Science of the Total Environment 468–469 (2014) 240–248

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Neural network integration of field observations for soil endocrine disruptor characterisation

M.J. Aitkenhead *, S.M. Rhind, Z.L. Zhang, C.E. Kyle, M.C. Coull

The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, Scotland, UK

HIGHLIGHTS

- A neural network was used to predict endocrine disrupting compound concentrations.
- Soil sample information from the Scottish Soils database was used in the modelling.
- We showed that concentration of several of the compounds could be predicted.
- We identify and quantify the most important input parameters for each compound.
- The method can be used for predicting the level of EDC concentration in the field.

ARTICLE INFO

Article history:
Received 18 June 2013
Received in revised form 2 August 2013
Accepted 2 August 2013
Available online 13 September 2013

Editor: Damia Barcelo

Keywords:
Soil
Endocrine disruptor
Neural network
National Soil Inventory of Scotland
Sample analysis

ABSTRACT

A neural network approach was used to predict the presence and concentration of a range of endocrine disrupting compounds (EDCs), based on field observations. Soil sample concentrations of endocrine disrupting compounds (EDCs) and site environmental characteristics, drawn from the National Soil Inventory of Scotland (NSIS) database, were used. Neural network models were trained to predict soil EDC concentrations using field observations for 184 sites. The results showed that presence/absence and concentration of several of the EDCs, mostly no longer in production, could be predicted with some accuracy. We were able to predict concentrations of seven of 31 compounds with r^2 values greater than 0.25 for log-normalised values and of eight with log-normalised predictions converted to a linear scale. Additional statistical analyses were carried out, including Root Mean Square Error (RMSE), Mean Error (ME), Willmott's index of agreement, Percent Bias (PBIAS) and ratio of root mean square to standard deviation (RSR). These analyses allowed us to demonstrate that the neural network models were making meaningful predictions of EDC concentration. We identified the main predictive input parameters in each case, based on a sensitivity analysis of the trained neural network model. We also demonstrated the capacity of the method for predicting the presence and level of EDC concentration in the field, identified further developments required to make this process as rapid and operator-friendly as possible and discussed the potential value of a system for field surveys of soil composition.

1. Introduction

Many chemicals that are used in the manufacturing of products as diverse as plastics, pesticides, cosmetics, paints and electrical equipment, amongst others, can act like steroid and other hormones or interfere with their action in animal tissues by binding to their receptors or otherwise affecting the normal functioning of those biological systems. Through these actions health and reproductive function can be compromised in humans and animals (Rhind, 2005, 2008, 2009). These compounds are collectively termed endocrine disrupting compounds (EDCs). EDCs are ubiquitous in the terrestrial environment. They are derived from either point or diffuse sources, mostly associated with human activities, such as manufacturing and waste recycling but many EDCs are semi-volatile organic contaminants (SVOCs) and so they can be transported long distances not only in gaseous phase but, in some cases, also in solid form, bound to particulates (Fu and Kawamura, 2010; Salapasidou et al., 2011), depending upon the physicochemical properties of the individual contaminants and atmospheric conditions (Teil et al., 2006; Wang et al., 2008; Zeng et al., 2009).

Rates of atmospheric deposition and washout are influenced by rainfall and temperature; in colder conditions, such as at higher altitudes or latitudes, there is greater particle scavenging and higher washout to terrestrial surfaces during precipitation (Thuren and Larsson, 1990; Teil et al., 2006; Callen et al., 2008; Wang et al., 2011). Once deposited, the accumulation of anthropogenic chemicals within soils varies greatly with soil characteristics and the specific chemicals themselves (Gibson et al., 2010) and the impact of EDCs in soil are greatly influenced by the quantity and composition of clay, organic matter and humic substances

* Corresponding author. Tel.: +44 1224 395257.
E-mail address: matt.aitkenhead@hutton.ac.uk (M.J. Aitkenhead).

0048-9697/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.scitotenv.2013.08.007

present (Senesi and Loffredo, 2008), although the nature of the relevant interactions is not well known. Yu and Liu (2013) demonstrate that the sorption rates of different pharmaceutical compounds in soils were highly variable and that degradation rates were also highly compound-specific. Soil microbial activity and organic matter content were identified as strong controlling factors in sorption and decomposition of the compounds studied.

Loffredo and Senesi (2000) also showed that different rates of adsorption of EDCs occurred at different depths within a soil profile, perhaps indicating a relationship between the age and reactivity of soil organic matter and its capacity to adsorb these compounds. Consequently, variation in concentrations of different EDCs was expected with soil type and other related environmental components such as vegetation type, climate and topography. The degree of variation was expected to differ with EDC type, according to their capacity for adsorption onto clays and soil organic matter.

Here, we attempt to determine the degree to which relatively basic, field observations can be used to evaluate the concentrations of a number of EDCs present in Scottish soils. We do not intend to replace the laboratory-based chemical analysis of samples, as it is highly unlikely that EDC concentration could be as accurately evaluated as through laboratory analysis, based on field observations alone. However, we do aim to show that using a few simple observations and measurements, it is possible to estimate EDC concentrations with an accuracy that would be of use not only to the field observer, but also to land managers and policymakers (e.g. as an approach to identification of priority areas for monitoring). It is known that soil factors such as texture are strongly related to more specific or subtle parameters within the soil, such as carbon turnover (Lee et al., 2006), biological activity and productivity (Cable et al., 2008), organic matter dynamics (Grandy et al., 2009) and nutrient status (Ige et al., 2007). Several of these factors, for example pH, organic matter content and texture, are known to be strongly related to specific characteristics, particularly those related to critical loads (the estimated upper limit for some pollutant beyond which significant harmful effects on the environment begin to occur) (Towers and Paterson, 1997; Kernan et al., 1998).

There are a few studies reporting gross geographic variations in soil levels of some EDCs on a continental or global scale (e.g. Meijer et al., 2003) but little is known of the pattern of variability in concentrations on a much smaller scale and how they are influenced by multiple factors, including soil and vegetation types, proximity to sources and chemical class. Thus, risks associated with exposure cannot be quantified properly. Liao et al. (2012) reported on the distribution of specific EDCs in industrialised areas and showed that concentrations decreased over time following reduction in use. Qiang et al. (2013) showed that pH could have an important impact on the sorption and subsequent removal of many EDCs, but that the effect could be positive or negative depending upon the hydrophobicity of the compounds in question. Gong et al. (2012) demonstrated strong links between EDC concentrations and dissolved organic carbon and particulate organic carbon in river waters. The binding of these compounds to organic material in soils is equally important. Additionally, Gong et al. (2012) showed that hydrophobicity of the compounds contributed strongly to their sorption onto organic material. These findings are in agreement with those of Murillo-Torres et al. (2012) who also argued that soils act as a filter and buffer for these pollutants, binding them for longer periods and leading to their increased degradation and thus reducing their release into groundwater and aquifers. Kumar et al. (2011) demonstrated an ecologically engineered artificial wetland system that reduced concentrations of estrogen compounds from wastewater. Each of these studies shows a relationship with individual factors, but there is little information about how EDC concentration is affected by multiple factors.

Determination of soil EDC profiles by analysis of large numbers of samples for multiple compounds is currently very labour-intensive and costly and so the development of a model designed to predict soil burdens, based on available data, would be valuable for both research and environmental protection purposes. While the number of different processes involved in the accumulation of specific endocrine disruptors within soil and the variable timescales over which the process occurs, make process-based simulation of EDC accumulation extremely difficult, it may be possible to detect and model underlying relationships between soil environmental conditions and endocrine disruptor accumulation using data-mining techniques, if sufficient data are available.

Neural networks for modelling relationships between parameters within complex systems have been applied, previously, to a range of environmental topics including soil parameterisation and characterisation (Marchant and Onyango, 2003; Kavdir, 2004) and landscape evaluation (Kuplich, 2006; Aitkenhead and Dyer, 2007). Recent work has shown that neural networks can capture the relationships between environmental parameters and soil characteristics, including colour and chemical composition (Aitkenhead et al., 2012, 2013). One particularly effective way of gaining an understanding of the processes captured by a neural network model is to carry out a partial derivative analysis of the neural network's input/output transformations. This can be achieved in many ways, but the most successful is the Connection Weights approach of Olden and Jackson (2002) and Olden et al. (2004). A simplified version of the Connection Weights approach (Aitkenhead et al., 2012), based on the original work, has also been shown to be effective.

The aims of this study were (a) to develop a neural network model for the prediction of EDCs using environmental factors that can influence soil properties and that are easily assessed in the field, and (b) to use this model to assess the relative importance of individual factors in the determination of soil concentrations of selected EDCs. Knowledge and improved understanding of how these factors relate to EDC concentrations will contribute to the understanding of patterns of EDC exposure in humans and animals and associated potential risk.

2. Methods

2.1. NSIS database development and characteristics

The second National Soils Inventory of Scotland (NSIS2), from which the data used in this study were derived, comprised a second survey of Scotland's soils designed to create a new dataset which included environmental, morphological and analytical data on a systematic basis, as well as field validation for mapping (Lilly et al., 2010).

Data generated from the survey and analysis included the following relevant information: site and environmental data, such as the elevation, vegetation, climate and slope; physical characteristics of the soil horizons, including horizon depth, colour and structure down to the parent material (or approximately 2 m depth, if the soil was deeper); and chemical analysis of the samples taken at each point. This included pH, carbon content, loss on ignition, particle size analysis (mineral samples only) and exchangeable cations. In addition to the NSIS survey and soil analysis work, analysis of endocrine disruptor compound (EDC) concentration was carried out.

For the EDC work, representative soil samples (0–5 cm depth) were collected at 20 km grid intersects throughout Scotland, using a soil corer (7.5 cm diameter $\times$ 5 cm depth). This component of the soil was considered likely to be exposed to the greatest degree of contamination resulting from atmospheric deposition of EDCs and represented the soil component to which most animals might be exposed, either because they live above the ground, where they may ingest soil or inhale re-volatilised EDCs (e.g. grazing animals), or because they live within the soil where they may be exposed through ingestion, inhalation or direct contact (e.g. soil invertebrates). Samples were collected from different regions of Scotland, during spring and summer, over a three year period (year 1: Central and East; year 2: South and West; year 3: North and West) They were wrapped in aluminium foil and stored at −20 °C until they were freeze dried and sieved (2 mm) to remove stones and large particles before analysis.