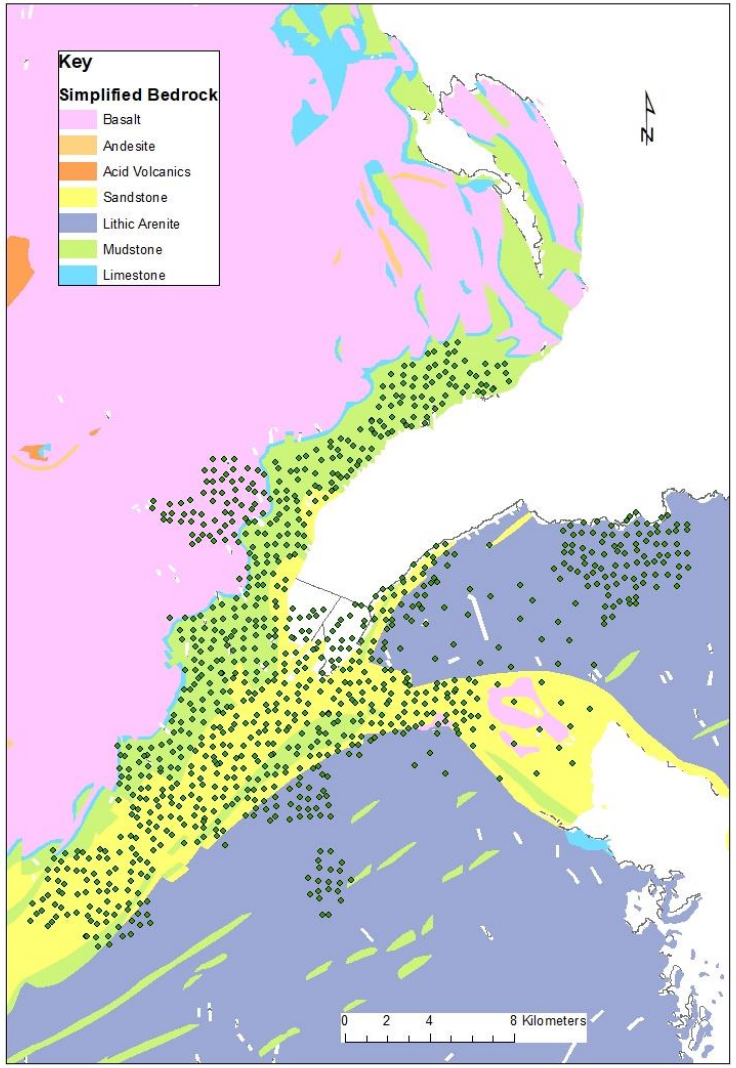
Dataset 3: Tellus Urban soil samples

QUEEN'S UNIVERSITY BELFAST

A guide to the Tellus data

Young, M. E. and Donald, A. W. (eds). 2013 Geological Survey of Northern Ireland (GSNI), Belfast. http://nora.nerc.ac.uk/509171/.





Tellus soil samples: 1000 urban sample points with XRF elemental analysis

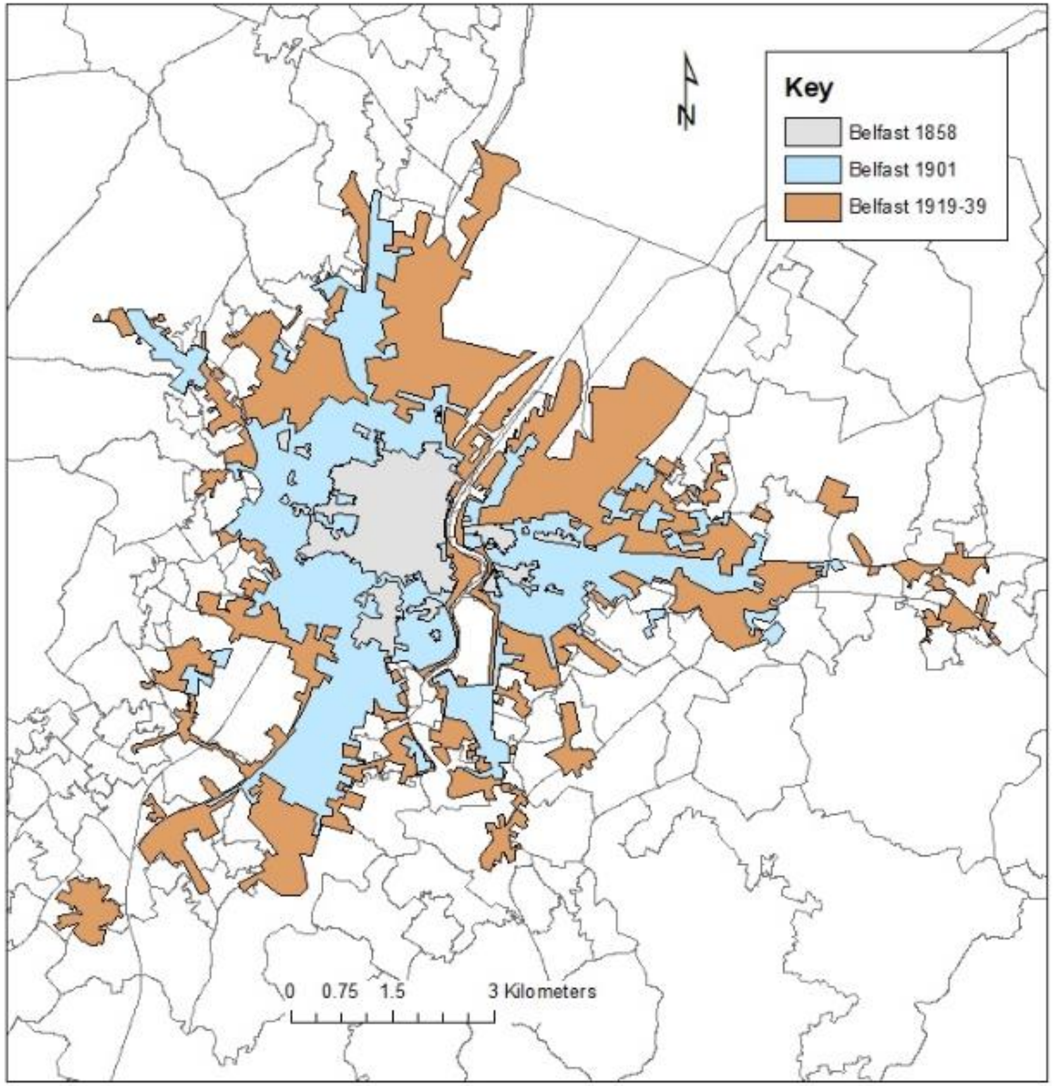
Natural sources for Potentially Toxic Elements (PTEs)

- Palaeogene basalts are potential source of cobalt (Co), vanadium (V), chromium (Cr) and nickel (Ni).
- Silurian greywacke and shales (lithic arenites) show elevated levels of arsenic (As) and molybdenum (Mo)

Mapping Urban Growth

Soils show evidence of legacy and modern day pollution

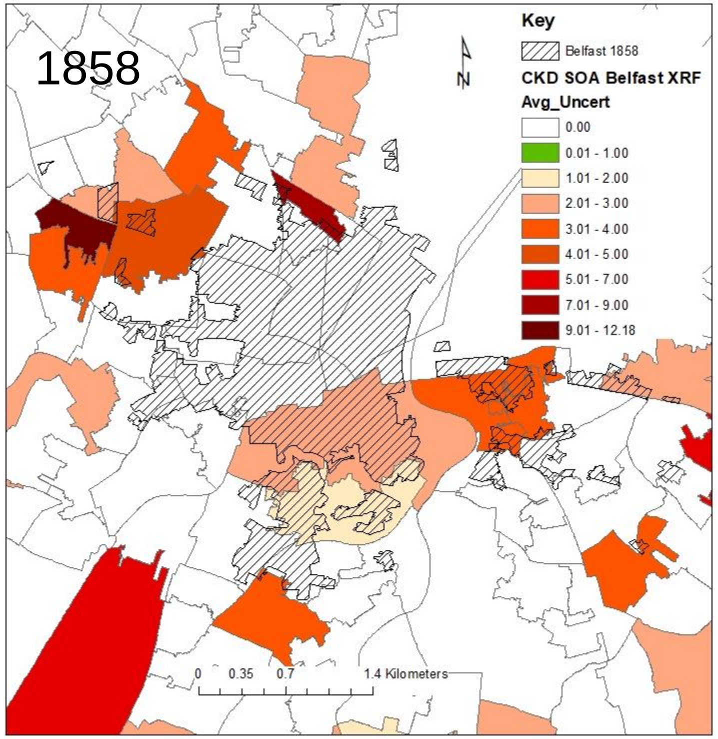
https://flashbak.com/belfast-1955

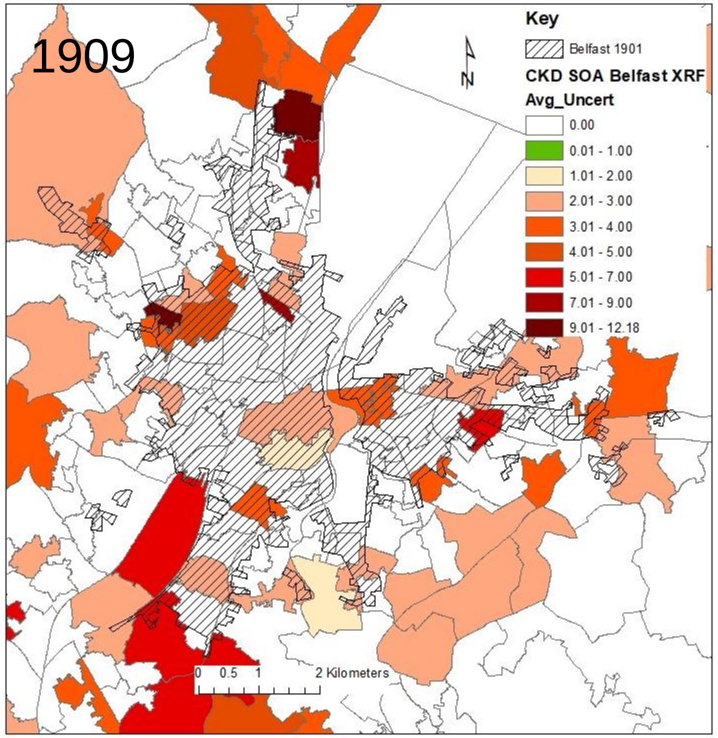


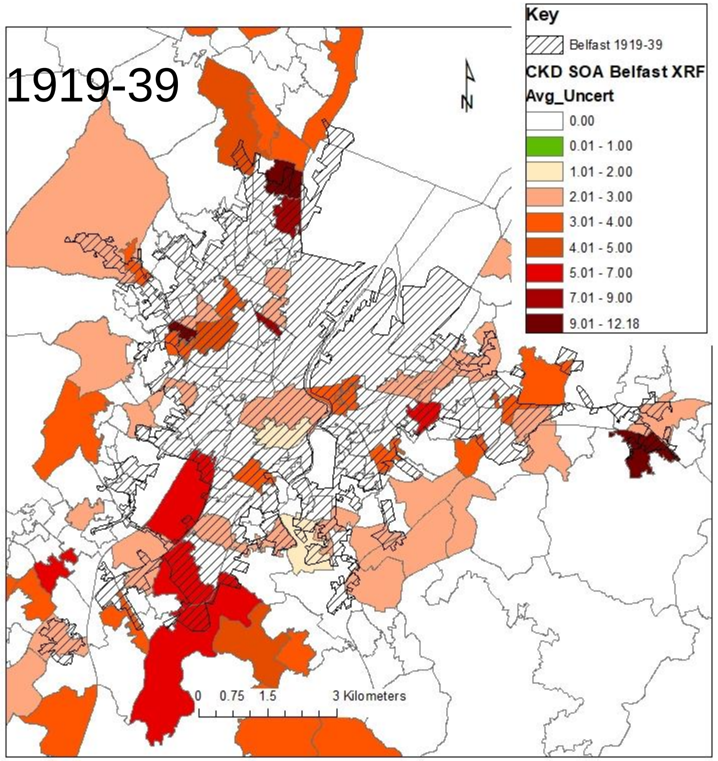
https://flashbak.com/belfast-1955

Anthropogenic sources for copper, (Cu), zinc (Zn), tin (Sn), antimony (Sb) and lead (Pb)

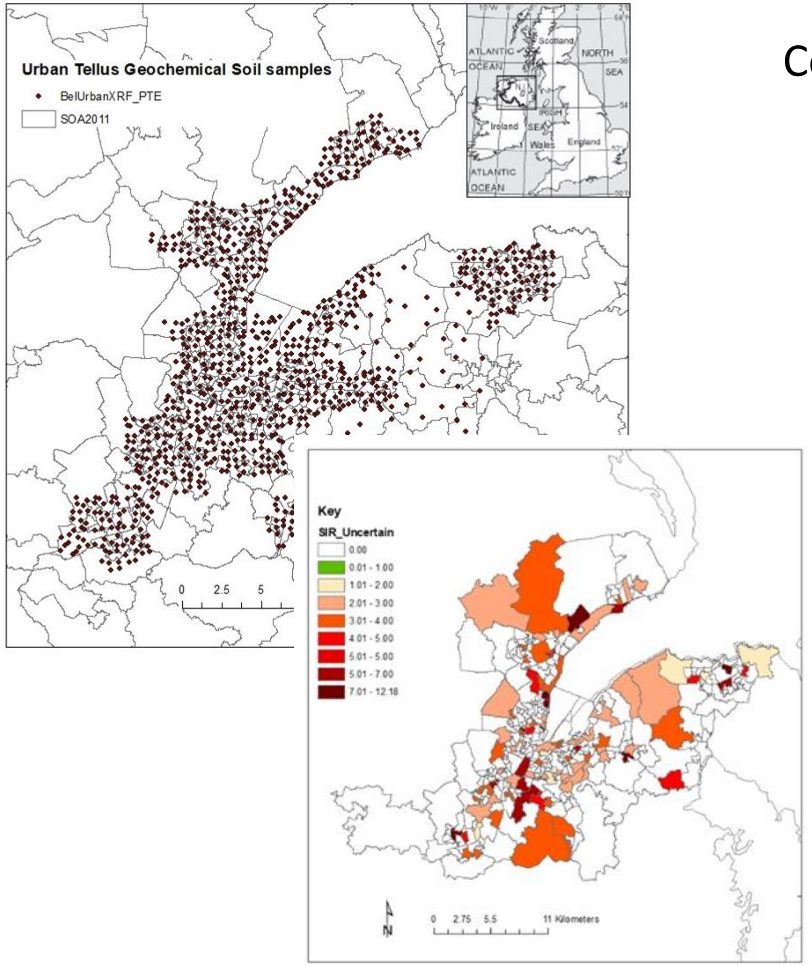
Mapping Urbanisation and CKDu



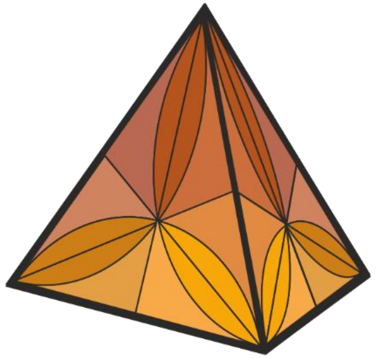




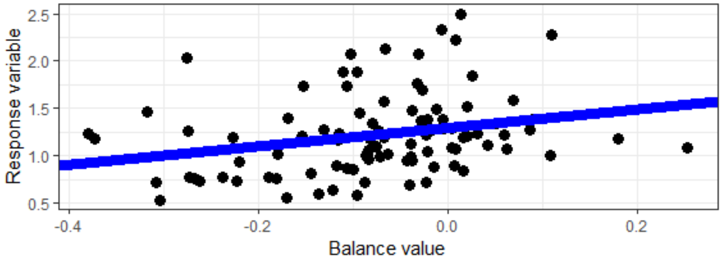
How do we model and test the relationship between different types of data ?



Constrained or closed relative data → Opened data



The Simplex



CoDA

Isometric log-ratio
Pairwise log-ratio
Balance approach

Coordinate space

Generalised linear regression
Spatial regression
Quantile regression
Tweedie model[1]
(allows for effects of zero inflation)

[1]Tweedie, 1884

CoDA Balance Approach

- A balance in CoDA using the R *selbal* algorithm (Rivera-Pinto et al. 2018) was used to identify components (MDMs and geochemical PTEs) whose relative abundance is associated with elevated incidences of CKD and CKDu (unknown aetiology) .
- An n-fold cross-validation (CV) procedure identifies the the "best" balance based on the coefficient of determination.
- Mean squared error (MSE) is used to determine the number of components included in the balance.

- The balance identified with the whole dataset is the most frequently identified in CV procedure.
- An associated regression model was used to calculate the mean response for the test set based on the balances identified in the parameter estimation step.

	%	Global	BAL 1	BAL 2	BAL 3
Avg_Ni	55				
Avg_Mo	45				
Avg_As_	45				
Avg_Cr	45				
Avg_Co	5				
FREQ	-	-	0.45	0.41	0.05

Results shown for Belfast (92 SOAs) SIRs of CKDu with soil PTEs and six individual domains of deprivation MDMs (comprising income, employment, health deprivation and disability, education, skills and training, access to services, and living environment)

There are three most common balances (with frequencies shown in the last row)

McKinley et al. 2020 Environ Geochem Health
https://doi.org/10.1007/s10653-020-00618-y

McKinley et al. 2020 Applied Computing and Geosciences.
https://doi.org/10.1016/j.acags.2020.100024

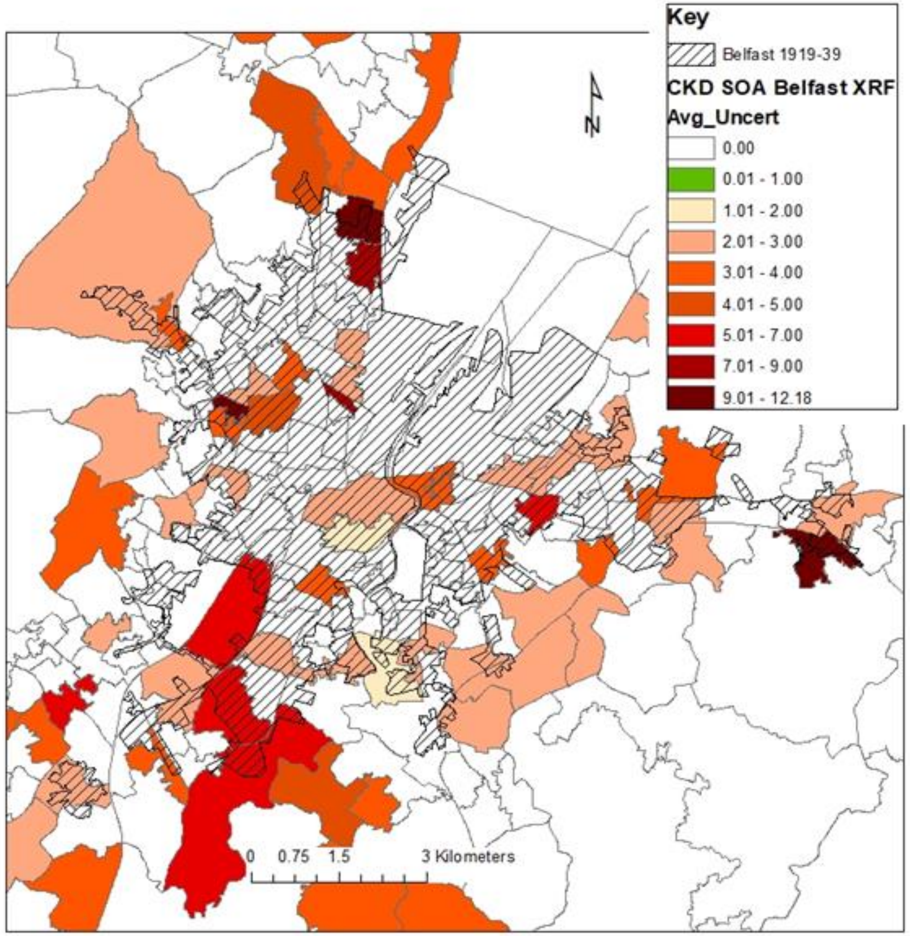
What do the results show? Urbanisation and our Health

QUEEN'S UNIVERSITY BELFAST

- The regression results suggest a correlation between **all ages of CKD incidences** (SIRs>16) and the **Multiple Deprivation Measure (MDM) domains of employment and income**[1]
- These MDMs have been used as **an indication of socio-economic factors such as smoking**[2]

- In **historical industrial Belfast (1919-39)** the strongest correlation for **CKDu** is found with an elemental balance of **copper (Cu) and antimony (Sb)**[3]
- Cu and Sb are linked to industrial areas (smelting or alloying with silver, lead and tin).

Historical Belfast 1919-39



[1]Employment and income 99% and 95% significance levels respectively
[2]Layte and Whelan 2009
[3]sample size 56 p-value =0.0371

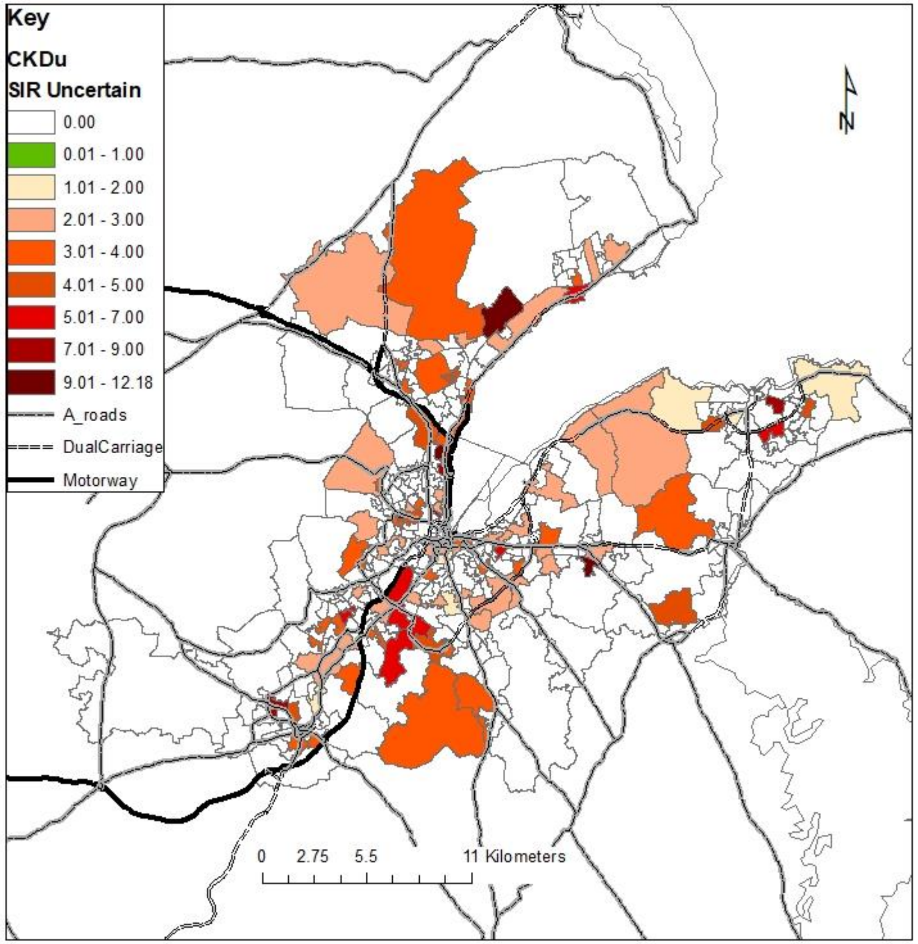
Northern Ireland Statistics and Research Agency (2017) NI Multiple Deprivation Measures 2017

What do the results show? The link with Air pollution

- For greater Belfast the strongest correlation with CKDu is found for an elemental balance of **arsenic (As) and molybdenum (Mo)**[1]

- **The transport networks** may help to shed light on the interpretation

- **Air pollution, traffic and brake wear emissions** have been cited as sources for heavy metals[2]
- **Both As and Mo** have been linked to **atmospheric pollution** deposition including **traffic pollution**[3]
- **Brake wear emissions** have been cited as a potentially important **source of Sb and Mo**[4]

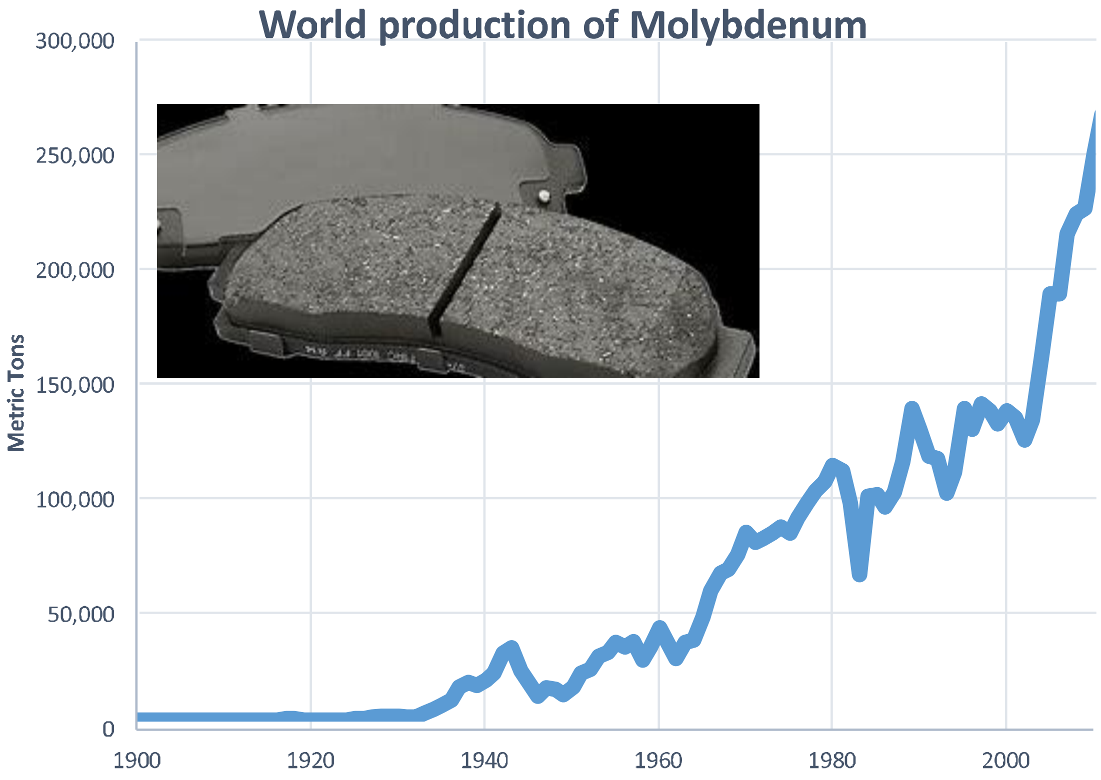


McKinley et al. 2020 Environ Geochem Health
https://doi.org/10.1007/s10653-020-00618-y

[1]sample size 340 (*p*-value =0.0391) - 95% confidence level
[2] Afsar et al. 2019
[3]Carrero et al. 2013
[4] Grigoratos & Martini 2015

What do the results mean?
Air pollution and kidney disease

- Research into **air pollution and kidney disease** is recent[1]
- Studies have shown that ultrafine particles (**including Pb, Mo and Sb**) may become **blood-borne and translocate** to other tissues such as the **liver, kidneys and brain** [1,2]
- **Soils show the evidence of air pollution deposition** and the potential impact of the **modern pollutants**
- The implications from this study are that PTEs in **urban soils** may be **used as a proxy for the availability of nephrotoxins for human intake from environmental pollution**

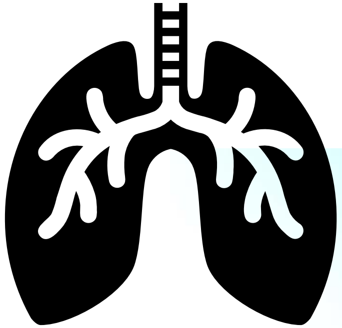


https://www.usgs.gov/centers/nmic

[1] Geiser & Kreyling 1999
[2] Oberdörster et al. 2005

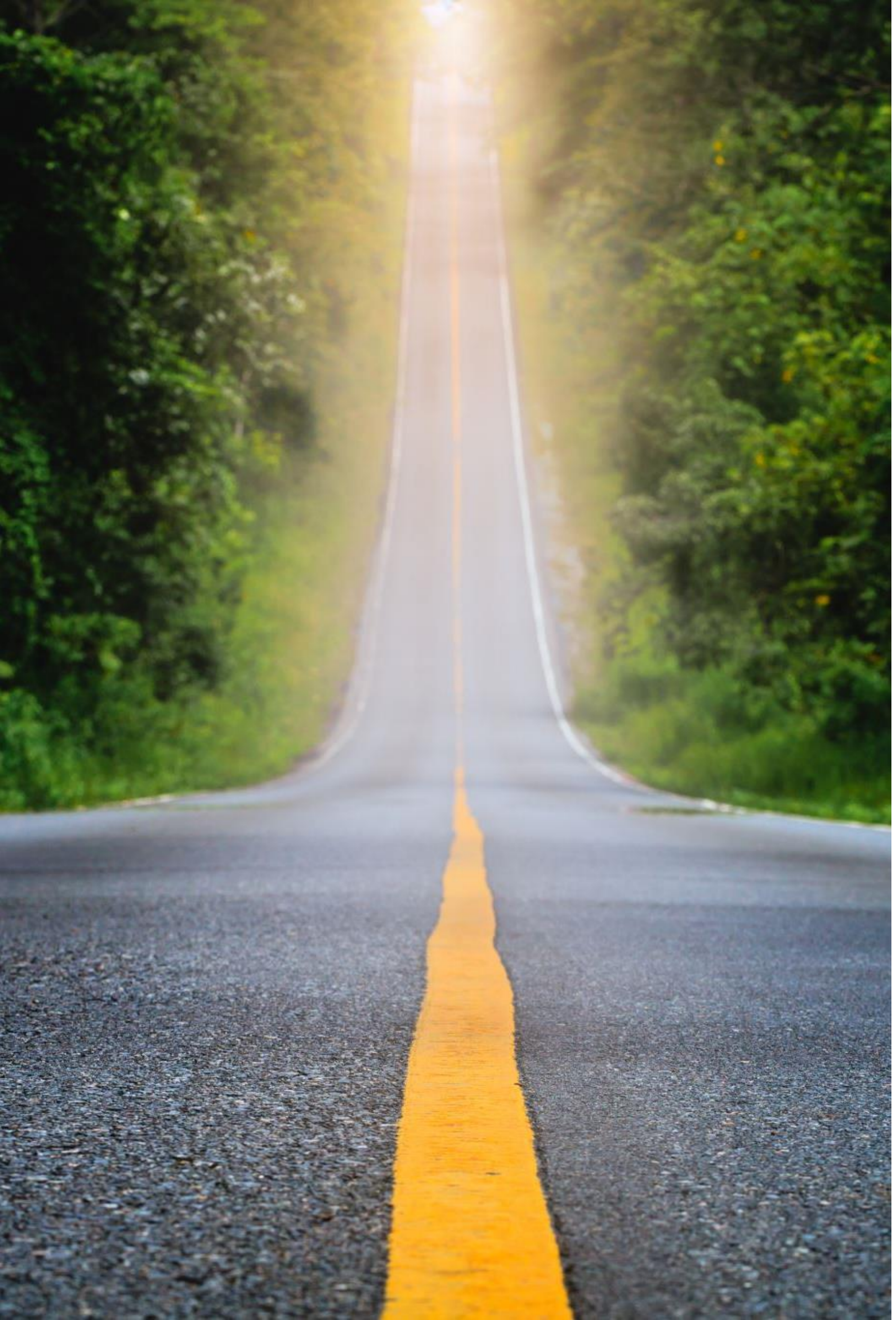
QUEEN'S UNIVERSITY BELFAST

Future work



- The preliminary findings support the argument that atmospheric pollution in the form of exposure deposition and associated toxic metals may negatively affect renal function.

- Further research is required to fully examine the impact of atmospheric pollutants and chronic kidney disease.



Future Opportunities

- What are the opportunities to develop machine learning approaches?
- How can we make sure that machine learning/AI approaches fully utilize the importance of spatial relationships?
- How to acknowledge the nature of nature in approaches?
- How can we collaborate to develop the geoscientist’s toolkit towards a more integrated, insightful interpretation?

Relevant references

- Caritat, P. de, Cooper, M., 2011a. National Geochemical Survey of Australia: The Geochemical Atlas of Australia. Geoscience Australia Record, 2011/20 (2 Volumes), 557.
- Gallagher, V., Grunsky, E. Fitzsimons, M., Browne, M., Lilburn, S., and Symons, J. (In Press) Tellus Regional Surface Water Geochemistry: environmental and mineral exploration applications. Journal of Geochemistry Exploration Environment and Analysis.
- Glennon, M., Harris, P., Ottesen, R.T., Scanlon, R.P. and O'Connor, P. (2014) The Dublin SURGE Project: Geochemical baseline for heavy metals in topsoils and spatial correlation with historic industry in Dublin, Ireland. Journal of Environmental Geochemistry and Health, 36(2), 235-254. Special Issue on the 9th International Symposium on Environmental Geochemistry.
- Lark, R.M., Ander, E.L., Cave, M.R., Knights, K.V., Glennon, M.M. and Scanlon, R.P. (2014) Mapping trace element deficiency by cokriging from regional geochemical soil data: a case study on cobalt for grazing sheep in Ireland. Geoderma, 226–227, 64–78
- McKinley, J, Deutsch, C, Neufeld, C, Patton, M, Cooper, M & Young, M 2014, 'Using geostatistical Bayesian Updating to integrate airborne radiometrics and soil geochemistry to improve mapping for mineral exploration' SAIMM, vol 114, pp. 575-586.
- McKinley, J., Tolosana Delgado, R., Hron, K., de Caritat, P., Grunsky, E., Reimann, C., Filzmoser P., and van den Boogaart K, G., 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
- McKinley, J.M, Mueller, U., Atkinson, P.M., Ofterdinger, U., Cox, S F., Doherty, R., Fogarty, D., Egozcue, J.J., Pawlowsky-Glahn, V. (2020) Chronic kidney disease of unknown origin is associated with social deprivation and environmental urbanisation in Belfast, UK., Environ Geochem Health. https://doi.org/10.1007/s10653-020-00618-y
- Pawlowsky-Glahn, V., Buccianti, A. (Eds.), 2011. Compositional Data Analysis. Theory and Applications. Wiley, Chichester.
- Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., O'Connor, P. (Eds.), 2014a. Chemistry of Europe's agricultural soils – Part A: Methodology and interpretation of the GEMAS data set. Geologisches Jahrbuch, Schweizerbarth, Hannover (Germany).
- Talebi, H., Mueller, U., Tolosana-Delgado, R., Grunsky, EC., McKinley, JM., de Caritat,. 2018. Surficial and Deep Earth Material Prediction from Geochemical Compositions. Natural Resources Research, Natural Resources Research 28 (3), 869-891Grunsky, E.C., 2010. The interpretation of geochemical survey data. Geochem. Explor. Environ. Anal. 10 (1), 27–74.

QUEEN'S UNIVERSITY BELFAST

Post Tellus Research Publications

P. E. ANDERSON et al. 2017. Refined model of incremental emplacement based on structural evidence from the granodioritic Newry igneous complex, Northern Ireland. **Geological Society of America Bulletin.**

P. E. ANDERSON et al. 2016. Zonation of the Newry Igneous Complex, Northern Ireland, based on geochemical and geophysical data. **Lithos.**

HOLLIS, S.P. et al. 2016. 'Using Tellus data to enhance targeting of volcanogenic massive sulphide mineralisation in the Tyrone Igneous Complex' in M.E. Young (ed.), **Unearthed.**

DEMPSTER, M. et al. 2016. 'Using soil geochemistry to investigate gold and base metal distribution and dispersal in the glaciated north of Ireland' in M. E. Young (ed.), **Unearthed.**

COOPER, M. R. et al. 2016. 'Shape and intrusion history of the Late Caledonian, Newry Igneous Complex, Northern Ireland' in M. E. Young (ed.), **Unearthed.**

Anderson, H. et al. 2016. 'Faults, intrusions and flood basalts: the Cenozoic structure of the north of Ireland' in M. E. Young (ed.), **Unearthed.**

HOLLIS, S.P. et al. 2015. Distribution, mineralogy and geochemistry of silica-iron exhalites and related rocks from the Tyrone Igneous Complex: implications for VMS mineralization in Northern Ireland. **Journal of Geochemical Exploration.**

HOLLIS, S. P. et al. 2014. Petrochemistry and hydrothermal alteration within the Tyrone Complex, Northern Ireland: implications for VMS mineralization in the British and Irish Caledonides **Mineralium Deposita.**

HOLLIS, S. P. et al. 2013. Stratigraphic, geochemical and U–Pb zircon constraints from Slieve Gallion, Northern Ireland: a correlation of the Irish Caledonian arcs. **Journal of the Geological Society, London.**

HOLLIS, S. P. et al. 2013. Evolution of the Tyrone ophiolite, Northern Ireland, during the Grampian–Taconic orogeny: a correlative of the Annieopsquotch Ophiolite Belt of central Newfoundland? **Journal of the Geological Society, London.**

COOPER, M. R. et al. 2013. A U-Pb age for the Late Caledonian Sperrin Mountains minor intrusions suite in the north of Ireland. **Journal of the Geological Society, London.**

DEMPSTER, M. et al. 2013. Principal Component Analysis of Geochemistry of Soils Developed on Till in Northern Ireland. **Journal of Maps.**

HOLLIS, S. P. et al. 2012. Episodic arc-ophiolite emplacement and the growth of continental margins: Late accretion in the Northern Irish sector of the Grampian-Taconic orogeny. **Geological Society of America Bulletin.**

COOPER, M. R. et al. 2012. Palaeogene Alpine tectonics and Icelandic plume-related magmatism and deformation in Ireland. **Journal of the Geological Society, London.**

COOPER, M. R et al. 2011. Age constraints and geochemistry of the Ordovician Tyrone Igneous Complex, Northern Ireland. **Journal of the Geological Society, London.**

References

- Aitchison, J. (1986) The Statistical Analysis of Compositional Data Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416p.
- Barsby, A., McKinley, J.M., Ofterdinger, U., Young, M., Cave, M.R. & Wragg, J. 2012, 'Bioaccessibility of trace elements in soils in Northern Ireland' Science of The Total Environment, vol 433, no. null, pp. 398-417., http://dx.doi.org/10.1016/j.scitotenv.2012.05.099
- Carson et al., Toxicology and Biological Monitoring of Metals in Humans, Lewis Pub. p. 21, 1987.
- Chayes, F. (1960). On correlation between variables of constant sum. Journal of Geophysical Research 65~(12), 4185–4193.
- Geiser M., & Kreyling W.G., 1999. Deposition and biokinetics of inhaled nanoparticles. Part Fiber Toxicol 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, Southhampton: Karger AG, Basel.
- Grigoratos, T. & Martini, G., 2015. Brake wear particle emissions: a review. Environ Sci Pollut Res (2015) 22:2491–2504 DOI 10.1007/s11356-014-3696-8
- Lewis, R., 2012. Understanding Chronic Kidney Disease: A guide for the non-specialist. England: M&K Update Ltd.
- McIlwaine, R., Doherty, R., Cox S. & Cox, M., 2017. The relationship between historial development and potentially toxic element concentrations in urban soils. Environmental Pollution Volume 220, Part B, January 2017, Pages 1036-1049 https://doi.org/10.1016/j.envpol.2016.11.040
- McKinley, J.M., Ofterdinger, U., Young, M., Barsby, A. & Gavin, A. 2013, 'Investigating local relationships between trace elements in soils and cancer data' Spatial Statistics, vol 5, pp. 25-41., http://dx.doi.org/10.1016/j.spasta.2013.05.001
- McKinley, J.M, Mueller, U., Atkinson, P.M., Ofterdinger, U., Cox, S F., Doherty, R., Fogarty, D., Egozcue, J.J., Pawlowsky-Glahn, V. (2020) Chronic kidney disease of unknown origin is associated with social deprivation and environmental urbanisation in Belfast, UK., Environ Geochem Health (2020). https://doi.org/10.1007/s10653-020-00618-y
- McKinley, J.M, Mueller, U., Atkinson, P.M., 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
- Oberdörster G, Oberdörster E, & Oberdörster J., 2005. Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. Environ Health Perspect 113:823–839
- Pawlowsky-Glahn, V. & Egozcue, J.J., 2001. Geometric approach to statistical analysis on the simplex. SERRA 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
- Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference (Eds. J. K. Ghosh and J. Roy), pp. 579–604. Calcutta: Indian Statistical Institute.
- 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. 2015; 16: 145. doi: 10.1186/s12882-015-0105-6 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4539684/
- Young, M. E. and Donald, A. W. (eds) 2013. A guide to the Tellus data. Geological Survey of Northern Ireland (GSNI), Belfast. The publication 'A Guide to the Tellus Data' can be downloaded from http://nora.nerc.ac.uk/509171/ .

QUEEN'S UNIVERSITY BELFAST

Tellus Data

- Department of Enterprise, Trade and Investment (N I)
- ‘Building Sustainable Prosperity’ fund of the EU Regional Development Programme

UKRR data

- The study has received ethics approval March 2018, NHS National Research Ethics Committee REC reference: 15/EM/0366.

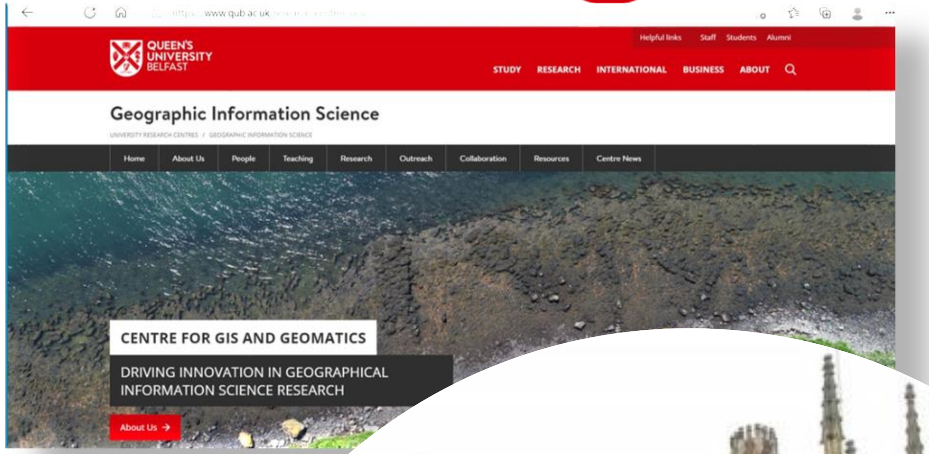
UKRR data

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.

QUEEN'S UNIVERSITY BELFAST

Thank you for listening



Prof Jennifer McKinley, C.Geol., FGS

Director of Centre for GIS and Geomatics
Geography
School of Natural and Built Environment,
Queen's University Belfast