



**QUEEN'S  
UNIVERSITY  
BELFAST**

## Investigating local relationships between trace elements in soils and cancer data

McKinley, J. M., Offerdinger, U., Young, M., Barsby, A., & Gavin, A. (2013). Investigating local relationships between trace elements in soils and cancer data. *Spatial Statistics*, 5, 25-41.  
<https://doi.org/10.1016/j.spasta.2013.05.001>

**Published in:**  
Spatial Statistics

**Document Version:**  
Publisher's PDF, also known as Version of record

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

### **Publisher rights**

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License <http://creativecommons.org/licenses/by/3.0/>, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited

### **General rights**

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

### **Open Access**

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>



ELSEVIER

Contents lists available at ScienceDirect

## Spatial Statistics

journal homepage: [www.elsevier.com/locate/spasta](http://www.elsevier.com/locate/spasta)

# Investigating local relationships between trace elements in soils and cancer data<sup>☆</sup>



Jennifer M. McKinley<sup>a,\*</sup>, Ulrich Ofterdinger<sup>b</sup>,  
Michael Young<sup>c</sup>, Amy Barsby<sup>a,1</sup>, Anna Gavin<sup>d</sup>

<sup>a</sup> School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, BT7 1NN, UK

<sup>b</sup> School of Planning, Architecture and Civil Engineering, Queen's University Belfast, Belfast, BT9 5AG, UK

<sup>c</sup> Geological Survey Northern Ireland (GSNI), Belfast, BT9 5BF, UK

<sup>d</sup> N. Ireland Cancer Registry, Centre for Public Health, School of Medicine, Dentistry & Biomedical Sciences, Queen's University Belfast, Mulhouse Building, Grosvenor Road, Belfast, BT12 6BJ, UK

## ARTICLE INFO

### Article history:

Received 30 December 2012

Accepted 21 May 2013

Available online 31 May 2013

### Keywords:

Cancer disease data

Soil geochemistry

Radon

Geographically weighted regression

Geostatistics

Arsenic

## ABSTRACT

Medical geology research has recognised a number of potentially toxic elements (PTEs), such as arsenic, cobalt, chromium, copper, nickel, lead, vanadium, uranium and zinc, known to influence human disease by their respective deficiency or toxicity. As the impact of infectious diseases has decreased and the population ages, so cancer has become the most common cause of death in developed countries including Northern Ireland. This research explores the relationship between environmental exposure to potentially toxic elements in soil and cancer disease data across Northern Ireland. The incidences of twelve different cancer types (lung, stomach, leukaemia, oesophagus, colorectal, bladder, kidney, breast, mesothelioma, melanoma and non melanoma (NM) skin cancer both basal and squamous) were examined in the form of twenty-five coded datasets comprising aggregates over the 12 year period from 1993 to 2006. A local modelling technique, geographically weighted regression (GWR) is used to explore the relationship between environmental exposure and cancer disease data. The results show comparisons of the geographical incidence of certain cancers (stomach and NM squamous skin cancer) in relation to concentrations of certain PTEs (arsenic levels in soils

<sup>☆</sup> This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

\* Corresponding author. Tel.: +44 2890973827.

E-mail address: [j.mckinley@qub.ac.uk](mailto:j.mckinley@qub.ac.uk) (J.M. McKinley).

<sup>1</sup> Posthumously.

and radon were identified). Findings from the research have implications for regional human health risk assessments.

© 2013 The authors. Published by Elsevier B.V. All rights reserved.

---

## 1. Introduction

### 1.1. How environmental factors affect health

Natural trace elements, mineral water and gases (such as radon) are present in the environment and these interact with the human body in both positive and negative ways. As recognised by Paracelsus (1493–1541 AC) “all substances are poisons; there is none which is not a poison; the right dose differentiates a poison and a remedy”. Medical geology or spatial epidemiology is concerned with the study of spatial patterns of disease incidence and mortality and the identification of potential causes of disease including environmental exposure or socio-demographic factors (Goovaerts, 2010). To date, the culmination of a broad body of research has recognised a number of potentially toxic elements (PTEs), such as arsenic (As), cobalt (Co), chromium (Cr), copper (Cu), nickel (Ni), lead (Pb), selenium (Se), vanadium (V), uranium (U) and zinc (Zn), known to influence human disease by their respective deficiency or toxicity. As the impact of infectious diseases has decreased and the population as a whole ages, so cancer has become the most common cause of death in developed countries. The risk of developing cancer is recognised as a combination of the person’s genetic makeup and environmental factors usually over long periods of time. Steingraber (2010) describes a study of cancer among adoptees that found correlations with their adoptive families but not within their biological ones. The concept that our genes work in communion with substances from the larger ecological world suggests that what runs in families does not necessarily run in the blood (Steingraber, 2010). Carcinogens fall into three groups—chemical, physical and biological. Chemical carcinogens, the largest group, include tobacco products, asbestos, benzene and the products of tobacco. Biological agents include infections such as Human Papilloma Virus, (HPV) causally linked with cervical cancer, and Human Immunodeficiency virus (HIV) linked with lymphomas. The best known example of physical carcinogens is high-energy radiation, including nuclear radiation and X-rays. Radiation is known as a ‘complete’ carcinogen because it can initiate, promote and progress a cancer. Chemical carcinogens occur in nature, in mineral ores, such as arsenic and others in foods (e.g. fungal contaminants). The history of cancer is long but our recognition of the agents that contribute to its occurrence has been slow to mature. A reflection that external or environmental agents could produce malignant change was noted by Pott, a London physician, in 1775, after observational studies prompted him to link scrotal cancer, common among chimney sweeps, to the soot that accumulated on their bodies (cited in Majno and Joris, 2004). Skin cancer was noted to be prevalent among workers exposed to arsenic fumes in copper smelters and tin foundries in Cornwall and Wales. Workers in cobalt mines in Saxony and the uranium mines in Bohemia were subject to a disease of the lungs later identified as cancer. Many of the causes of cancer including the effects of lifestyle and environmental factors are still not well understood. Investigating the geographical differences in cancer incidence may shed light on variations in cancer risk factors between populations (Carsin et al., 2009).

### 1.2. Aim of the work

The term ‘environmental medicine’ has been used for studies into how environmental factors affect health (Alexander and Boyle, 1996). While it is estimated that occupational and environmental exposures account for 2% and less than 1% of total cancers respectively, it has also been assessed that 80% of all cancers are environmentally induced when cancer from dietary carcinogens, polluted air and water along with exposure to cosmic and solar radiation is included. Environmental exposure rates concentrate in subgroups of the population (Bofetta, 2004), deeming them worthy of investigation despite the small percentages. The aim of this research is to explore spatial correlations between PTEs in soils and epidemiological data using a comprehensive soil geochemical dataset for Northern

Ireland generated from the Tellus Survey (GSNI) in conjunction with cancer data provided by the Northern Ireland Cancer Registry (NICR). Bioaccessibility is the fraction of the pseudo-total soilborne PTE concentration that is soluble in the gastrointestinal tract and is available for absorption (Wragg et al., 2009). A conjunctive study by Barsby et al. (2012) undertook bioaccessibility testing of selected soil samples from the Tellus Survey following the Unified BARGE Method (BARGE INERIS, 2011) and investigated the spatial variability of pseudo-total and bioaccessible PTE concentrations. A secondary aim of this study is to compare observed lithological and pedological associations from the Barsby et al. (2012) study with spatial correlations between PTEs in soils and epidemiological data.

## 2. Material and methods

There are several factors that culminate to make Northern Ireland an important exemplar for the UK and Ireland and indeed for advancing the science of medical geology globally. The key opportunities stem firstly from Northern Ireland's complex geology (Fig. 1a) that forms a microcosm for that encountered across the UK and Ireland and secondly from the combination of available data sources: comprehensive soil geochemistry data generated by GSNI Tellus Survey (conducted over the period 2004–2006) and cancer data collected from 1993 onwards and maintained by NICR. In bringing these two datasets together, this research offers the potential to examine the impact of one aspect of the environment on public health. The accuracy and inclusivity of the datasets enables spatial investigation into the influence of geographical distribution, total concentration and level of bioaccessibility of trace element in soil on human epidemiology in Northern Ireland. The indications of spatial correlations found have implications outside Northern Ireland for regional human health risk assessments.

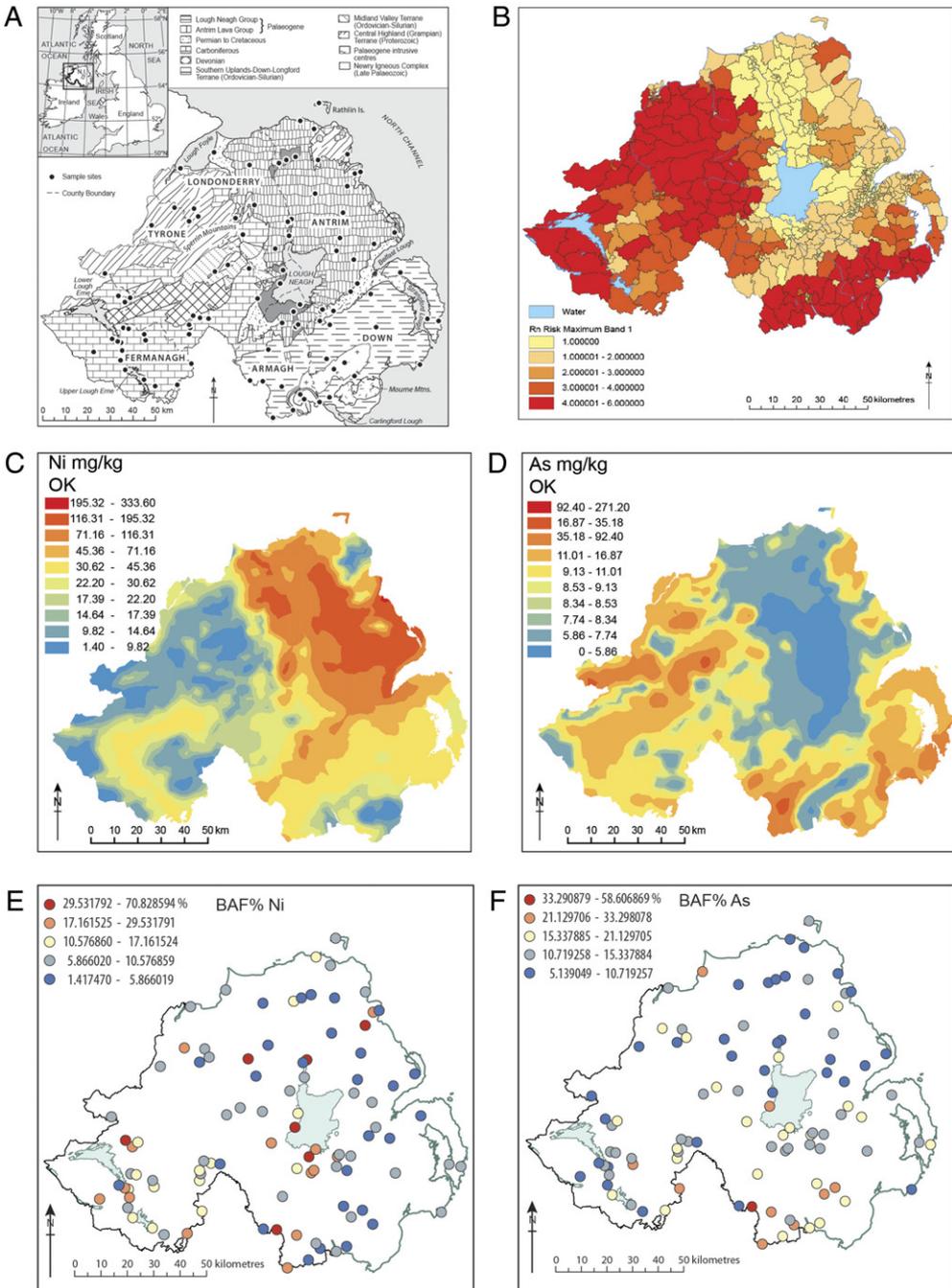
### 2.1. Soil geochemistry

The GSNI Tellus Survey, completed between 2004 and 2006, provides a unique dataset combining comprehensive spatial soil sampling coverage with an extensive suite of soil geochemical analysis (Smyth, 2007). This study uses data from the ground based geochemical survey comprising 13,860 soil samples taken at a 20 cm depth and collected on a regular grid of one sample site every 2 km<sup>2</sup> (GSNI, 2007) following the G-BASE sampling regime established by British Geological Survey (BGS). The soil samples used for this study were analysed for 60 elements and inorganic compounds using pressed pellet X-ray Fluorescent Spectrometry (XRF) using Wavelength Dispersive XRF Spectrometry (WD-XRF) and Energy Dispersive/Polarised XRF Spectrometry (ED-XRF). The sampling and analysis regimes for the geochemical surveys included in the Tellus Survey are detailed in Smyth (2007). The Tellus data set provides the basis for a comprehensive study with relevancy for the whole of the UK.

### 2.2. Cancer data

Cancer is now the most common cause of death in Ireland as a whole (CSO, 2011) and moreover, overall cancer incidence is expected to increase by 20% between 2010 and 2020, and by 30% between 2010 and 2030 (NCR, 2011) mainly due to population ageing. Cancer is predominantly a disease of the elderly with cancers in children accounting for around 0.5% of the total cancer incidence (Quinn et al., 2005, 8). Cancer mortality is also projected to increase, although not to the same extent (Carsin et al., 2009, 11). In Northern Ireland, between 1993 and 2003, 8,640 new cancer cases were diagnosed every year. The 2001 census for Northern Ireland records the population as 1,685,267 (NISRA, 2004), this equates to 0.51% of the population being diagnosed with cancer annually. The overall cancer mortality rates in Ireland are slightly below the European average in men and above average in women (Boyle et al., 2003).

Disease data for this research were provided by the NICR. This registry was re-established in May 1994 and replaced an older, incomplete paper based registry established in 1959. The registry receives and collects electronic data information on all neoplasms diagnosed in Northern Ireland including non melanoma (NM) skin cancers and has an extensive programme of quality assurance of the registry data



**Fig. 1.** (A) Simplified geology; (B) Maximum radon risk; (C) OK map of total Ni (mg/kg) in soils; (D) OK map of total As (mg/kg) in soils; (E) BAF % Ni in soils; (F) BAF % As in soils.

Source: Adapted from Barsby et al. (2012).

(NICR, 2008). The information contained in the database is a uniquely valuable resource and makes Northern Ireland a prime location in which to carry out this research study.

Spatial variation observed in the incidences of different cancer diseases does not necessarily mean that the spatial location itself causes cancer, rather as Carsin et al. (2009) stress it is more likely to reflect socio-economic differences and lifestyle factors in the population, geographical differences in exposure to risk factors and variations in access to, or uptake, of cancer services. The influence of population density must also be taken into account. Areas with a population of >20 persons/hectare consistently were found to record a higher risk of cancer than sparsely populated areas with <1 person/hectare (Carsin et al., 2009). This does not negate the importance of the geographical aspect of the land on population settlement in that access to resources may influence where populations initially locate. A further issue is population mobility, from the foetal development stage until diagnosis it is entirely possible for a human to come into contact with a whole range of possible carcinogens from a variety of sources. However, the proposal of exploring the relationship between soil constituents with cancer disease is arguably well suited to Northern Ireland, as an area with traditionally little migration of people thus providing greater potential for prolonged exposure at one location or area. Furthermore as the risk of developing cancer increases with age, an older population (> 60 years) is likely to have less exposure to imported and processed foods and beverages.

### 3. Theory/calculations

#### 3.1. *Aspatial and spatial analyses of geochemical data*

Following exploratory aspatial analysis (Table 1) of pseudo-total PTE concentrations, including As, Co, Cd, Cr, Ni, Pb, Se, V, U and Zn, spatial mapping techniques were used to identify areas of elevated levels of the PTEs. It is rare that environmental data are normally distributed. Although the application of geostatistics is more straightforward for distributions that are not too skewed, it does not require the variable to follow a normal distribution except in specific situations, such as multiGaussian kriging and sequential Gaussian simulation. Moreover, undertaking a normalisation process may force the data to follow fictitious constraints, which in the case of geochemistry may have little in common with real geological processes. For this reason normalisation was not applied to the geochemical data for this study. Gstat (Pebesma and Wesseling, 1998; Pebesma, 2004), an open source programme for multivariate geostatistical modelling, was used to estimate variograms for the PTEs. Mathematical models, selected from the set of authorised models provided by McBratney and Webster (1986), were fitted to the experimental variograms using the weighted least squares functionality of Gstat. Parameters of the fitted models (Table 2) were used to provide information on the maximum scale of spatial variation of the PTEs (Jensen et al., 1996; Gringarten and Deutsch, 2001; McKinley et al., 2004). The nugget:sill ratio was computed to determine the proportion of random to spatially structured variation observed for the geochemical elements. The coefficients of the modelled variograms using Gstat were used for ordinary kriging (OK), conducted using the geostatistical functionality of ArcGIS version 10 for all PTEs of interest.

#### 3.2. *Standardisation of disease data*

The incidence of twelve different cancer types (lung, stomach, leukaemia, oesophagus, colorectal, bladder, kidney, breast, mesothelioma, melanoma and non melanoma basal and squamous), were provided by the NICR in the form of twenty-five anonymised coded datasets comprising aggregates over the 12 year period from 1993 to 2006. The aggregated datasets were provided in the form of postcode, gender and disease code. In addition, NICR provided a non-cancer disease dataset and a fabricated dataset to test the effectiveness of the methodology. Initially a blind analysis was conducted on the disease coded data. The 2001 Census provided population age and gender structure on a census ward basis. These data were not available on a postcode scale, so the disease events were assigned by postcode (the lowest level unit of geographical unit) to the corresponding census ward code using a 'point in polygon' method which matched the relevant latitude and longitude information for postcodes from the Central Postcode Directory NISRA (2004) against digital boundaries supplied by the Ordnance Survey of Northern Ireland. A function was applied to the data to account for population density and age distribution variations across the area of investigation. Methods for age standardisation are outlined in Bland (1996, 291). Age standardisation takes the frequency of disease by age

**Table 1**

Summary statistics for pseudo-total concentrations of individual potentially toxic elements (PTEs) in mg/kg as measured by X-ray Fluorescent Spectrometry (XRF).

mg/kg	Arsenic (As)	Cadmium (Cd)	Chromium (Cr)	Nickel (Ni)	Lead (Pb)	Selenium (Se)	Uranium (U)
Method detection limit (MDL)	0.9	0.5	3.0	1.4	1.3	0.2	0.5
Minimum	0.9	0.5 <sup>a</sup>	4.1	1.4	2.2	0.2	0.5
Maximum	271.2	63.3	1229	333.6	18,756	7.8	142.9
Median	8.7	0.5	94.1	29.1	28.8	0.7	2.3
Interquartile range	6.5–11.7	< MDL–0.7	56.7–161	14.0–56.1	22.2–41.0	0.6–1.0	1.7–2.8

<sup>a</sup> 57% samples below MDL.

**Table 2**

Coefficients of the modelled variograms.

Element	Modelled parameters							
	$c_0$	$c_1$	$c_2$	a1 (m)	a2 (m)	$c_0 + c_1 + c_2$	$c_0:c_0 + c_1 + c_2$	%
As	64.17	19.40	14.17	7436	42311.9	83.56	0.7679	76.79
Cd	0.42	0.17		10000		0.59	0.7124	71.24
Co	54.61	152.78		73603.7		207.39	0.2633	26.33
Cr	3773.63	1590.93	11760.50	10370.4	89597.6	17125.06	0.2204	22.04
Cu	1001.33	690.74		60425.1		1692.07	0.5918	59.18
Ni	741.33	1220.30	3284.49	50817.8	73913.7	5246.12	0.1413	14.13
Pb	55056.60	12324.10		17736		67380.70	0.8171	81.71
Se	0.11	0.06	0.06	10005.6	39528.8	0.22	0.4965	49.65
V	1250.32	3644.78		81864.3		4895.10	0.2554	25.54
Zn	1521.50	2078.69		95924.3		3600.19	0.4226	42.26
Zr	2659.82	1752.3	2059.53	9175.9	50526.5	6471.65	0.4110	41.10

(and gender) group in each administrative unit (in this case ward, the population age distribution in each unit, and the standard population. For this study Northern Ireland 2001 census data were used for 582 electoral districts or wards. Following the methodology of Bland (1996) the disease frequency was divided by the category specific population to give the age specific rate. This was multiplied by the proportion of the age group in the standard population, to give the expected frequency if the population exhibited the same age profile as the standard population. The standardised rate therefore is the sum of the expected frequencies divided by the sum of the standard population. This gave the Age Standardised Incidence Rate (ASIR's) of disease by gender for 582 electoral wards. The approach used in this research is described in Donnelly and Gavin (2007).

### 3.3. Mapping disease clustering

Following the age standardisation process and preliminary analysis of the mapped ASIR outputs a local cluster analysis was undertaken. Disease clustering involves the analysis of 'unusual' aggregations of disease to identify if areas of elevated incidences of disease are present (Lawson, 2006). A global spatial analysis involves the complete study area to examine whether a disease has a natural tendency to cluster. The aim of a local clustering method is to identify the locations of any clusters (Alexander and Boyle, 1996). In this study the local technique Moran's  $I$  was used to identify statistically significant spatial outliers in the disease datasets. The local Moran's  $I$  index ( $I$ ) is a relative measure and can only be interpreted within the context of significance levels. A positive value for  $I$  indicates that a geographical unit has neighbouring geographical units with similarly high or low attribute values and therefore forms part of a cluster. A negative value for  $I$  indicates that a geographical unit has neighbouring geographical units with dissimilar values; this geographical unit is an outlier. In either instance, the significance level for the geographical unit must be small enough for the cluster or outlier to be considered statistically significant. Results are only reliable if the input class contains at least 30 geographical units. This study had 582 input classes (geographical ward units). All results of the local cluster analysis shown are statistically significant for a significance level of 0.05.

### 3.4. Spatial relationships between disease data and explanatory factors

Ecological analysis involves the analysis of relations between the spatial distribution of disease incidence and measured explanatory factors (Lawson, 2006). More often the analysis is undertaken at an aggregated spatial level and compared to explanatory factors measured at regional or other levels of aggregation. This can lead to misconceptions of relations that are more related to the change of aggregation level than the aggregated data itself (Lawson, 2006). In an effort to avoid this, summary statistics of PTEs were calculated on a ward basis to compare directly with disease data.

Regression analysis may be useful to investigate factors affecting a population. Cook and Pocock (1983) investigated the relationship between cardiovascular incidence in the UK and potentially explanatory variables including water hardness, climate, pollution, location, socioeconomic and genetic factors. Brugge et al. (2007) studied the link between respiratory health of school children and volatile organic compounds in the external atmosphere. Geographically Weighted Regression (GWR) is a spatial regression technique that is increasingly used in geography and other disciplines. In this study GWR was used to provide a local model of the relation between disease data and PTEs by fitting a regression equation to every geographic unit. GWR works in a similar way to a moving window regression in that all the data points within a window are used to calibrate a model (Fotheringham et al., 2002). This is repeated for all possible regression points, each data point being weighted by its distance from the regression point in question. GWR can be used as a tool for exploring patterns in data and generating an exploratory hypothesis for further testing, rather than as a tool for testing a priori hypotheses (Fotheringham et al., 2002). An example of the application of GWR to disease modelling is provided by Lawson et al. (2003) in a study of the spatial distribution of incidences of lip cancer in Eastern Germany. Spatially variable regression coefficients were found between the log relative risk of lip cancer and the population employed in a covariate of agriculture, forestry and fishing.

A potential problem with a GWR approach is that there may be only one attribute value at each of the locations for each of the target (e.g., diseases) and explanatory variables (PTEs in soil). To provide sufficient observations with which to fit a local regression model for each, a moving window is considered around the location of the variables (Brunsdon et al., 1999). The weighted least-squares (WLS) approach to fitting regression models provides a means by which to vary the influence of individual measurements on the fitted model. The weights usually are chosen to inversely reflect some measure of uncertainty, so that more uncertain observations are assigned less weight. Brunsdon et al. (1999) consider a range of possible weighting functions. The essential idea of GWR, whatever weighting function is chosen, is to give more weight to observations close to the location of observations at which the regression model is desired than to those observations that are farther away. For this research BioMedware Space-Stat<sup>TM</sup> (Version 3.6.20) was used to perform GWR using a maximum-likelihood approach to calculate local model regression parameters and variances. Within Space-Stat<sup>TM</sup> GWR is extended to non-linear regression procedures including Poisson regression with parameter values and parameter variances calculated from a weighted log-likelihood formulation. Poisson aspatial regression and Poisson GWR using 20 nearest neighbours were used to assess how the relationship between disease data and the PTEs vary across Northern Ireland. The kernel bandwidth can be a constant (or fixed) distance or a constant (or fixed) number of nearby observations (i.e. an adaptive distance). In this study, different weighting schemes (equal weights, Gaussian, bisquare-adaptive, and bisquare-fixed) were compared, and a cross-validation procedure was used to calculate the optimum bandwidth. For adaptive bandwidths, the optimum bandwidth was calculated as 361.39 m using a cross-validation procedure. The local parameter estimates were mapped using ArcGIS version 10, to investigate the potential dependent disease data and explanatory PTE variables.

## 4. Results

### 4.1. Geostatistical analysis of geochemical data

Spherical models were fitted to the experimental variograms, and the nugget variance ( $c_0$ ), the sill variance of the spatially dependent component ( $c$ ) and the range ( $a$ ), were estimated from the modelled variograms. The modelled variograms (Table 2) indicate spatial structure for all of the PTEs

except U, which demonstrated a highly variable distribution across the region. The coefficients of these variograms were used as inputs for ordinary kriging (OK).

Spatial structure at different ranges of correlation is observed for individual pseudo-total PTE concentrations. Shorter correlation distances ( $\leq 20$  km) are indicated for Cd, Cr, Se and Pb. Longer correlation distances ( $> 60$  km) are indicated for the PTEs of Co, Cu, V and Zn. For PTE concentrations of As and Cr two distinctive correlation distances or ranges are observed (shorter ranges 7–15 km, longer ranges  $> 40$  km). Spatial structure with two longer ranges is recorded for Ni concentrations ( $\sim 50$  km and 70 km). The nugget:sill ratio ( $c_0:c_0+c$ ) varies from 14% (Ni) to 82% (Pb) suggesting that for some of the PTEs (Table 1; As, Cr, Cu and Pb) there is considerable unresolved variation at scales finer than the resolution of the Tellus survey ( $2 \text{ km}^2$ ). Previous studies (Jordan et al., 2007; Barsby et al., 2012) have observed positively skewed distributions of untransformed soil geochemical data. It should be noted that the observed large nugget effects for several of the PTEs in this study (Pb and Cr) may be related to skewed distributions. The spatial pattern of the PTEs is linked with the distribution of geological and soil parent material. The presence of several scales of spatial structure may be related to the presence of more than one lithological or pedological source for several of the data (including As, Ni, V, Cr, Co and Cu). Barsby et al. (2012) found that outliers for Ni, Cr and V concentrations were predominantly related to samples derived from igneous/basalt soil parent material. The variable distribution for U concentrations has been related to samples derived from soil parent material typically associated with elevated U concentrations, including shales, granites and limestones (Alloway, 2005).

#### 4.2. Mapping PTEs

The OK outputs maps (shown for Ni and As) demonstrate lithological associations with the spatial distribution of PTE concentrations. The presence of the basaltic lavas of the Palaeogene Lava Group (Fig. 1a) has a strong control over the spatial distribution of PTE concentrations of Ni (Fig. 1c), Cr, Co, V and Cu and to a lesser degree for Cd and Zn. The spatial distribution of elevated concentrations of Ni, V, Cr, Co and Cu suggest a secondary lithological control in the occurrence of these PTEs associated with the Southern Upland Terrain. PTE concentrations of Cd, Zn and Pb show outlier areas with elevated values that do not closely correspond to specific lithological and pedological associations.

Arsenic is notably absent from basaltic lavas and overlying deposits. The distribution of As (Fig. 1d) is more closely associated with sandstone and shales of the Midland valley and Southern Upland Terrains (Ordovician–Silurian) where elevated concentrations of this PTE exceed the current UK soil guidelines values (SGVs) for As in soils with a residential or allotment land use (EA, 2009a).

Radon, the radioactive gas formed from the natural radioactive decay of U, is not measured in the Tellus soil survey, but the distribution of soil U demonstrates visual correlation with known areas of elevated indoor radon, including the area around Co Tyrone and the southeast part of Co. Down (GSNI, 2007). Soils overlying the granitic bedrock of the Mourne Mountains Complex show the highest U values (Table 1; max  $142.9 \text{ mg/kg}$ ) in Northern Ireland. The spread of elevated U anomalies from granite bedrock onto Lower Palaeozoic outcrop of the Hawick Group in southeast Co. Down may be due to the effects of river drainage and glacial overprinting during the Last Glacial Maximum (GSNI, 2007). Soil values on the eastern coastal fringe of the Mourne mountains suggest transport of uranium downslope towards the coast.

A correlation of elevated soil U is also noted associated with limestone bedrock which may reflect its natural accumulation in organic-rich rocks (GSNI, 2007). In Northern Ireland, this association is evident overlying the Early Carboniferous Armagh Group in south Co. Fermanagh, where elevated soil U values ( $> 4.0 \text{ mg/kg}$ ) coincide with the outcrop of limestone and mudstone-dominated Carboniferous formations. Extensive NW–SE orientated zones of elevated soil U levels present in Co. Fermanagh correspond to the outcrop of Palaeogene dolerite dykes. However, the spatial extent of these elevated U zones, suggest that the dykes are not the only source of the soil U anomalies.

The spatial distribution of radon concentrations, provided by GSNI and the Health Protection agency (HPA), (Fig. 1b; mapped by ward for maximum risk bands) are variable across Northern Ireland. Radon concentrations ( $> 25\%$  radon risk) are indicated for the granitic bedrock of the Mourne Mountains Complex which also showed the highest U values in overlying soils (Table 1;

max 142.9 mg/kg) in Northern Ireland. Wards in counties Tyrone and Fermanagh also show high concentrations of radon.

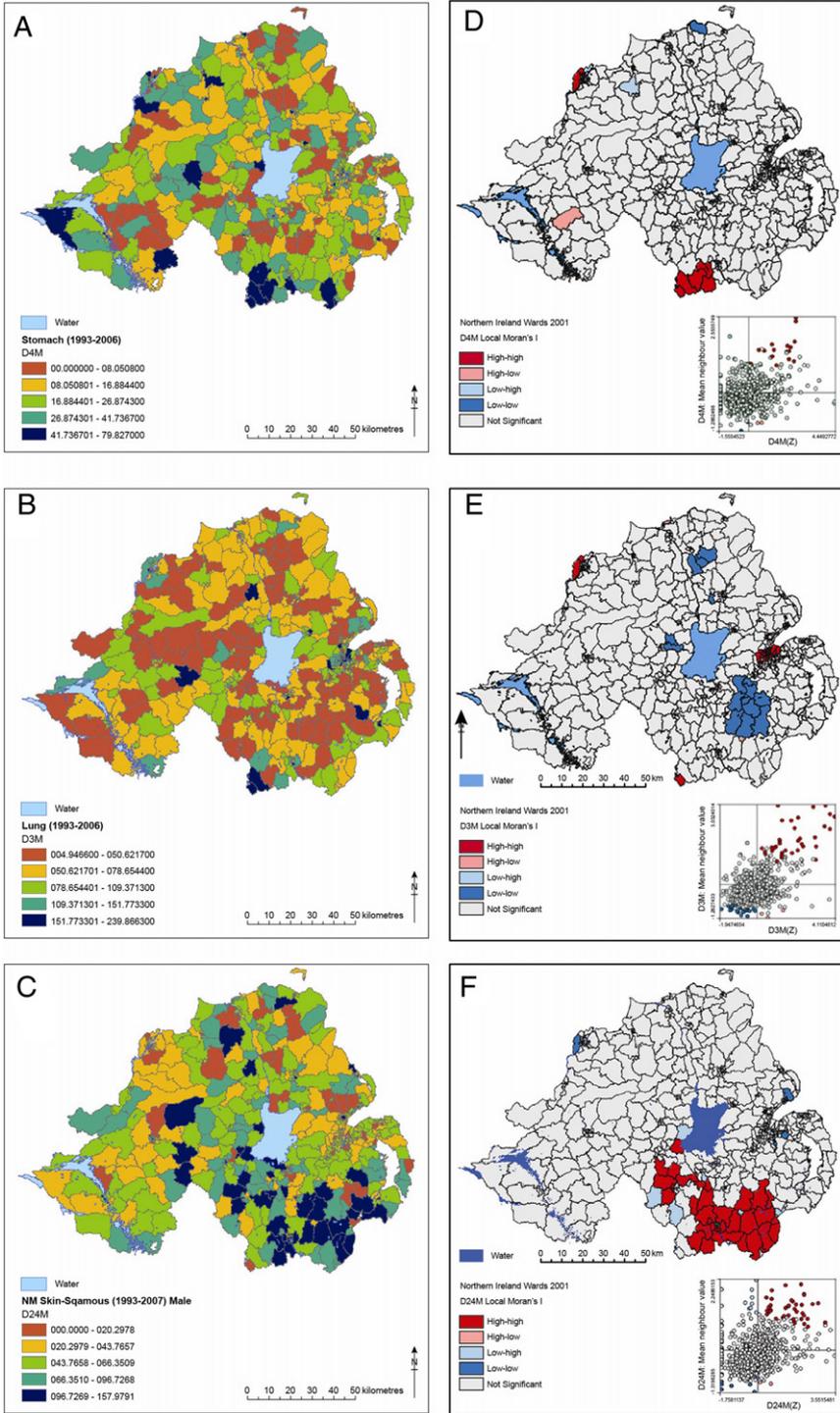
#### 4.3. Mapping age standardised incidence rates (ASIRs)

The ASIRs demonstrate categorised incidence frequencies of each ward (standardised to account for population and age) for each disease dataset by gender. Mapping ASIRs indicates spatial variability for the disease datasets of stomach, leukaemia, oesophagus, colorectal, bladder, kidney, breast cancer and mesothelioma. Results are shown for male incidences of stomach, lung and NM skin cancer (Fig. 2). The outputs demonstrate the usefulness of the standardisation process in that wards containing urban centres of high population do not dominate the spatial distribution of disease incidence. It should be noted that this may also be attributed to the fact that the rates for these diseases may be more stable and less likely to be as extreme as in sparsely populated areas. Between 2000 and 2004 stomach cancer was recorded as the fifth most common cancer in males comprising 3.3% of cases with an average of 148 males diagnosed each year (Donnelly and Gavin, 2007). Mapping ASIRs for 1993–2006, shows similarities between incidence rates of stomach cancer for males and females for the twelve year period. High incidence rates are shown for men for a number of wards in the southwest (county Fermanagh) and southeast (county Armagh) of Northern Ireland (Fig. 2a).

Lung cancer data were provided for three time periods; two different 6 year periods 1993–1999 and 2000–2006 and the complete 12 year period of 1993–2006. Mapping ASIRs indicated an increase in the number of wards with high incidence rates, most evident for female incidence rates. Along with the urban centres of Belfast and Londonderry, several wards were found to show high incidence rates of lung cancer for both male and female for the twelve year period. Highest ASIRs for both male and female across the 12 year period from 1993–2006 occur in several wards within Co Tyrone and Co Down and high male incidences rates are indicated for several wards in southern Co. Armagh (Fig. 2b). The highlighted wards have low to moderate Multiple Deprivation Measure (MDM) ranks. Donnelly and Gavin (2007) report higher incidence and mortality rates of lung cancer in the 40% most deprived areas of Northern Ireland than across Northern Ireland.

In total 10 of the NICR datasets comprised aggregated data of incidences of skin cancer. The results show variability in the incidence of skin cancer across Northern Ireland. From 2000 to 2004 malignant Melanoma of the skin was the eighth most common cancer in females and the tenth most common cancer in males (Donnelly and Gavin, 2007). Skin cancer is more commonly associated with affluence than deprivation, reflecting the risk factor of ultraviolet radiation which comes with an increased access to holidays in sunnier climates (NCR, 2003). Previous work has shown that Melanoma of the skin has a strong geographical pattern, similar for men and women in that, areas of higher relative risk were recorded along the east coast of Northern Ireland (NCR, 2007). Highest ASIRs for Melanoma for the 12 year period (1993–2006) in this research are recorded for females with a spatially dispersed pattern of wards with the highest category of incidence rates (ASIRs of 34–66 per 100,000 persons). Wards on the east coast of Northern Ireland (e.g. Newry and Mourne) show high ASIRs for both males and females.

In both 2001 and 2007, Non Melanoma skin (NMS) cancer was recorded as the most common cancer in males and females accounting for about 27% of male and 25% of female cancers of all malignant neoplasms (NCR, 2003, 2007). The results from this study examining records from 1993 to 2002 indicated that for all three cancers, the age-specific rates for both males and females increased with age, except for NMS cancer in men aged 75 years and over, where the rates were seen to decrease. Over the 12-year period a 62% increase was observed in the overall number of skin cancer and a 20% increase in the number of patients. Although a commonly diagnosed cancer, NMS cancer is rarely a cause of cancer mortality. The most commonly diagnosed NMS cancers (NCR, 2007) were basal cell carcinomas (64%), followed by squamous cell (29%). The number of cases and the age-specific rates increased with increasing age. Basal Cell Carcinoma is the most common and least dangerous form of skin cancer and accounts for over half of skin cancers. It grows slowly, mainly affecting people over 60 years, and rarely spreads, but can cause local damage and requires treatment (NCR, 2007). Squamous Cell Skin Carcinoma occurs mainly in older people and accounts for over one quarter of skin cancers (NCR, 2007). Previous research has demonstrated a geographical variation in the incidence of



**Fig. 2.** ASIR mapping for male incidences of (A) stomach cancer data (1993–2006); (B) lung cancer (1993–2006); (C) NM squamous skin cancer (1993–2007); Moran's *I* mapping for (D) stomach cancer data (1993–2006); (E) lung cancer (1993–2006); (F) NM squamous skin cancer (1993–2007).

**Table 3**

Summary table of local Moran's  $I$  index ( $I$ ): Cluster of high values (HH), cluster of low values (LL), outlier in which a high value is surrounded primarily by low values (HL), and outlier in which a low value is surrounded primarily by high values (LH) ( $p$ -value = 0.05).

Cancer/ disease type (time period)	Disease code	Gender	No. HH	No. HL	No. LH	No. LL	No. N/S	Total
Lung (1993–1999)	1	M	28	2	2	9	541	582
		F	37	1	3	23	518	582
Lung (2000–2006)	2	M	28	2	2	18	532	582
		F	25	3	3	22	529	582
Lung (1993–2006)	3	M	35	3	2	20	522	582
		F	42	3	3	25	509	582
Stomach (1993–2006)	4	M	16	2	2	2	560	582
		F	15	4	6	12	545	582
Leukaemia (1993–2006)	5	M	2	3	0	10	567	582
		F	0	4	3	0	575	582
Oesophagus (1993–2006)	6	M	5	2	3	13	559	582
		F	0	5	1	6	570	582
Colorectal (1993–2006)	7	M	8	2	2	6	564	582
		F	10	2	4	10	556	582
All January cancers excluding Non melanoma (NM) Skin cancer (1993–2006).	8	M	4	2	5	9	562	582
		F	6	0	2	3	571	582
Bladder (1993–2006)	10	M	0	3	1	12	566	582
		F	2	2	2	6	570	582
Kidney (1993–2006)	11	M	0	4	4	6	568	582
		F	2	0	4	4	572	582
Breast (1993–2006)	12	F	4	2	4	2	570	582
All NM skin (1993–1999)	13	M	31	3	4	11	533	582
		F	18	6	4	10	544	582
Mesothelioma (1993–2006)	14	M	32	14	8	1	527	582
		F	1	26	6	0	549	582
All NM skin (2000–2006)	15	M	10	1	13	7	551	582
		F	13	5	3	8	553	582
Melanoma (1993–2006)	16	M	4	1	3	7	567	582
		F	3	3	5	8	563	582
All NM skin (1993–1997)	17	M	4	6	2	5	565	582
		F	2	5	1	6	568	582
All NM skin (1998–2002)	18	M	9	3	8	5	557	582
		F	7	6	5	7	557	582
All NM skin (2003–2007)	19	M	5	3	1	1	572	582
		F	2	5	4	1	570	582
NM skin basal (1993–1997)	20	M	4	3	3	8	564	582
		F	5	3	2	2	570	582
NM skin basal (1993–2007)	21	M	6	2	3	9	562	582
		F	4	4	2	9	563	582
NM skin-Squamous	23	M	43	2	11	8	518	582
		F	16	4	6	8	548	582
NM skin-Squamous (1993–2007)	24	M	39	2	11	10	520	582
		F	23	3	5	11	540	582

NMS cancer in that incidences for males and females were much higher in Newry and Mourne, and Craigavon District Councils. The ASIR maps for basal skin cancer show spatial variability in wards with highest incidence rates. Mapping the ASIRs for squamous skin cancer for both sexes (shown for male ASIRs 1993–2006 Fig. 2c) indicates highest incidences rates in the Southeast of Northern Ireland.

#### 4.4. Results of local cluster analysis

In the maps and scatterplots results from Moran's  $I$  analysis (Table 3) the cluster/outlier type distinguishes between a statistically significant (at  $\alpha = 0.05$  level) cluster of high values (HH), cluster of low values (LL), outlier in which a high value is surrounded primarily by low values (HL),

and outlier in which a low value is surrounded primarily by high values (LH). There are a number of clusters of high and low values for the different disease datasets, although the majority of wards are “not significant” in terms of clustering of incidences of diseases for the period from 1993 to 2007. Output maps are provided for ASIRs for stomach cancer, lung and NM squamous skin cancer and these results are discussed in more detail. Several wards in county Armagh showed high ASIRs for stomach cancer for the period 1993–2006. These wards are highlighted as HH clusters (high incidences of stomach cancer surrounded by wards of similar high occurrences of this disease) (Fig. 2d). Several wards within the urban centres of Belfast and Londonderry showed high ASIRs for lung cancer for the period 1993–2006. These wards are highlighted as HH clusters (Fig. 2e; high incidences of lung cancer surrounded by wards of similar high occurrences of this disease). The most interesting results from Moran’s  $I$  analysis are shown for NM skin cancer data. Mapped ASIRs demonstrated highest incidences rates for NM squamous skin cancer for both sexes were located in the Southeast of Northern Ireland. Results from Moran’s  $I$  analysis indicates that wards located in this area show HH clusters (Table 3: 48 wards highlighted as HH clusters). Mapped ASIRs differentiated for male and female showed that although overall incidence rates are higher for females, a large portion of the pattern displayed (Fig. 2f) is controlled by a high male incidence rate. The southeast of Northern Ireland showed a grouping of wards with high incidence rates for males. These same wards show HH clusters for male incidence rates (39 wards show HH clusters) and female rates (23 wards show HH clusters) for NM squamous skin cancer. The results for NM Basal skin disease showed spatial variability in the location of wards with highest incidence rates distributed across Northern Ireland and a much lower number of clusters (6 wards show HH clusters for male incidences of NM basal skin compared with 39 HH clusters for NM squamous skin cancer).

#### 4.5. Investigation of relations between disease data and environmental hazards (PTEs)

The results indicate that there are spatially varying patterns of both elevated PTEs and incidences of diseases across Northern Ireland. To explore the spatial correlations and to model any spatially varying relationships between PTEs in soils and epidemiological data, GWR provides a local model of the relationship between soil geochemistry and disease data by fitting a regression equation to a subset of the dataset within a window of each target area.

##### 4.5.1. Use of GWR to test for correlations between the diseases and PTEs

Aspatial Poisson regression was applied to all datasets as an initial exploratory investigation. Aspatial Poisson regression demonstrated a weak global relationship between male stomach cancer data and As (correlation coefficient  $r = 0.075$ ) and between male NM squamous skin cancer data and maximum radon risk (correlation coefficient  $r = 0.233$ ) for a significance level of 0.05. In global regression geographic location is not taken into account. Such regression models are appropriate only where the relations between properties (e.g. PTEs and disease data) can reasonably be modelled as invariant over space, i.e., there is no variability with geographic distribution. Where the relations between variables are expected to change with geographical location, as investigated in this study, a spatially non-stationary model can be applied to allow variation in the parameters of the model from place to place.

##### 4.5.2. GWR results

Poisson GWR was conducted for the disease datasets of stomach cancer and NM skin cancer as the dependent variables for As and maximum radon risk, respectively, as potential explanatory variables. Using a maximum weighted likelihood approach within SpaceStat™ the regression parameters, parameter variances, residuals, standard errors and the “local model”  $r^2$  were calculated. The outputs shown are produced using Poisson GWR with 20 nearest neighbours and a bisquare-adaptive weighting scheme. The coefficient standard error was mapped in all cases. This measures the reliability of each coefficient estimate. Confidence in estimates is higher when standard errors are small in relation to the actual coefficient values. Large standard errors may indicate problems with local multicollinearity which indicates redundancy due to two or more variables indicating the same control (Wheeler, 2007).

**4.5.2.1. Investigating incidences of stomach cancer and As.** Several wards within South Armagh showed clusters of high male and female incidences of stomach cancer. GWR indicates a grouping of moderate correlations (Fig. 3a;  $r$  values = 0.44–0.69) suggesting an interesting association between stomach cancer and As for the same group of wards in South Armagh. Intercept values show a pattern of largest values along the southeastern coast and in county Fermanagh. Where intercept values are large (Fig. 3b), this suggests stomach cancer incidence values vary irrespective of As. The standard errors (Fig. 3c) for stomach cancer (1993–2006) and As show that over- and underpredictions are randomly distributed. A grouping of the highest observed correlation values (Fig. 3a;  $r$  values range from 0.27 to 0.69) between lung cancer incidences and maximum As (for each ward) exhibits an interesting association with highest recorded pseudo-totals concentrations of As (Fig. 1e; 35.18–271.2 mg/kg). This coincides with the Dalradian metamorphic rocks of Co. Tyrone, and the organic-rich and pyritised Carboniferous rocks of Co. Fermanagh. In Co. Tyrone, highest correlation values ( $r$  values = 0.44–0.69) are associated with the As-bearing Devonian Shanmullagh Formation in the Fintona Block, along with the broader zone of As anomalies across the Fintona Block suggested as response to glacial dispersion of materials (GSNI, in preparation).

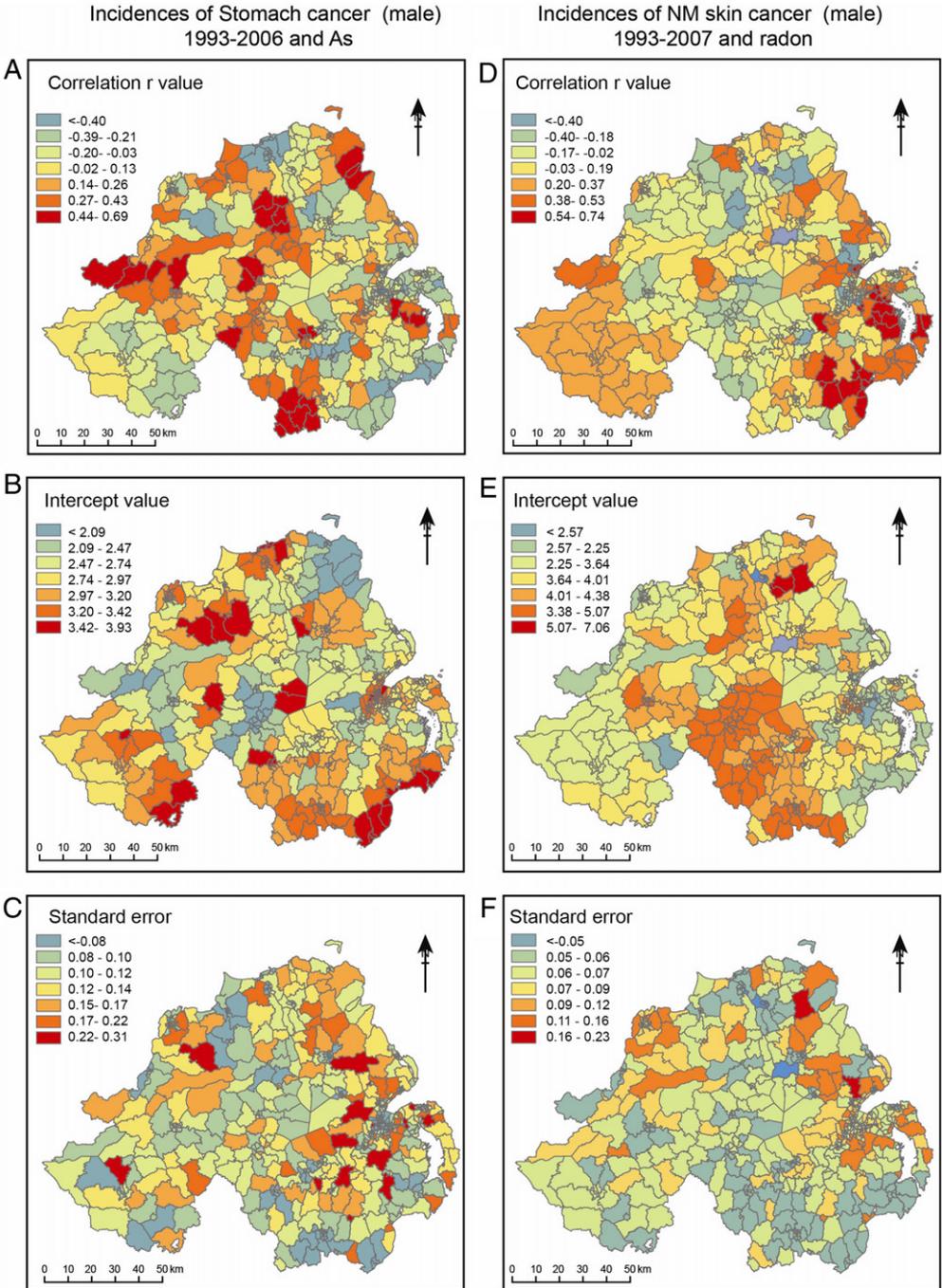
**4.5.2.2. Investigating incidences of NM squamous skin cancer and radon.** Findings from GWR analysis correspond with the occurrence of high ASIRs for skin cancer for both male and females in the southwest of Northern Ireland. Highest correlation values are found for wards in this area between NM squamous skin cancer and maximum risk band for radon (Fig. 3d;  $r$  values = 0.54–0.74). This correlates with proximity to the south coastal areas but also with the mountainous area of the Mournes. Largest intercept values (Fig. 3e) are observed in county Armagh and through central Northern Ireland. This suggests incidences of NM squamous skin cancer vary in these areas irrespective of radon levels. The standard errors for NM skin-squamous cancer (1993–2006) and radon show that over- and underpredictions are randomly distributed (Fig. 3f).

## 5. Discussion

Soil geochemistry concentrations in shallow soils in Northern Ireland reveal elevated Ni and Cr concentrations across a significant area of the Antrim Plateau (Barsby et al., 2012). These levels exceed the current UK soil guidelines values (SGVs) for Ni in soils (EA, 2009b). On the basis of total concentration, these elevated Ni and Cr concentrations would cause concern for human health risk assessment. Bioaccessibility testing by Barsby et al. (2012) indicated that the PTEs of Ni (Fig. 1d) and Cr in basaltic lavas of the Palaeogene Lava Group are not readily bioaccessible and as a result do not present a risk to human health. Results from Moran's  $I$  cluster analysis and GWR did not indicate a correlation between the PTEs of Ni and Cr and the disease data studied in this research.

Studies have indicated a link between stomach cancer and exposure to Pb, As, Cd and Zn, although these studies did not account for quantitative data on smoking (Selevan et al., 1985; Fu and Boffetta, 1995). The strongest relationship indicated in this study for stomach cancer was observed with As for wards in South Armagh. Very high (>43.7 mg/kg) concentrations of As are recorded in soils over the Hawick Group of Co. Armagh and Co Down, possibly reflecting the presence of Late Caledonian lamprophyre dykes or veinlets of galena following cleavage planes in the bedrock (GSNI, in preparation). High levels of bioaccessible As (Fig. 1f; BAF %) were found to be associated with the metasediments of the Gala Group of the Southern Uplands-Down Longford Terrain (Barsby et al., 2012). GSNI (in preparation) record that no anthropogenic influence on the concentration or distribution of soil As has been observed for Northern Ireland. PTE concentrations of Cd, Zn and Pb show outlier areas which exceed the respective SGV value but these do not closely correspond to specific lithological and pedological associations. Results for Pb, Cd and Zn from this study did not show any clear relationship with disease data. A potential relationship between stomach cancer and elevated levels of bioaccessible As in soils overlying the Hawick and Gala Groups of the Southern Uplands-Down Longford Terrain is very tentative and requires more further research.

Previous research has strongly indicated that exposure to radon is associated with an increased frequency of lung cancer. Although mining sources create greatest levels of exposure, low level radon exposure in the home environment is thought to account for approximately 10% of lung cancers



**Fig. 3.** GWR results for Stomach cancer data (1993–2006) and arsenic (mg/kg) (A) correlation value; (B) intercept and (C) standard error; GWR results for NM Squamous Skin cancer (1993–2007) and radon (D) correlation value; (E) intercept and (F) standard error.

(Belpomme et al., 2007) with the duration of the exposure being generally longer (Bofetta, 2004). Autier (2007) describes how most of the ionising radiation affecting humans originates from medical X-rays and background radiation including cosmic radiation and terrestrial radiation from radon decay products. Levels of exposure to radon in the range of  $1000 \text{ Bq/m}^3$  or more have been associated with lung cancer (Autier, 2007). Granitoid parent rocks have been linked to increased radionuclides in drinking water. In Northern Ireland, soils overlying granite bedrock of the Mourne Mountains Complex show the highest U values ( $>5.5 \text{ mg/kg}$ ). The spread of U anomalies from granite bedrock onto Lower Palaeozoic outcrop have been attributed to river drainage patterns and glacial overprinting accounting for the distribution of elevated U soil concentrations to extend from the Eastern Mournes Complex onto the Hawick Group near Annalong in southeast Co. Down (GSNI, in preparation). The mapped total airborne radiation shows a comparable spatial distribution to mapped U soil concentrations with highest U values corresponding to elevated gamma radiation dosage from granite bedrock of the Mournes Complex, Co. Down. The association between lung cancer and radon is not evident from the findings of this research. Results from Moran's  $I$  cluster analysis for lung cancer data concur with previous findings by Donnelly and Gavin (2007) in that clusters of high incidences of lung cancer reflect more deprived areas linked to higher levels of tobacco use within Northern Ireland. GWR analysis in the current research showed a pattern of low but significant correlation between radon and NM squamous skin cancer for both sexes in the Southeast of Northern Ireland. This correlates with proximity to the south coastal areas but also with the mountainous area of the Mournes. Results for basal skin cancer do not show an obvious association observed with radon concentrations. Previous research (Boyle et al., 2003; Quinn et al., 2005; Gavin et al., 2012), encompassing data on skin cancer from the Republic of Ireland and the NICR, concurs with higher levels of skin cancers found for coastal areas. The results also showed that affluence was a key indicator confirming research undertaken previously by NICR in 2001.

Hoey et al. (2007) describe how the incidence of skin cancer, both melanoma and nonmelanoma skin cancer, is rising and stress the importance of an evaluation of trends in skin cancer to allow better planning of the future development of skin cancer services (NICE). Previous studies have noted an association between radon emission and NM skin cancers. Etherington et al. (1996) studied incidences of 14 major cancers in Devon and Cornwall in relation to local radon levels. They found that incidence rates for cancers were similar across all domestic radon categories, including the results for lung cancer where radon had been claimed to be a risk factor. The exception was found to be NM skin cancers where the rate of incidence was found to be elevated in high-radon areas. Similar findings relating high radon levels with skin cancers have been reported in the literature by Eatough and Henshaw (1991, 1992) and Sevcova et al. (1978). Eatough and Henshaw (1992) calculate the theoretical dose of radiation received at the basal layer of the skin as  $2.5 \text{ mSv/year}$  for the population average radon concentration of  $20 \text{ Bq/m}^3$ .

The findings from this research suggesting an association between elevated radon areas and NM squamous skin cancer are by no means conclusive but may indicate that further work in this area is warranted to confirm or refute this apparent relationship.

## 6. Conclusions

This research explores the relationship between environmental exposure to potentially toxic elements in soil and cancer disease data across Northern Ireland. As such this work is part of the multidisciplinary field of science that has been termed medical geology examining the relationship between the geological environment and health issues in humans. The study therefore adds to the breadth of previous work in this field and more specifically provides a spatial framework that has not been fully exploited within this area to date. There are several ways to approach the influence of occupational and environmental exposures on cancer incidence. This research addresses the spatial distribution of exposure to diseases through geocoded incidences of cancers and the lithological and pedological associations of potentially toxic elements through soil geochemistry. The incidence of twelve different cancer types (lung, stomach, leukaemia, oesophagus, colorectal, bladder, kidney, breast, mesothelioma, melanoma and non melanoma skin cancer both basal and squamous, were examined in the form of twenty-five coded datasets comprising aggregates over the 12 year period

from 1993 to 2006. Through the use of GIS and geostatistical techniques the spatial relations between these epidemiological, soil and geological data types have been investigated. In addition, the fraction of the PTEs of the soil that are bioaccessible or soluble in the gastrointestinal tract have been considered to further refine the relationship between trace element abundance and human epidemiology in Northern Ireland.

Local Moran  $I$  analysis identified clustering of wards with high and low incidences of the different cancers and outlier wards of high or low incidences surrounded by wards exhibiting very different disease statistics. The use of GWR allowed this analysis to be taken further and enabled the relationship between different cancers and PTEs in the soil and radon to be investigated. The results show comparisons of the geographical incidence of certain cancers (stomach cancer and NM squamous skin cancer) in relation to concentrations of certain PTEs (arsenic levels in soils and radon were identified). As acknowledged by previous studies it is important to bear in mind that the spatial variation observed in the incidences of the different cancer diseases included in this study, does not mean that the spatial location itself causes the cancer. The influence of socio-economic differences and lifestyle factors in the population of Northern Ireland need to be taken into account in assessing any associations found. This is an area that has the opportunity for further work within this research. Findings from this work have implications for regional human health risk assessments.

## Acknowledgements

This research was funded by the British Geological Survey (BGS) University Funding Initiative (BUFI). The NICR and GSNI are gratefully thanked for the use of the datasets. Colin Fox, NICR, is thanked for preparation of the anonymised cancer disease datasets. The NICR is funded by the Public Health Agency for N. Ireland. The Tellus Project was carried out by GSNI and funded by the Department for Enterprise, Trade and Investment (DETINI) and The Rural Development Programme through the Northern Ireland Programme for Building Sustainable Prosperity. Bioaccessibility testing was carried out at the BGS laboratories in Keyworth Nottingham. Mark Cave and Joanna Wragg are gratefully thanked for their assistance in these analyses. Maura Pringle is thanked for assistance with the preparation of figures. This project was given ethics approval by the NHS National Research Ethics Committee.

## References

- Alexander, F.E., Boyle, P., 1996. *Methods for Investigating Clustering of Disease*. IARC Scientific Publications, p. 135.
- Alloway, B.J., 2005. Bioavailability of elements in soil. In: Selinus, O., Alloway, B.J., Centeno, J.A., Finkelman, R.B., Fuge, R., Lindh, U., et al. (Eds.), *Essentials of Medical Geology*. Elsevier Academic Press, Amsterdam.
- Autier, P., 2007. *Attributable Causes of Cancer in France in the Year 2000*. IARC, Lyon.
- BARGE–INERIS, 2011. UBM Procedure for the Measurement of Inorganic Contaminant Bioaccessibility from Solid Matrices.
- Barsby, A., McKinley, J., Ofterdinger, U., Young, M., Cave, M.R., Wragg, J., 2012. Bioaccessibility of trace elements in soils in Northern Ireland. *Science of the Total Environment* 433, 398–417.
- Belpomme, D., Irigaray, P., Hardell, L., Clapp, R., Montagniere, L., Epstein, S., Sasco, A.J., 2007. The multitude and diversity of environmental carcinogens. *Environmental Research* 105, 414–429.
- Bland, M., 1996. *An Introduction to Medical Statistics*. Oxford Medical Publications, Oxford.
- Bofetta, P., 2004. Epidemiology of environmental and occupational cancer. *Oncogene* 23, 6392–6403.
- Boyle, P., Golia, S., Daly, P., Johnston, P.G., Robertson, G., 2003. Cancer mortality in Ireland, 1926–1995. *Annals of Oncology* 14, 323–332. <http://dx.doi.org/10.1093/annonc/mdg062>.
- Brugge, D., Durant, J.L., Rioux, C., 2007. Near-highway pollutants in motor vehicle exhaust: A review of epidemiologic evidence of cardiac and pulmonary health risks. *Environmental Health* 6, 23. <http://dx.doi.org/10.1186/1476-069X-6-23>.
- Brunsdon, C., Aitkin, M., Fotheringham, S., Charlton, M.E., 1999. A comparison of random coefficient modelling and geographically weighted regression for spatially non-stationary regression problems. *Geographical and Environmental Modelling* 3, 47–62.
- Carsin, A-E., Sharp, L., Comber, H., 2009. *An Atlas of Cancer in Ireland 1994–2003*. The National Cancer Registry, Cork, ISBN: 978-0-9554970-1-8, <http://www.ncr.ie>.
- Central Statistics Office (CSO), 2011. *The Statistical Yearbooks of Ireland*. CSO, Cork.
- Cook, D.G., Pocock, S.J., 1983. Multiple regression in geographical mortality studies with allowance for spatially correlated errors. *Biometrics* 39, 361–371.
- Donnelly, D.W., Gavin, A.T., 2007. Survival of cancer patients in Northern Ireland: 1993–2004. In: *Northern Ireland Cancer Registry*. Queen's University, Belfast, 168 + Summary 6.
- Eatough, J.P., Henshaw, D.L., 1991. Radon dose to the skin and the possible induction of skin cancers. *Radiation Protection Dosimetry* 39, 33–37.

- Eatough, J.P., Henshaw, D.L., 1992. Radon and thoron associated dose to the basal layer of the skin. *Physics in Medicine and Biology* 37, 955–967.
- Environment Agency (EA), 2009a. Soil Guideline Values for Arsenic in Soil. Environment Agency, Bristol.
- Environment Agency (EA), 2009b. Soil Guideline Values for Nickel in Soil. Environment Agency, Bristol.
- Etherington, D.J., Pheby, D.F.H., Bray, F.I., 1996. An ecological study of cancer incidence and radon levels in South West England. *European Journal of Cancer* 32A (7), 1189–1197.
- Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons, Chichester, U.K.
- Fu, H., Boffetta, P., 1995. Cancer and occupational exposure to inorganic lead compounds: a meta-analysis of published data. *Occupational & Environmental Medicine* 52 (2), 73–81.
- Gavin, A., Boyle, R., Donnelly, D., Donnelly, C., Gordon, S., McElwee, G., O'Hagan, A., 2012. Trends in skin cancer knowledge, sun protection practices and behaviours in the Northern Ireland population. *European Journal of Public Health* 22 (3), 408–412. <http://dx.doi.org/10.1093/eurpub/ckr087>.
- Geological Survey Northern Ireland (GSNI), 2007. Tellus project overview. <http://www.bgs.ac.uk/gsnitellus/index.html>. Accessed 03/04/08, 1632.
- GSNI. Tellus: User Guide and Atlas. Geological Survey of Northern Ireland, Belfast (in preparation).
- Goovaerts, P., 2010. How do multiple testing correction and spatial autocorrelation impact areal boundary analysis? *Spatial and Spatio-temporal Epidemiology* 1 (4), 219–229.
- Gringarten, E., Deutsch, C.V., 2001. Teacher's aide variogram interpretation and modeling. *Mathematical Geology* 33 (4), 507–534.
- Hoey, S.E.H., Devereux, C.E.J., Murray, L., Catney, D., Gavin, A., Kumar, S., Donnelly, D., Dolan, O.M., 2007. Skin cancer trends in Northern Ireland and consequences for provision of dermatology services. *British Journal of Dermatology* 156, 1301–1307. <http://dx.doi.org/10.1111/j.1365-2133.2007.07936.x>.
- Jensen, J.L., Corbett, P.W.M., Pickup, G.E., Ringrose, P.S., 1996. Permeability variograms, geological structure, and flow performance. *Mathematical Geology* 28, 419–435.
- Jordan, C., Zhang, C., Higgins, A., 2007. Using GIS and statistics to study influences of geology on probability features of surface soil geochemistry in Northern Ireland. *Journal of Geochemical Exploration* 93, 135–152.
- Lawson, A.B., 2006. Disease cluster detection: a critique and a Bayesian proposal. *Statistics in Medicine* 25, 897–916.
- Lawson, A.B., Browne, W.J., Vidal-Rodiero, C.L., 2003. Disease Mapping with WinBUGS and MLwiN. John Wiley & Sons, Inc, New York.
- Majno, G., Joris, I., 2004. Cells, Tissues, and Disease: Principles of General Pathology, 2nd ed. Oxford University Press.
- McBratney, A.B., Webster, R., 1986. Choosing functions for semivariograms of soil properties and fitting them to sample estimates. *Journal of Soil Science* 37, 617–639.
- McKinley, J.M., Lloyd, C.D., Ruffell, A.H., 2004. Use of variography in permeability characterisation of visually homogeneous sandstone reservoirs with examples from outcrop studies. *Mathematical Geology* 36 (7), 761–779.
- National Cancer Registry (NCR), 2007. Annual Statistical report. <http://www.ncri.ie/pubs/pubs.shtml>.
- National Cancer Registry (NCR), 2003. Annual Statistical report. <http://www.ncri.ie/pubs/pubs.shtml>.
- National Cancer Registry (NCR), 2011. Annual Statistical report. <http://www.ncri.ie/pubs/pubs.shtml>.
- National Institute for Health and Clinical Excellence (NICE). <http://www.nice.org.uk/>.
- Northern Ireland Cancer Registry (NICR), 2008. <http://www.qub.ac.uk/researchcentres/nicr/AboutUs/>, 2008. Accessed 03/04/08.
- Northern Ireland Statistics and Research Agency (NISRA), 2004. Registrar General Annual Reports 2001; 2003. Northern Ireland Statistics and Research Agency, Belfast. <http://www.nisra.gov.uk/>.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* 30 (7), 683–691.
- Pebesma, E.J., Wesseling, C.G., 1998. Gstat, a program for geostatistical modelling, prediction and simulation. *Computers & Geosciences* 24, 17–31.
- Quinn, M., Wood, H., Cooper, N., Rowan, S., 2005. Cancer atlas of the United Kingdom and Ireland 1991–2000. ONS SMPS No. 68. Palgrave MacMillan, Basingstoke.
- Selevan, S., Landrigan, P.J., Stern, F.B., Jones, J.H., 1985. Mortality of lead smelter workers. *American Journal of Epidemiology* 122, 673–683.
- Sevcova, M., Seve, J., Thomas, J., 1978. Alpha irradiation of the skin and the possibility of late effects. *Health Physics* 35, 803–806.
- Smyth, D., 2007. Methods used in the Tellus Geochemical Mapping of Northern Ireland. British Geological Survey (OR/02/022).
- Steingraber, S., 2010. Living Downstream: An Ecologist's Personal Investigation of Cancer and the Environment, second ed. Da Capo Press.
- Wragg, J., Cave, M., Taylor, H., Basta, N., Brandon, E., Casteel, S., et al. 2009. Interlaboratory Trial of a Unified Bioaccessibility Procedure. OR/07/027.
- Wheeler, D., 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A* 39 (10).