Spatially Resolved Distribution Models of POP Concentrations in Soil: A Stochastic Approach Using Regression Trees

ARTICLE *in* ENVIRONMENTAL SCIENCE AND TECHNOLOGY · DECEMBER 2009

Impact Factor: 5.33 · DOI: 10.1021/es902076y · Source: PubMed

8 AUTHORS, INCLUDING:

Jiri Jarkovsky
Masaryk University
116 PUBLICATIONS **974** CITATIONS

Milan Sanka
Masaryk University
23 PUBLICATIONS **242** CITATIONS

Ivan Holoubek
Masaryk University
216 PUBLICATIONS **3,508** CITATIONS

Jana Klanova
Masaryk University
163 PUBLICATIONS **2,547** CITATIONS

Environ. Sci. Technol. 2009 *43,* 9230–9236

Spatially Resolved Distribution Models of POP Concentrations in Soil: A Stochastic Approach Using Regression Trees

KLÁRA KUBOŠOVÁ,[†] JIŘÍ KOMPRDA,[†] JIŘÍ JARKOVSKÝ,[†,§] MILAN SÁŇKA,[†] ONDŘEJ HÁJEK,[‡] LADISLAV DUŠEK,[†,§] IVAN HOLOUBEK,[†] AND JANA KLÁNOVÁ*,[†]

RECETOX (Research Centre for Environmental Chemistry and Toxicology), Kamenice 126/3, 625 00 Brno, Czech Republic, Department of Botany and Zoology, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic, and Institute of Biostatistics and Analyses, Masaryk University, Kamenice 126/3, 625 00 Brno, Czech Republic

Received July 24, 2009. Revised manuscript received October 5, 2009. Accepted October 26, 2009.

Background concentrations of selected persistent organic pollutants (polychlorinated biphenyls, hexachlorobenzene, *p,p'*-DDT including metabolites) and polyaromatic hydrocarbons in soils of the Czech Republic were predicted in this study, and the main factors affecting their geographical distribution were identified. A database containing POP concentrations in 534 soil samples and the set of specific environmental predictors were used for development of a model based on regression trees. Selected predictors addressed specific conditions affecting a behavior of the individual groups of pollutants: a presence of primary and secondary sources, density of human settlement, geographical characteristics and climatic conditions, land use, land cover, and soil properties. The model explained a high portion of variability in relationship between the soil concentrations of selected organic pollutants and available predictors. A tree for hexachlorobenzene was the most successful with 76.2% of explained variability, followed by trees for polyaromatic hydrocarbons (71%), polychlorinated biphenyls (68.6%), and p,p'-DDT and metabolites (65.4%). The validation results confirmed that the model is stable, general and useful for prediction. The stochastic model applied in this study seems to be a promising tool capable of predicting the environmental distribution of organic pollutants.

Introduction

The attention of many environmental scientists has been focused on monitoring and analysis of persistent organic pollutants (POPs) in various environmental compartments in the last decades. Information on the POP concentrations in soils, however, originated mostly from the short-term surveys focused on specific sources, whereas long-term data from the large-scale monitoring networks were scarce.

* Corresponding author phone: +420 549495149; fax: +420 549492840; e-mail: klanova@recetox.muni.cz.

[†] RECETOX.

[‡] Department of Botany and Zoology, Masaryk University.

[§] Institute of Biostatistics and Analyses, Masaryk University.

To overcome this lack of monitoring data and predict the soil concentrations of selected pollutants, a number of modeling techniques has been applied ranging from the standard multivariate techniques such as principal component analysis (*1, 2*) to spatial analyses such as kriging. A kriging approach has been frequently used for prediction of heavy metals in soils (*3*−*5*) but the application of spatial statistical analysis to POPs was far less common. It was, for example, applied to analysis of sediments (*6*) or HCH residues in topsoils (*7*). The spatial methods for prediction of POP concentrations in soils based on geographical information system (GIS) have been further developed with increasing amount of available soil concentration data. The GIS derived model of POP distribution on the European scale was, for instance, introduced by Pistocchi (*8*).

There are two different ways to predict spatial pattern: deterministic and stochastic approach, both with some advantages and disadvantages. Generally, deterministic models (e.g., box models) require a detailed description of the chemical properties of compounds (such as K_{ow}) and their distribution processes in the environment to predict a spatial distribution of specific POPs. On the contrary, stochastic models are based on the statistical analysis of available monitoring data in combination with environmental parameters of the sampling sites.

In the study presented here, a nonparametric regression technique (CART: classification and regression trees) (*9*) was applied for prediction of background concentrations of POPs in soils of the Czech Republic, and at the same time, the most important environmental parameters controlling POP distribution in the environment were identified. Such data mining approaches have been used before for the spatial prediction of heavy metal concentrations (*10, 11*) but they are still quite rare. These methods seem to be very promising as they provide certain advantages when compared to the classical statistical methods. They are more robust, and also less demanding when it comes to distribution of variables compared to other parametric regression techniques.

Material and Methods

Soil Data Set. The POP concentrations in soils of the Czech Republic were collected from several projects. The most important data sources were two nation-wide soil monitoring systems: Basal monitoring of agricultural soils conducted by the Central Institute for Supervising and Testing in Agriculture (CISTA) and Basal monitoring of soils in protected areas conducted by the Agency for Nature Conservation and Landscape Protection (ANCLP CR). A database was complete with data from several short- and long-term projects conducted by the Research Center for Environmental Chemistry and Ecotoxicology of Masaryk University at various spatial and time scales.

Although the sampling procedures slightly differed among the monitoring projects, the main principles were comparable and allowed for compilation of the joint database. The analytical methods were consistent throughout the data set and were described earlier (*12*). Only samples collected in the 2006−2007 period were processed using data on the top layer of soil, i.e., plow layer of arable land, top 20 cm of grassland soil, and top 20 cm of forest soil including the overlaying organic horizon. All the sites with incomplete measurements were excluded from the processing. Few samples available from urban and industrial sites were excluded as well since the main focus of all three projects was on background soils. A half of the value of detection limit was substituted whenever soil concentration was under

10.1021/es902076y CCC: $40.75 Published on Web 11/05/2009

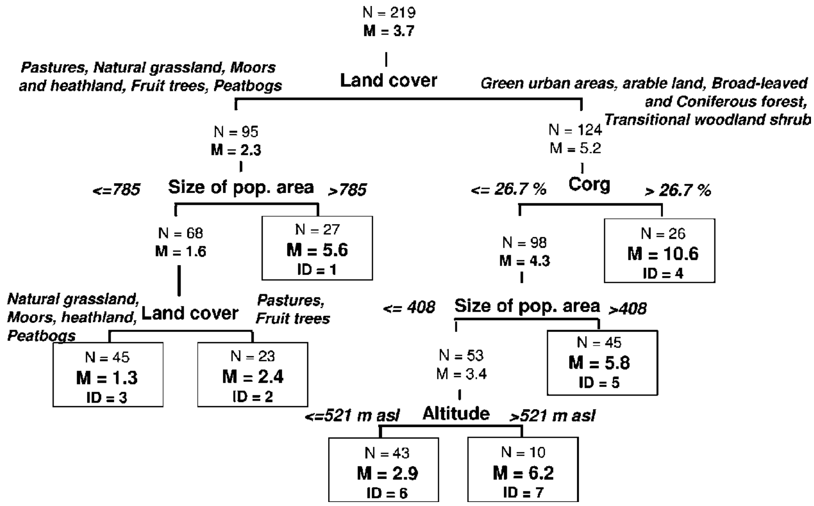

FIGURE 1. Regression tree for PCBs (N is the number of sampling sites; M is the geometric mean of concentration (ng g^{-1}); ID is the terminal node identifier (sequence number).

the limit of detection. This was 0.01 ng g^{-1} for organochlorines, 0.05 ng g^{-1} and 0.1 ng g^{-1} for lighter and heavier PAHs, respectively.

Geographical coordinates of the sampling sites, land use, soil type, and soil organic carbon content were determined at the individual sites beside the soil concentrations of POPs. Other parameters were derived using GIS. A map of the sampling sites is provided in Supporting Information (SI) Figure S2.

Predictors in the Model. A set of parameters with a potential to affect the environmental distribution of POPs was carefully selected. Four groups of predictor variables were utilized in the model: (i) markers of anthropogenic activities such as distances from industries, populated areas and roads, concentrations of NO_x and particulate matter in ambient air, or risk levels associated with the old ecological burdens (sites with historical contamination from unknown or discontinued sources); (ii) climatic factors such as mean annual temperature or precipitation; (iii) soil properties such as soil type (*13*) or content of organic carbon (C_{org}); (iv) land cover (*14*). (See details on the GIS predictors in SI).

A square grid was generated for the Czech Republic using ArcGIS 9.2 (80 033 squares of 1 × 1 km in total) prior to modeling, and environmental predictors were assigned to each square. In addition to variables determined directly in the field (land cover, soil type, and organic carbon content), detailed description of the sampling sites was completed with a set of parameters derived from the GIS.

Regression Analysis. Models for prediction of the POP concentrations were developed using CART (*9*) (More information can be found in SI). Regression tree was created in a process called binary recursive partitioning which is an iterative hierarchic process of splitting data into increasingly homogeneous subgroups. Tree pruning and 10-fold cross validation was applied in this study for estimation of the best tree size and stability of the model. A parameter of explained variability (Rv) was used to evaluate prediction capability of the tree (see SI). As each split of the tree was performed hierarchically on independent subset (node), multicolinearity of variables did not cause the serious problems known from the parametric regression models.

Results and Discussion

Distribution details of regression trees for selected groups of compounds can be seen in Figures 1−4.

Polychlorinated Biphenyls. Polychlorinated biphenyls (PCBs) are the chemicals of anthropogenic origin manufactured and heavily used in the former Czechoslovakia until the 1980s. High PCB concentrations can be still occasionally found in various environmental matrices in the Czech Republic, PCB levels in human breast milk are among the highest in Europe (*15, 16*).

A sum of five indicator congeners (PCB 101, 118, 153, 138, 180) in the soil samples collected from 219 sampling sites was used for the analysis. Regression tree consisted of 12 nodes including 7 terminal nodes (Figure 1). The most important splitting factor of the root node was the land cover. The right branch with higher PCB levels contained soils collected near roads and rail networks, soils from arable land, broad-leaved and coniferous forests, and green urban areas. The left branch containing soils from grasslands and natural pastures manifested two times lower PCB concentrations than the right one.

The next splitting factor of the left branch was the size of populated area. There was nearly four times higher PCB concentration in the heavily populated area terminal node (5.6 ng g^{-1}). The last two terminal nodes were separated by the land cover. Higher concentration (2.4 ng .g^{-1}) belonged to pastures and orchards (ID = 2). The lowest PCB concentration in the whole tree (1.3 ng g^{-1}) was observed in soils collected from natural grassland, moors and peatbogs (ID = 3).

The sampling sites in the right branch of the tree were split first according to C_{org}. Samples containing more than 26% of organic carbon were separated to terminal node 4 and they represented forest soils. Node 4 had the highest mean PCB concentration among all nodes (10.4 ng g^{-1}). The following splitting factor in this branch was the size of populated area. Larger settlements meant also higher concentrations (5.8 ng g^{-1}), this node contained mainly arable land. The altitude as the last splitting factor separated the sampling sites above 520 m above sea level (a.s.l.) with PCB concentrations two times higher (6.2 ng g^{-1}) from the sites below this altitude (2.9 ng g^{-1}). As the sampling sites at higher altitudes were mostly coniferous woods, a scavenging effect of the trees and litter deposition to the forest floor should be considered as important factors affecting PCB concentrations in soils. Climatic conditions (lower temperatures in mountain areas) can also play a significant role in partitioning of PCBs as they affect a deposition rate.

Polyaromatic Hydrocarbons. Polyaromatic hydrocarbons (PAHs) are typical byproducts of various combustion processes. Enhanced levels were identified at the sites close to residential and industrial areas. A sum of 12 PAHs (fluorene,

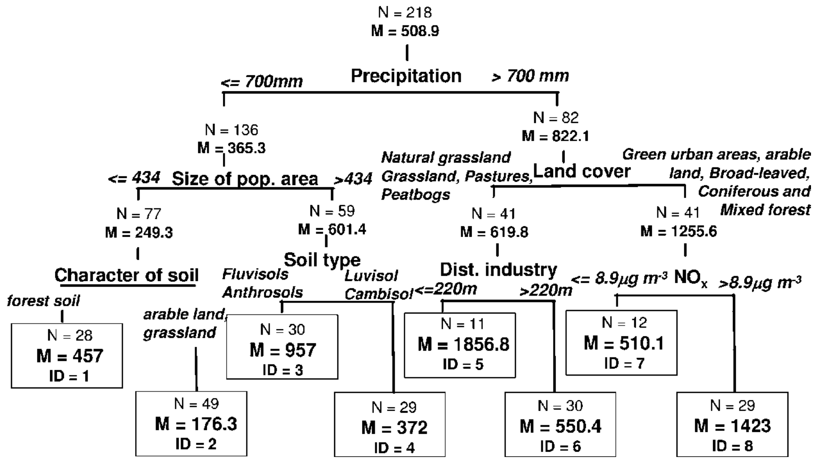

FIGURE 2. Regression tree for PAHs (N is the number of sampling sites; M is the geometric mean of concentration (ng g^{-1}); ID is the terminal node identifier (sequence number).

phenanthrene, fluoranthene, pyrene, benzo(a)anthracene, chrysene, benzo(b)fluoranthene, benzo(k)fluoranthene, benzo(a)pyrene, indeno(123 cd)pyrene, dibenzo(ah)anthracene, benzo(ghi)perylen) determined in each of 218 soil samples was used for development of prediction model. Only compounds highly (more than 0.9) correlated with each other as well as their sum were assessed. Most volatile PAHs were excluded because of their lower correlation (their measured levels were most probably affected by soil handling and sample preparation) as well as different behavior in the environment. An effect of anthropogenic activities can be observed in resulting PAH regression tree. Nevertheless, a splitting factor at the first level was precipitation (Figure 2). It is not surprising considering that PAHs are significantly associated with solid atmospheric particles for which a wet deposition is the most effective removal mechanism.

The left branch of the tree representing the sites with lower precipitation was further split according to the size of populated area and high PAH concentrations were related to high population density. Terminal nodes in the subsequent steps were defined by a character and type of soil. In heavily populated areas, the high PAH concentrations were found in Fluvisols (possibly as a result of contamination of alluvial soils by mobilized sediments during the floods (*17*)), and in Anthrosols affected directly by the anthropogenic activities (957 ng g^{-1}). On the contrary, the mean concentration of PAHs was almost three times lower in Cambisols and Luvisols.

A group of samples from less populated regions was further divided according to the soil character. The mean concentration of PAHs in forest soils was more than two times higher (457 ng g^{-1}) than the one in arable soils and grasslands (176.3 ng g^{-1}).

The right branch of the tree representing the sites with higher precipitation resulted in the terminal nodes with generally higher PAH concentrations. The land cover as a first decision factor separated the samples from pastures and natural grasslands from the samples collected in urban areas, coniferous woods and mixed forests. Grasslands were divided in the next step according to their distance from the industry. Node 5 with the highest PAH concentration in the entire study (1 856.8 ng g^{-1}) was characterized by distance from the industrial sources lower than 220 m. NO_x concentration in ambient air as an indicator of the anthropogenic activities was the splitting factor for the woods and parks. High PAH concentrations (1 423 ng g^{-1}) were observed in node 8 representing the sites with high NO_x levels. The terminal nodes 6 (grasslands far from industry) and 7 (woods with low NO_x concentrations) had very similar PAH concentrations in soils (550 and 510 ng g^{-1}, respectively).

PAH regression tree confirmed a crucial role of industry and other primary anthropogenic sources in the soil contamination process. However, it also identified precipitation as the main transfer route of atmospheric PAHs to soils.

DDTs. *p,p′*-DDT was widely used as organochlorine pesticide in the second half of the 20th century. It is very persistent in soils and not very volatile in temperate regions. It can be degraded—among others—to *p,p′*- DDE and *p,p′*-DDD. Soil concentrations of DDTs (sum of *p,p′*-DDT, DDD, and DDE) at 180 sampling sites were used for development of regression tree. The main decision factor was the soil character which separated grassland soils with DDT concentrations five times lower from remaining soils (Figure 3). It corresponds to the fact that DDT was preferably used for protection of the agricultural crops and forests.

In the group of agricultural, forest, and anthropogenic soils (right branch), the next decision factor was the soil type. Higher mean DDT level was found in Fluvisols, Luvisols, Arenosols, and Podzols (24.3 ng g^{-1}). Contamination of Fluvisols has been discussed earlier. As all remaining soil types are typical for the forests (Podzols also for the high altitudes), their contamination is probably related to former pesticide application in forests. In the group of Chernozems and Cambisols, the final splitting factor was a level of risk which separated sampling sites close to the old dumps (25.9 ng g^{-1}) from the sites more remote (8.4 ng g^{-1}). It seems that the sampling sites directly connected to former application, storage, and disposal of pesticides fell to this branch of the tree.

In the left branch (grassland sites), temperature and land cover were the next decision factors. Soils with the highest DDT concentrations in this branch (7.0 ng g^{-1}) can be found in node 1 and are characterized by higher temperatures. The altitude and size of populated area were surrogates for temperature in this case as higher temperature means the lowland and also the site closer to former application of DDT in agriculture. For the low temperature sites, the final decision factor was the land cover. It separated two nodes with very low DDT concentrations (0.34 ng g^{-1} and 1.2 ng g^{-1}), and, as expected, revealed somewhat higher concentrations in pastures when compared to natural grasslands.

Hexachlorobenzene. Regression tree of hexachlorobenzene (HCB) with eight terminal nodes (Figure 4) resulted from the statistical analysis of 192 soil samples. HCB production was terminated in former Czechoslovakia in 1968, and its application as a pesticide was banned in 1977 while 10 000 t of HCB per year was still produced in the 1970s worldwide (*18, 19*). This can explain mean concentration of HCB in arable land persisting at the level of 3.2 ng g^{-1} while

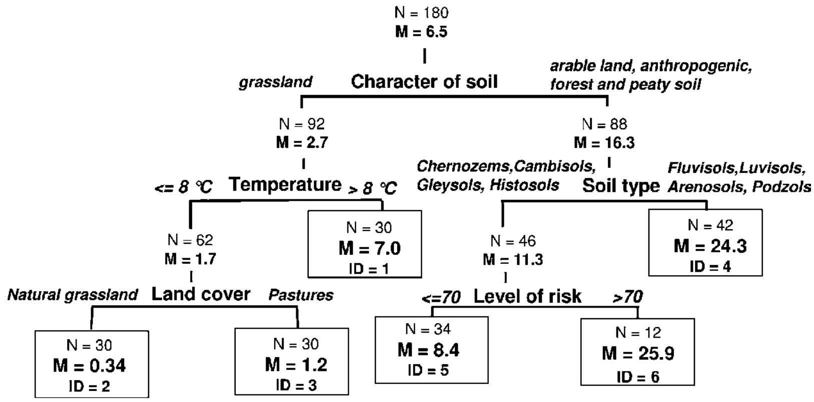

FIGURE 3. Regression tree for DDTs (N is the number of sampling sites; M is the geometric mean of concentration (ng g^{-1}); ID is the terminal node identifier (sequence number).

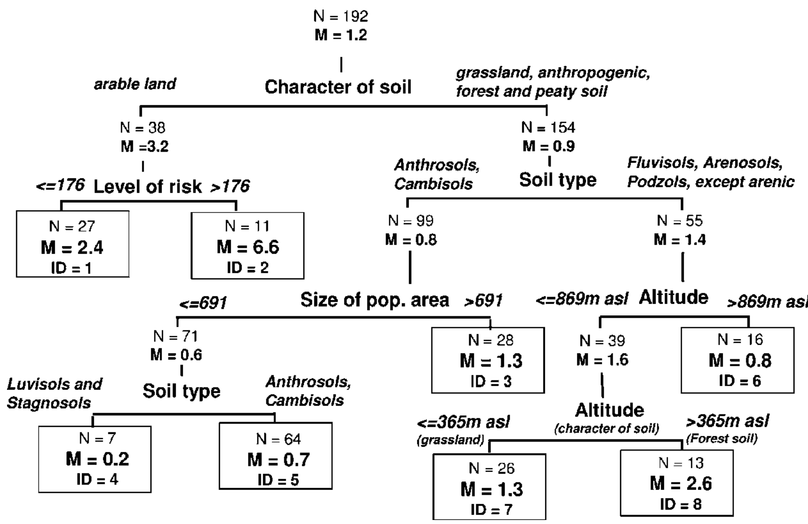

FIGURE 4. Regression tree for HCB (N is the number of sampling sites; M is the geometric mean of concentration (ng g^{-1}); ID is the terminal node identifier (sequence number).

those of other soils were generally lower. In the branch of arable soils, the level of risk (presence of the old ecological burden) was the next splitting parameter dividing the subgroup into two terminal nodes. While terminal node 1 corresponded to common agricultural soils (mean concentration of 2.4 ng g^{-1}), terminal node 2 showed the highest mean concentration of HCB (6.6 ng g^{-1}) and was represented by sampling sites in the vicinity of industrial areas and old dumping sites.

The right side of the tree was more connected to current sources of HCB (long-range transport and combustion). A soil type was the second most important predictor in this branch of the tree and it distinguished between Fluvisols, Arenosols and Podzols on the one hand, and Anthrosols and Cambisols on the other hand.

Accumulation of chemical substances in the alluvial areas during the floods can be responsible for elevated levels of HCB in Fluvisols, and the high C_{org} can explain the same in case of Arenosols and Podzols. This branch was split into three terminal nodes separated by the altitude. The highest mean concentration of HCB in this branch was found in the range of 365−869 m asl (2.6 ng g^{-1}), whereas the mean concentration was lower below 365 m asl (1.3 ng g^{-1}). The lowest observed concentrations (0.8 ng g^{-1}) were found in altitudes above 869 m a.s.l., in the low populated areas distant from industry. Long-range transport should be considered as the most important source of HCB to these regions. Such difference, however, should not be only explained by elevation since node 8 mostly contained forest soils. As described earlier, scavenging effect of vegetation and deposition of HCB in colder areas could be the processes responsible for these findings. Cambisol (left side of this branch) is a soil type covering more than a half of forested areas in the country. Slightly higher HCB concentrations were found in populated lowlands (1.3 ng g^{-1}). Remaining samples in this branch were separated to terminal nodes by the size of populated area, and subsequently by the soil type. Not surprisingly, HCB concentration found in the populated areas (1.3 ng g^{-1}) was higher when compared to the low populated areas which were further separated according to the soil type. Lower concentrations measured in Luvisols (0.2 ng g^{-1}) when compared to Cambisols (0.7 ng g^{-1}) can be explained by lower organic carbon content in Luvisol. However, the low number of samples from this category has to be considered.

Based on these results, two possible sources of HCB to soils can be identified in the Czech Republic: the old burdens connected to past application of HCB as a pesticide in arable soil (left side of the tree), and the present sources, especially industrial emissions and combustion processes.

Based on regression trees, an importance of available predictors for estimation of the soil concentrations was calculated for each of the analyzed groups of POPs (Figure 5).

Results of regression trees have to be interpreted with a great care since some variables can be surrogates of others. As mentioned earlier, a total importance of the individual variables can be calculated (Figure 5) estimating their influence on distribution of POPs in soils. For detail interpretation, however, the trees are still needed. An

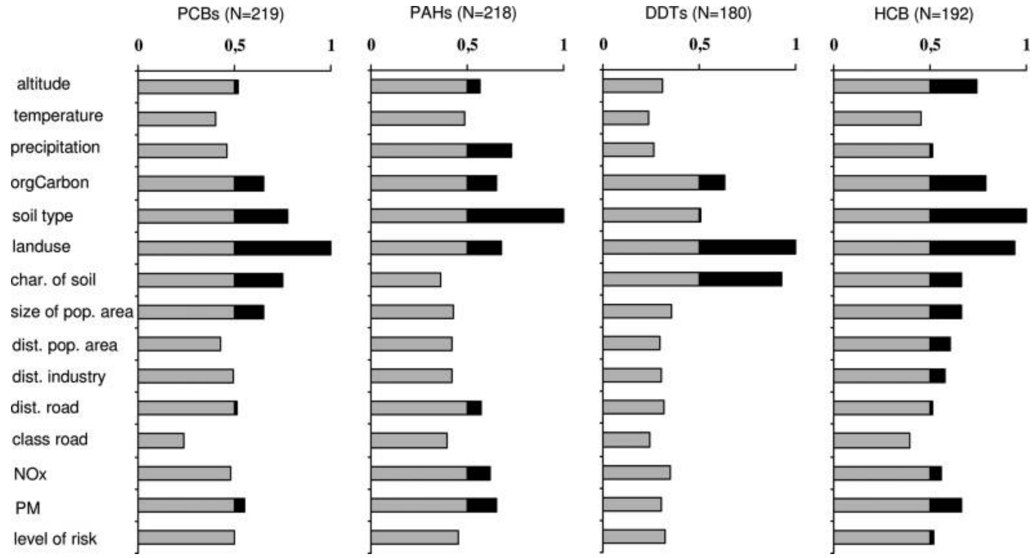

FIGURE 5. Importance of the individual predictors for estimation of the soil concentrations of selected groups of compounds calculated from regression trees (black color = importance ≥ 0.5).

assessment of impacts of predictors in the individual nodes indicated that some aspects are common to all substances.

Anthropogenic Activities. An impact of human activities could be detected in all regression trees but mechanisms differed for various groups of compounds. Higher DDT concentrations were linked directly to the areas of pesticide application and the old dumps. It applies partially to HCB as well but levels of HCB as the more volatile compound were also affected by long-range transport (*20, 21*). An impact of long-range transport of HCB from the primary (combustion) and secondary (soil volatilization) sources was observed in forest soils. Scavenging effect of coniferous trees with consequent litter fall (*22*), vertical and horizontal deposition were responsible for the air−soil transport of HCB, whereas high C_{org} in forest soils together with a strong affinity of HCB to soil organic carbon (*23, 24*) resulted in retention of HCB in this matrix.

To certain extent, these processes affected also the environmental distribution of other substances as PCBs and PAHs. Direct effects of combustion and traffic were observed for PAHs; short distance to industrial activities as well as higher precipitation level was good predictors of elevated pollution.

An impact of the population density was confirmed for all investigated chemicals with the exception of DDTs. It is in a good agreement with numerous studies presenting decreasing POP concentrations with increasing distance from the populated areas (*25*−*28*). Some authors refer to time dependence of the air concentrations affected by a distance from the city, and temperature changes as a "urban pulse" (*29*) or "hallo effect" (*30*).

Soil Organic Carbon Content, Soil Type and Character. The amount of soil organic carbon is another important factor affecting distribution of POPs. In this study, it was an important splitting factor in PCB tree separating samples with C_{org} higher than 26.7% to a terminal node containing soils with the highest PCB concentrations which is in agreement with previously published studies (*31*).

It has to be considered that other predictors such as a soil type, character and land cover could be surrogates for C_{org}. The effects of the land use, vegetation scavenging and affinity to C_{org} are parallel and in many cases not easy to distinguish. There was a typical split in all regression trees connected to soil types. Fluvisols separated alluvial areas to nodes with high concentrations of POPs. Physicochemical properties and origin of various groups of pollutants also play an important role in the separation process. PAHs, for instance, are unintentional byproduct of various combustion processes. HCB is partially of the same origin but it has also been produced and applied as a pesticide. It can be hypothesized that Fluvisols are separated from anthrosols in the HCB tree because they are more affected by agricultural application of HCB, whereas contamination of Anthrosols originates in industrial activities. In the PAH tree, however, they stay together. The role of Cambisols was, however, hard to distinguish since they cover more than a half of the area of both forests and agricultural soils.

Geographic and Climatic Conditions. Geographical predictors allowed for the alternative interpretations because they often acted as surrogates of climatic or anthropogenic predictors. In HCB and DDT trees, the altitude was connected to increasing distance from populated areas, in PCB tree to the forests growing at higher altitudes. In case of PAHs, precipitation was the most influential factor for soil concentrations as it controls deposition of heavy PAHs on soil surfaces.

The maps (resolution of 1 × 1 km) predicting a spatial distribution of selected POPs in soils of the Czech Republic were generated based on regression trees (Figure 6).

The highest soil concentrations of PCBs were predicted in the mountain areas along the borderline of the Czech Republic and in the central highlands. It has been published earlier that the highest PCB concentrations were measured in urban and industrial areas but such regions were not included in this study and are shown as the white spots in the maps. Elevated PCB concentrations in the mountain soils were the result of specific land cover (coniferous forests) and soil type (high organic carbon) (*31*).

Mountain forest soils were predicted to have the highest concentrations of PAH as well. In this case, however, there was a great difference between the northern and southern parts of the country. While the southern regions are mostly rural, low populated areas with little industry, majority of industry is concentrated to the north of the country. This area between the Czech Republic, Germany and Poland called the "black triangle" is heavily affected by the industrial activities from all three countries (*32*). Thus, the northern mountains scavenge the atmospheric pollution from the large industrial region. The red protrusion reaching from the northern border inland marks the area with the highest annual mean precipitation in the Czech Republic.

In agreement with the previously published data (*31*), HCB and DDT maps showed higher concentrations in lowlands when compared to mountain soils. Red color

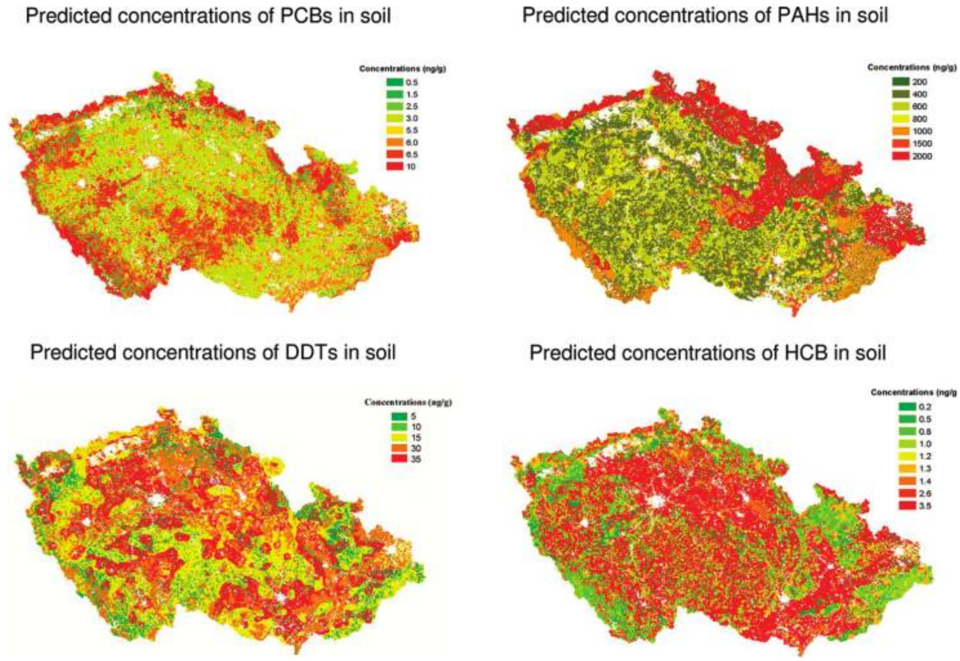

FIGURE 6. The maps predicting POP concentrations in soils. The white spots represent areas with the lack of soil samples (large urban areas, water bodies, and traffic networks).

marks the areas of former application of these substances in agriculture and can be found especially along the major rivers (alluvial soils). HCB distribution in soils was more uniform than the one of DDTs. It can be explained by higher volatility and long-range transport potential of HCB.

The stochastic model applied in this study seems to be a promising tool capable of predicting the environmental distribution of organic pollutants. The selected predictors addressed specific conditions affecting a behavior of the individual groups of pollutants: presence of primary and secondary sources, density of human settlement, geographical characteristics and climatic conditions, land use, land cover, and soil properties.

A percentage of explained variability as well as an overall fit of the model could be determined due to the fact that the model was based on consistent and representative data set, and consequently tested by cross validation.

The models resulting from the above-described analyses explained a high portion of variability in relationship between soil concentrations of selected POPs and available predictors. HCB tree was the most successful with 76.2% (71.9−80.5%) of explained variability, followed by trees for PAHs (71%; 66.8−78.2%), PCBs (67.6%; 62.9−71.1%), and DDTs (65.4%; 56.4−67.5%). Moreover, the models showed a high rate of correct predictions for unknown cases during the cross validation.

This is an advantage when compared to deterministic distribution models (i.e., box models) as they suffer higher uncertainties associated to the testing procedures (selection of input parameters, description of physical processes, or interactions among variables). An application of regression trees made it possible to assess impacts of the individual parameters on the environmental distribution of POPs: to distinguish between the effects of current sources of POPs and the old burdens, for instance. The results of the model, however, are not always explicit and have to be interpreted with a great care. The applicability of this stochastic model in other regions is limited to the areas with similar geographical and climatic conditions and historical usage of these compounds. This potential has to be further explored.

Acknowledgments

This project was supported by the Czech Ministry of Education, Youth and Sport (MSM 0021622412), and EC (7FP project isoSOIL).

Supporting Information Available

All predictors are listed and the statistical methods explained. Concentration ranges in the individual terminal nodes are provided in Figure S1, a map of the sampling sites in Figure S2. This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Caia, Q.; Mob, C.; Wua, Q.; Katsoyiannisc, A.; Zenga, Q. The status of soil contamination by semivolatile organic chemicals (SVOCs) in China: A review. *Sci. Total Environ.* **2008**, *389*, 209–224.

(2) Skrbic, B.; Durisic-Mladenovic, N. Principal component analysis for soil contamination with organochlorine compounds. *Chemosphere* **2007**, *68*, 2144–2152.

(3) Cattle, J. A.; McBratney, A. B.; Minasny, B. Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. *J. Environ. Qual.* **2002**, *31*, 1576–1588.

(4) McGrath, D.; Zhang, C.; Cartona, O. T. Geostatistical analyses and hazard assessment on soil lead in Silvermines area, Ireland. *Environ. Pollut.* **2004**, *127*, 239–248.

(5) Zhang, C. Using multivariate analyses and GIS to identify pollutants and their spatial patterns in urban soils in Galway, Ireland. *Environ. Pollut.* **2006**, *142*, 501–511.

(6) Ouyang, Y.; Nkedi-Kizza, P.; Mansell, R. S.; Ren, J. Y. Spatial distribution of DDT in sediments from estuarine rivers of central Florida. *J. Environ. Qual.* **2003**, *32*, 1710–1716.

(7) Li, B. G.; Cao, J.; Liu, W. X.; Shen, W. R.; Wang, X. J.; Tao, S. Geostatistical analysis and kriging of hexachlorocyclohexane residues in topsoil from Tianjin, China. *Environ. Pollut.* **2006**, *142*, 567–575.

(8) Pistocchi, A. A GIS-based approach for modeling the fate and transport of pollutants in europe. *Environ. Sci. Technol.* **2008**, *42*, 3640–3647.

(9) Breiman, L.; Friedman, J.; Stone, C. J.; Olshen, R. A. *Classification and Regression Trees*; Chapman and Hall: New York, 1984.

(10) Pesch, R.; Schröder, W. Integrative exposure assessment through classification and regression trees on bioaccumulation of metals, related sampling site characteristics and ecoregions. *Ecol. Inf.* **2006**, *1*, 55–65.
(11) Zhang, X.; Lin, F.; Jiang, Y.; Wang, K.; Wong, M. T. F.; Zhang, X.; et al. Assessing soil Cu content and anthropogenic influences using decision tree analysis. *Environ. Pollut.* **2008**, *156*, 1260–1267.
(12) Holoubek, I.; Klánová, J.; Jarkovský, J.; Kubík, V.; Helešic, J. Trends in background levels of persistent organic pollutants at Kosetice observatory, Czech Republic. Part II. Aquatic and terrestric environments 1988−2005. *J. Environ. Monit.* **2007**, *9*, 564–571.
(13) World Reference Base for Soil Resources 2006. In *World Soil Resources Reports No. 103*; FAO: Rome, 2006.
(14) Bossard, M.; Feranec, J.; Otahel, J. *CORINE Land Cover Technical Guide—Addendum 2000*, Technical Report No. 40; European Environment Agency: Copenhagen Denmark, 2000; Available at http://www.eea.eu.int.
(15) Schoula, R.; Hajšlová, J.; Bencko, V.; Poustka, J.; Holadová, K.; Vízek, V. Occurrence of persistent organochlorine contaminants in human milk collected in several regions of Czech Republic. *Chemosphere* **1996**, *33*, 1485–1494.
(16) Černá, M.; Spěváčková, V.; Batáriová, A.; Šmíd, J.; Čejchanová, M.; Očadlíková, D.; Bavorová, H.; Beneš, B.; Kubínová, R. Human biomonitoring system in the Czech Republic. *Int. J. Hyg. Environ. Health* **2007**, *210*, 495–499.
(17) Podlešáková, E.; Nìmeèek, J.; Hálová, G. Burden of Alluvial soils by risk substances. *Rostlinná výroba* **1994**, *40*, 69–80.
(18) Jones, K. *Hexachlorbenzene—Sources, Environmental Fate and Risk Characterization*; Eurochlor Publication: Brussels, 2005.
(19) Barber, J. L.; Sweetman, A. J.; Wijk, D.; W.; Jones, K. C. Hexachlorbenzene in the global environment: Emissions, levels, distribution, trends and processes. *Sci. Total Environ.* **2005**, *349*, 1–44.
(20) Wania, F.; Mackay, D. A global distribution model for persistent organic chemicals. *Sci. Total Environ.* **1995**, *160/161*, 211–232.
(21) Wania, F.; Mackay, D. Tracking the distribution of persistent organic pollutants. *Environ. Sci. Technol.* **1996**, *30*, 390A–6A.
(22) Klánová, J.; Čupr, P.; Baráková, D.; Šeda, Z.; Andìl, P.; Holoubek, I. Can pine needles indicate trends in the air pollution levels at remote sites? *Environ. Pollut.* **2009**, DOI: 10.1016/j.envpol/2009.15.030.
(23) Mackay, D. *Multimedia Environmental Models: The Fugacity Approach*, 2nd ed.; Lewis Publishers: New York, 2001.
(24) Beurskens, J. E. M.; de Wolf, J.; Van den Heuvel, H. High levels of chlorinated aromatic compounds in deep Rhine sediments with special reference to microbial transformations. *Land. Degrad. Rehabil.* **1993**, *4*, 367–71.
(25) Wyrzykowska, B.; Hanari, N.; Orlikowska, A.; Bochentin, I.; Rostkowski, P.; Falandysz, J.; Taniyasu, S.; Horii, Y.; Jiang, Q.; Yamashita, N. Polychlorinated biphenyls and -naphthalenes in pine needles and soil from Poland—Concentrations and patterns in view of long-term environmental monitoring. *Chemosphere* **2007**, *67*, 1877–1886.
(26) Ren, N.; Que, M.; Li, Y.; Liu, Y.; Wan, X.; Xu, D.; Sverko, E.; Ma, J. Polychlorinated biphenyls in chinese surface soils. *Environ. Sci. Technol.* **2007**, *41*, 3871–3876.
(27) Chrysikou, L.; Gemenetzis, P.; Kouras, A.; Manoli, E.; Terzi, E.; Samara, C. Distribution of persistent organic pollutants, polycyclic aromatic hydrocarbons and trace elements in soil and vegetation following a large scale landfill fire in northern Greece. *Environ Int.* **2008**, *34*, 210–22.
(28) Ray, S.; Khillare, P. S.; Agarwal, T.; Shridhar, V. Assessment of PAHs in soil around the International Airport in Delhi, India. *J. Hazard. Mater.* **2008**, *156*, 9–16.
(29) Jamshidi, A.; Hunter, S.; Harzati, S.; Harrad, S. Concentrations and chiral signatures of polychlorinated biphenyls in outdoor and indoor air and soil in a major U.K. conurbation. *Environ. Sci. Technol.* **2007**, *41*, 2153–2158.
(30) Stow, J. P.; Sovab, J.; Reimer, K. J. The relative influence of distant and local (DEW-line) PCB sources in the Canadian Arctic. *Sci. Total Environ.* **2005**, *342*, 107–118.
(31) Holoubek, I.; Dušek, L.; Sáňka, M.; Hofman, J.; Čupr, P.; Jarkovský, J.; Zbíral, J.; Klánová, J. Soil burdens of persistent organic pollutants—Their levels, fate and risk. Part 1. Variation of concentration ranges according to different soil uses and locations. *Environ. Pollut.* **2009**, DOI: 10.1016/j.envpol/2009.15.031.
(32) Kukučka, P.; Klánová, J.; Sáňka, M.; Holoubek, I. Soil burdens of persistent organic pollutants - Their levels, fate and risk. Part II. Are there any trends in PCDD/F levels in mountain soils? *Environ. Pollut.* **2009**, DOI: 10.1016/j.envpol/2009.15.029.

ES902076Y