# REFINING OF ASTHMA PREVALENCE SPATIAL DISTRIBUTION AND VISUALIZATION OF OUTDOOR ENVIRONMENT FACTORS USING GIS AND ITS APPLICATION FOR IDENTIFICATION OF MUTUAL ASSOCIATIONS

**Pavla Škarková[1], Radek Kadlubiec[1], Michal Fischer[1], Jana Kratěnová[2], Miloš Zapletal[1, 3], Jiří Vrubel[1]**

[1]EKOTOXA Ltd., Opava, Czech Republic
[2]National Institute of Public Health, Prague, Czech Republic
[3]Faculty of Philosophy and Science, Silesian University in Opava, Opava, Czech Republic

## SUMMARY

*Aim:* This study presents a procedure of complex assessment of the environment impact on asthma prevalence. This approach is also applicable for any other disease which is supposed to be associated with the quality of the outdoor environment.

*Methods:* The input data included asthma prevalence values from the National Institute of Public Health (NIPH) cross-section questionnaire survey (13,456 children) and annual reports on activities of all paediatricians in the Czech Republic (2,072 surgeries); concentrations of $PM_{10}$, $PM_{2.5}$, $NO_2$, $SO_2$, $O_3$, benzene, benzo(a)pyrene, As, Cd, Pb and Ni; emissions of total suspended particles, $SO_2$, $NO_x$, CO, VOC, $NH_3$; traffic intensity; land cover (anthropogenic area, urban greenery, arable land, grassland, other agricultural land, forests); proportion of cultivation of individual agricultural crops (17 categories); and proportion of individual woods (15 categories). Using the Geographical Information Systems (GIS) analysis the integration of all source data through one spatial unit was achieved and complete data sets were compiled to be subjected to statistical analysis (combination of factor analysis with logistic/linear regression).

*Results:* In this study, the approach of combined use of GIS analyses and statistical evaluation of large input data sets was tested. The asthma prevalence demonstrated positive associations with the air pollution ($PM_{10}$, $PM_{2.5}$, benzene, benzo(a)pyren, $SO_2$, Pb, Cd) and the rate of agricultural use of land (growing oats, rye, arable fodder crops). Conversely, there was a negative association with the occurrence of natural forests (ash, poplar, fir, beech, spruce, pine). No significant associations were observed with the distance from the road, traffic intensity and $NO_2$ concentration.

*Conclusions:* These findings suggest that the outdoor quality may be one of the crucial factors for asthma prevalence.

*Key words:* asthma, children, GIS, environmental factors, air pollution, land cover, outdoor

**Address for correspondence:** P. Škarková, EKOTOXA Ltd., Otická 761/37, 746 01 Opava, Czech Republic. E-mail: pavla.skarkova@ekotoxa.cz

## INTRODUCTION

Asthma is a heterogeneous disease, usually characterized by chronic airway inflammation. It is defined by the history of respiratory symptoms such as wheeze, shortness of breath, chest tightness and cough that vary over time and in intensity, together with variable expiratory airflow limitation. 1–18% of the populations in different countries suffer from this chronic respiratory disease. It is generally believed that a disease inception and persistence is driven by gene-environment interactions. The most important of these interactions may occur in early life and even *in-utero*. Multiple environmental factors, both biological and sociological, may be important in the development of asthma (1).

In the Czech Republic, numbers of registered asthmatics progressively increase especially in children, as annually monitored by the Institute of Health Information and Statistics of the Czech Republic (IHIS). 1.94% of asthmatic children (aged 0–14 years) and 1.68% of asthmatics among youth (aged 15–19 years) were registered by general practitioners for children and youth in the Czech Republic in 2000. In 2012, the proportion increased to 4.28% in children and 7.29% in youth. The highest values were reported from Hradec Kralové and Karlovy Vary Regions, the lowest from Zlín Region (2, 3). A prevalence study of the National Institute of Public Health (NIPH) presents asthma occurrence in children aged 5, 9, 13 in the year 1996 as 3.9%, in 2001 as 5.1%, in 2006 as 8.2%, and in 2012 as 9.6% (4). Higher asthma occurrence is generally thought to be related to the deterioration of environmental conditions, as proved by a number of studies. The most often discussed hazard factors are air pollutants emitted from the traffic and industry. There are connections between asthma occurrence and particulate matter less than 10 μm in aerodynamic diameter ($PM_{10}$) and smaller (5–10), $NO_2$ (5, 11–14) and $SO_2$ (6). Generally, the quality of the environment is also affected by other parameters like land use and agricultural production (15, 16).

A very useful tool for the assessment of connections between disease occurrence and quality of the environment is the Geo-

graphical Information Systems (GIS). GIS is used for health data processing, analysis of geographical distribution and variation of diseases mapping, monitoring and management of health epidemics. In addition, this system makes it possible to combine the spatial localization of the monitored disease and layers with information about the quality of the environment (7, 10, 14, 16–19).

In this paper, we present an application of GIS analyses for refining asthma prevalence spatial distribution in the Czech Republic and for visualization of environmental factors which can be connected with the disease occurrence. All data were converted into the grid $1 \times 1$ km and joined with the grid number as a unique indicator. The result of data integration was a data set which contains the information about disease prevalence and all monitored characteristics of the outdoor environment for each spatial unit (grid). This data set served for statistical evaluation of associations between asthma prevalence and particular factors of the environment.

## MATERIAL AND METHODS

### Asthma Prevalence Data

Data from NIPH cross-sectional questionnaires as well as data annually reported by all general practitioners for children and youth and gathered by IHIS were processed for this study. The asthma occurrence was visualized based on the child's residence or paediatrician surgery address.

The NIPH cross-section questionnaire survey was performed in five-year intervals in the framework of the Environmental Health Monitoring System in the Czech Republic, when anamnesis (including the information about asthma occurrence) of a representative sample of children aged 5, 9, 13 and 17 was obtained from the surgeries of selected paediatricians. According to the methodology of given survey, respondents were localized only by the address of responsible nursing paediatricians in 2006 and 2011, and in the year 2012, the collected data was also supplemented by the child's residence address for a further specification (Table 1).

Annual reports on activities of general practitioners for children and youth in the Czech Republic for the year 2010 filled out for the purposes of IHIS were used. It comprised 2,072 surgeries (all general practitioners in CR). The proportion of asthmatics registered at a given physician was calculated from the total number of registered patients.

The spatial presentation of IHIS data was made via surgeries localization as mean asthma prevalence values in municipalities. The spatial interpolation of these data was used to create a surface in the extent of the Czech Republic. A municipality was defined as an elementary spatial unit for which a mean value of asthma prevalence was calculated. Municipalities with at least one paediatrician were included into the interpolation process. The applied interpolation method was Empirical Bayesian Kriging (EBK) which represents an advanced level of classical kriging geostatistics used for describing spatial patterns and interpolation values for locations where no samples were available. It automates building a valid kriging model via calculating parameters through a process of subsetting and simulations. EBK accounts for the error introduced by estimating the underlying semivariogram. Other kriging methods calculate the semivariogram from known data locations and use this single semivariogram to make predictions at unknown locations. This process implicitly assumes that the estimated semivariogram is the true semivariogram for the interpolation region. By not taking the uncertainty of semivariogram estimation into account, other kriging methods underestimate the standard errors of prediction (20, 21).

### Outdoor Environment Factors

#### Air Pollution

Five-year average air pollution concentrations in the period of 2008–2012 published by the Czech Hydrometeorological Institute (CHMI) (22) were used for description of the air quality. These are model values of air pollution concentrations of $PM_{10}$ and $PM_{2.5}$, $NO_2$, $SO_2$, $O_3$, benzene, benzo(a)pyrene, arsenic, cadmium, lead, and nickel for the Czech Republic with the resolution of $1 \times 1$ km square.

Since square grids of Ekotoxa and CHMI are in different coordinate systems, the data had to be converted by spatial method of the nearest distance before further processing.

#### Emissions

Information on amounts of pollutants emitted in a given spatial unit was acquired from the Register of Emissions and Air Pollution Sources (REZZO), administered by CHMI. Emission data of large stationary sources (REZZO 1), medium stationary sources (REZZO 2) and small stationary sources (REZZO 3) for 2010 were processed (Table 2).

**Table 1.** *Characteristics of NIPH questionnaire survey*

| Year | Number of municipalities | Number of paediatricians | Number of children | Localization |
|------|--------------------------|--------------------------|--------------------|--------------|
| 2006 | 18 | 61 | 6,953 | Paediatrician surgery address |
| 2011 | 10 | 31 | 3,554 | Paediatrician surgery address |
| 2012 | 10 | 28 | 2,949 | Residence address of respondent |

**Table 2.** *Characteristics of emission data*

| Year | Type of source | Pollutants | Number | Localization |
|------|----------------|------------|--------|--------------|
| 2010 | REZZO 1–2 (large and medium sources) | TSP, $SO_2$, $NO_x$, CO, VOC, $NH_3$ | 56,522 | Source coordinates |
| 2010 | REZZO 3 (small sources) | TSP, $SO_2$, $NO_x$, CO | 6,384 | Municipality |

REZZO – Register of Emissions and Air Pollution Sources
TSP – Total Suspended Particles
VOC – Volatile Organic Compounds

The sources REZZO 1 and 2 are point elements for which affiliation to the square $1 \times 1$ km was determined by spatial analyses. The final emission value for individual squares was set as a sum of values of all sources located in the square. Data of REZZO 3 sources are provided at the level of the municipality so they are of spatial character. The emission value of REZZO 3 was determined for the square $1 \times 1$ km based on the spatial location of square centre in the relevant municipality.

### Traffic Intensity

Data source for assessment of traffic intensity was the nationwide traffic counts on all motorways, roads of categories I and II and on selected roads of category III of the Czech Republic in 2010 provided by the Road and Motorway Directorate (RMD) of the Czech Republic. Linear data of traffic counts 2010 were converted into the square grid $1 \times 1$ km. In the first phase of conversion, the linear data were divided by $1 \times 1$ km grid. The calculation of length (in km) was completed for such newly formed counting segments. In the next phase the length of segment was used as weight for a traffic intensity calculation in the given square. The result is a value of the sum of passing vehicles in 24-hours.km$^{-2}$ for each square. For example in $1 \times 1$ km square which contains the only one 0.5 km length road with traffic intensity of 100 vehicles the final number is 50 vehicles 24-hours.km$^{-2}$.

### Land Cover

The source of data for land cover analyses was the Fundamental Base of Geographic Data (ZABAGED®) administered by the Czech Office for Surveying, Mapping and Cadastre. It is an unified seamless geographic database of land cover in the extent of the Czech Republic in the scale of 1 : 10,000. Percentage representation of individual land cover categories was determined by spatial analyses of geographical layer intersections in the squares of $1 \times 1$ km. Total of 25 categories of different types of land cover was processed and aggregated into 6 key categories (anthropogenic area, urban greenery, arable land, grassland, other agricultural land, and forests).

### Sowing

Data source was a database of cultivated crops of the Czech Statistical Office Agrocenzus 2010. In total there are 53 categories of cultivated crops in differentiation for individual districts of the Czech Republic (76 districts). The number of categories was reduced to 17 the most important ones. The recalculation for percentage representation of cultivation of individual agricultural crops in the district area was made for each district. The final values of relevant district, where the monitored respondent or physician's surgery is located, were entered into the matrix.

### Species Composition of Forests

Data source was the Institute for Economic Forest Adjustment and its databases of species composition. In total there are 83 categories of wood species in spatial accuracy of municipalities with extended powers (ORP) (206 ORP). The number of categories was reduced to 15 the most important ones. The recalculation for percentage representation of individual woods in the area of ORP was made. The final values of relevant ORP, where the monitored respondent or physician's surgery is located, were entered into the matrix.

### Statistical Evaluation

All source data were converted into the grid $1 \times 1$ km and interconnected with the grid number. The result of data integration was a data set which contains the information about disease prevalence and all monitored characteristics of the outdoor environment (64 characteristics) for each spatial unit (grid). This data set served for statistical evaluation of the following models:
NIPH data – localization on the child's residence address (Model 1)
NIPH data – localization on the paediatrician's address (Model 2)
IHIS data – localization on the paediatrician's address (Model 3).

Logistic regression was used for evaluation of the Model 1 and 2 data sets because the dependent variable was binary (presence/absence of asthma in a particular child). The multivariate linear regression for the Model 3 data set was used because the dependent variable was already fully quantified (the number of asthmatics at paediatricians). In all cases there was at first a linear independence of individual variables of the outdoor environment and their transformations into the independent factors secured by factor analysis (Principal Component Analysis using a varimax rotation), using the artificially created new vectors of variables – latent variables (LV).

The subsequent evaluation by logistic/linear regression was already performed with the latent variables. Influences of original observable variables were retrospectively derived from results of the regression made on the latent variables.

The results of logistic regression (Models 1 and 2) are presented as odds ratios (OR), the results of linear regression are presented as regression coefficients (b), which express how one unit of the given (observable) variable affects the percentage of asthma occurrence. For the observable variables odds ratios (OR) were calculated (23). Adjusted regression coefficients were also calculated as products of correlation coefficients and regression coefficients of the corresponding latent variables. 5% statistical significance level was considered in all statistical analyses. The evaluation was made by the statistical software UNISTAT® 6.0.

### RESULTS

The source data of different precision and formats were converted into a single form using geoinformation technologies to make possible to compile a complete set of the information on the outdoor environment and asthma prevalence for every grid by integrating individual layers. A high number of the layers was processed for each category of environmental data (e.g. 11 pollutants for air pollution, 17 crops for sowing). In Figures 1–8, one representative for each category is visualized only.

The statistical significance of individual regression models was tested as a whole. The Models 1 and 2 were statistically significant ($p < 0.001$). Model 3 does not meet the criterion of statistical significance ($p = 0.197$), but it may be still taken as a relatively acceptable value and, considering this fact, it is possible to continue with interpretation of individual variables within the regression model.

In the statistical evaluation of data for Model 1, 15 latent variables were obtained by factor analysis from 64 original observable variables. These parameters, which are mutually orthogonal, were then entered into a logistic regression as independent variables. The dependent variable was information about the presence/
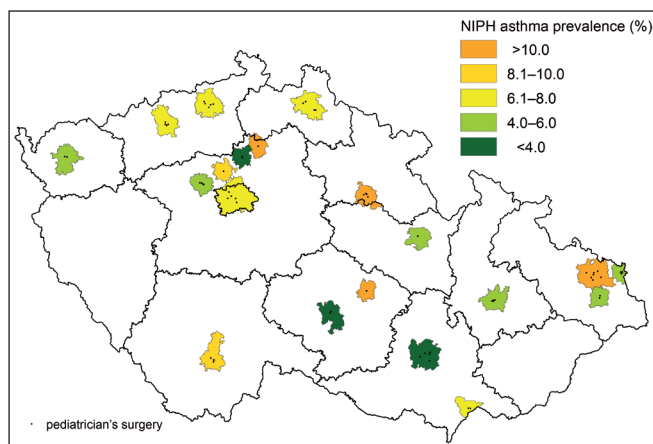
*Fig. 1.* Mean asthma prevalence in children from the catchment areas of paediatricians in municipalities (based on the National Institute of Public Health data for 2006–2012).
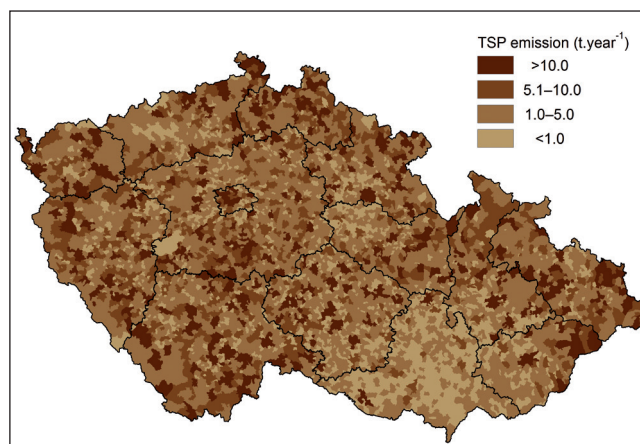


*Fig. 2.* Mean asthma prevalence in children in municipalities based on the Institute of Health Information and Statistics data for 2010. The surface was interpolated from mean asthma prevalence data in municipalities with paediatricians via Empirical Bayesian Kriging (EBK). EBK represents an advanced level of classical kriging geostatistics used for describing spatial patterns and interpolation values for locations where no samples were available, municipalities without paediatricians in this case.



*Fig. 3.* Five-year mean concentrations of $PM_{10}$ (2008–2012) in grid 1 × 1 km (based on the Czech Hydrometeorological Institute geodata).



*Fig. 4.* Annual total suspended particles emission (based on the Register of Emissions and Air Pollution Sources of the Czech Hydrometeorological Institute).
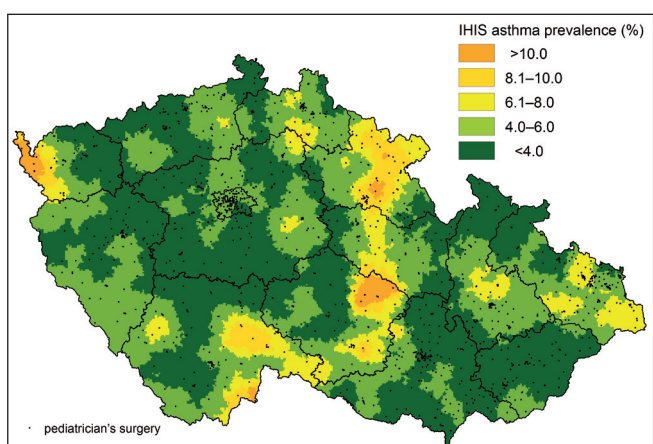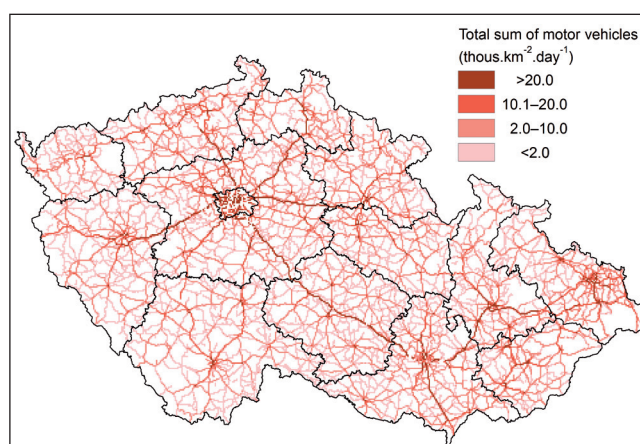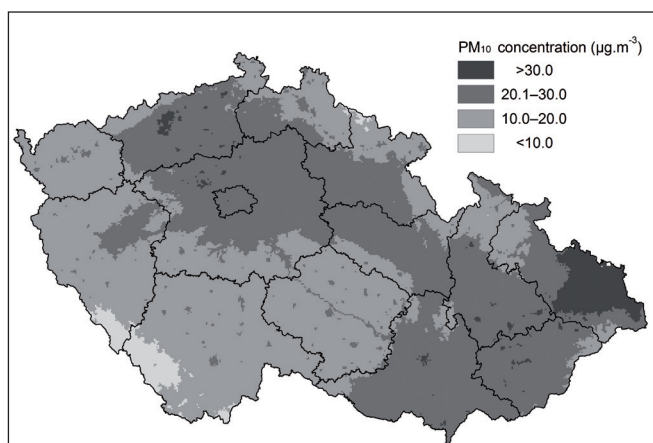


*Fig. 5.* Total number of motor vehicles in grid 1 × 1 km (based on the Road and Motorway Directorate geodata).
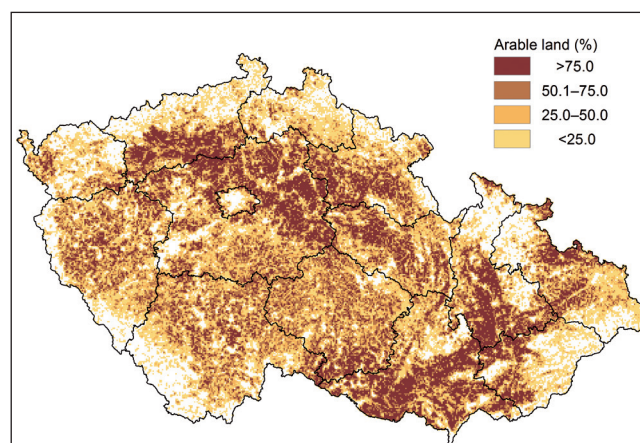


*Fig. 6.* Percentage proportion of arable land in grid 1 × 1 km (based on the Fundamental Base of Geographic Data).

absence of asthma in a particular child. Latent variables LV1, LV 7 and LV 13 were statistically significant (in Table 3 only statistically significant LV are pictured). For these latent variables, the main original observable variables (factor loadings higher than 0.5) are specified in the Table.
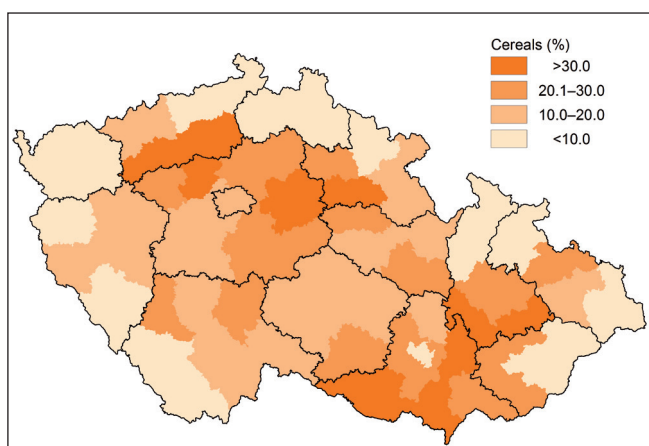
*Fig. 7. Percentage proportion of cereals area in districts (based on the Czech Statistical Office Agrocenzus 2010 database).*
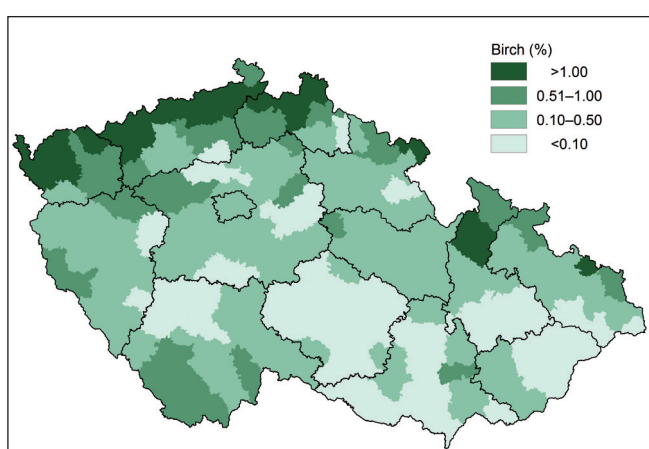


*Fig. 8. Percentage proportion of birch area in municipalities with extended powers (based on the Institute for Economic Forest Adjustment database).*

LV1 and LV13 may be described as an extensively cultivated agricultural area. Out of the original observable variables, it is predominantly characterized by percentage of the area, where wheat (OR = 1.041) and rape (OR = 1.110) are produced. In case of LV 13, there is a percentage occurrence of other agricultural areas (OR = 1.013) (areas with other uses than arable land – hop-garden, vineyard, orchard). LV7 may be described as cultural landscape. Out of the original observable variables, it is especially characterized by the percentage occurrence of lime trees (OR = 1.440). All the above mentioned variables are positively associated with the asthma prevalence (Table 3).

In the statistical evaluation of data for Model 2, 13 latent variables were obtained by factor analysis from 64 original observable variables (Table 4). Variables LV3, LV6, LV8, LV9, LV11 and LV13 were the only statistically significant. Even when LV3 was statistically significant, its OR was equal to 1.000 and particular variables show no positive and negative associations with the asthma prevalence. LV6 represents again agricultural landscape with production of oats (OR = 1.707), fodder crops (OR = 1.053), and green maize (OR = 1.116). Higher asthma occurrence was determined in areas with higher proportions of the above listed crops. It was the other way around in case of variables LV8, 9, 11 and 13, which indicate natural forests and are represented by, e.g. maple (OR = 0.634), ash (OR = 0.669), alder (OR = 0.727), or birch (OR = 0.863). Lower asthma occurrence was associated with such areas.

In the statistical evaluation of the national-wide IHIS data (Model 3), the original 64 observable variables were reduced by factor analysis to 21 latent variables. The statistically significant variables were LV1, 4, 6, 9, 10, 16, 18, 20 and 21 (Table 5).

The variables LV9 and LV20 represent again the agricultural landscape and they are positively associated with the asthma prevalence in the same way as in case of the evaluation of Models 1 and 2, e.g. percentage of the area used for production of oats (b = 5.618) and rye (b = 6.589). Among the statistically significant

*Table 3. Results of Model 1*

| Latent variable | p | Regression coefficient | OR | Observable variable | Factor loadings | Adjusted regression coefficient | Adjusted OR |
|---|---|---|---|---|---|---|---|
| LV1 | < 0.001 | 0.043 | 1.044 | Wheat incl. spelt | 0.930 | 0.040 | 1.041 |
| | | | | Rape and turnip rape | 0.884 | 0.104 | 1.110 |
| | | | | Industrial crops | 0.877 | 0.074 | 1.077 |
| | | | | Cereals | 0.874 | 0.024 | 1.025 |
| | | | | Pulses for grain | 0.812 | 0.483 | 1.621 |
| | | | | Sugar beet | 0.807 | 0.130 | 1.139 |
| | | | | Green maize | 0.717 | 0.168 | 1.183 |
| | | | | Arable fodder crops | 0.630 | 0.074 | 1.076 |
| | | | | Maize for grain | 0.598 | 0.091 | 1.095 |
| | | | | Rye | 0.540 | 0.896 | 2.449 |
| | | | | Fallow land | 0.517 | 0.150 | 1.162 |
| LV7 | 0.012 | 0.436 | 1.547 | Lime-tree | 0.836 | 0.364 | 1.440 |
| | | | | Willow | 0.664 | 3.886 | 48.700 |
| | | | | Poplar | 0.608 | 0.818 | 2.266 |
| | | | | Pine | 0.558 | 0.078 | 1.081 |
| LV13 | 0.013 | 0.014 | 1.014 | Other agricultural area | 0.932 | 0.013 | 1.013 |

*Table 4. Results of Model 2*

| Latent variable | p | Regression coefficient | OR | Observable variable | Factor loadings | Adjusted regression coefficient | Adjusted OR |
|---|---|---|---|---|---|---|---|
| **LV3** | **< 0.001** | **0.000** | **1.000** | Total number of motor vehicles | −0.931 | 0.000 | 1.000 |
| | | | | Urban greenery | −0.889 | 0.000 | 1.000 |
| | | | | Anthropogenic area | −0.849 | 0.000 | 1.000 |
| | | | | $NO_2$ annual mean concentration | −0.843 | 0.000 | 1.000 |
| | | | | Number of freight vehicles | −0.741 | 0.000 | 1.000 |
| | | | | Forest cover | 0.620 | 0.000 | 1.000 |
| **LV6** | **< 0.001** | **0.604** | **1.829** | Oats | 0.886 | 0.535 | 1.707 |
| | | | | Arable fodder crops | 0.778 | 0.052 | 1.053 |
| | | | | Green maize | 0.762 | 0.110 | 1.116 |
| | | | | Temporary grassland on arable land | 0.741 | 0.391 | 1.478 |
| | | | | Other cereals (buckwheat, millet, sorghum, triticale) | 0.572 | 0.215 | 1.240 |
| **LV8** | **< 0.001** | **−0.563** | **0.570** | Maple | 0.811 | −0.456 | 0.634 |
| | | | | Ash | 0.745 | −0.402 | 0.669 |
| | | | | Alder | 0.710 | −0.319 | 0.727 |
| | | | | Birch | 0.675 | −0.147 | 0.863 |
| | | | | Grassland | 0.533 | −0.012 | 0.988 |
| **LV9** | **< 0.001** | **−0.452** | **0.636** | Hornbeam | 0.908 | −0.411 | 0.663 |
| | | | | Other agricultural area | 0.642 | −0.028 | 0.972 |
| | | | | Oak | 0.617 | −0.055 | 0.946 |
| | | | | Ni – annual mean concentration | 0.594 | −0.196 | 0.822 |
| **LV11** | **0.011** | **−0.081** | **0.922** | Pine | 0.720 | −0.058 | 0.943 |
| **LV13** | **0.002** | **−0.001** | **0.999** | VOC – emission from large and medium sources (REZZO 1, 2) | 0.444 | 0.000 | 1.000 |
| | | | | Average ozone concentrations per growing season | 0.404 | 0.000 | 1.000 |

and positively associated variables, there are also other ones, in comparison to the previous models, which may be characterized as air pollution factors. LV1 may be described as an air-polluted area with high concentrations of $PM_{10}$, benzene, $PM_{2.5}$, lead, benzo(a)pyrene, $SO_2$ and cadmium. This variable is also related to LV4 – area with a high proportion of emissions of REZZO3 sources – local household heating (emissions of $SO_2$, CO, total suspended particles and $NO_x$).
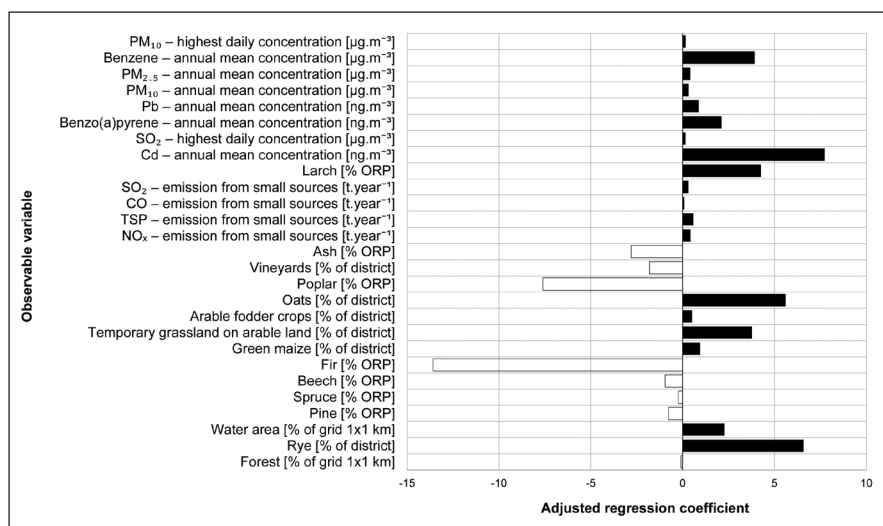


*Fig. 9. Regression coefficients for main observable variables and the asthma occurrence (Model 3).*

*Table 5. Results of Model 3*

| Latent variable | p | Regression coefficient | Observable variable | Factor loadings | Adjusted regression coefficient |
|---|---|---|---|---|---|
| LV1 | 0.006 | 0.184 | $PM_{10}$ – highest daily concentration | 0.948 | 0.174 |
| | | | Benzene – annual mean concentration | 0.931 | 3.925 |
| | | | $PM_{2.5}$ – annual mean concentration | 0.928 | 0.428 |
| | | | $PM_{10}$ – annual mean concentration | 0.925 | 0.340 |
| | | | Pb – annual mean concentration | 0.889 | 0.883 |
| | | | Benzo(a)pyrene – annual mean concentration | 0.864 | 2.129 |
| | | | $SO_2$ – highest daily concentration | 0.736 | 0.162 |
| | | | Cd – annual mean concentration | 0.705 | 7.739 |
| | | | Larch | 0.579 | 4.271 |
| LV4 | < 0.001 | 0.331 | $SO_2$ – emission from small sources (REZZO 3) | 0.972 | 0.322 |
| | | | CO – emission from small sources (REZZO 3) | 0.971 | 0.103 |
| | | | TSP – emission from small sources (REZZO 3) | 0.968 | 0.586 |
| | | | $NO_x$ – emission from small sources (REZZO 3) | 0.689 | 0.436 |
| LV6 | 0.062 | −3.197 | Ash | 0.877 | −2.802 |
| | | | Vineyards | 0.811 | −1.802 |
| | | | Poplar | 0.606 | −7.595 |
| LV9 | 0.003 | 6.932 | Oats | 0.810 | 5.618 |
| | | | Arable fodder crops. total | 0.554 | 0.521 |
| | | | Temporary grassland on arable land | 0.543 | 3.778 |
| | | | Green maize | 0.502 | 0.957 |
| LV10 | < 0.001 | −17.063 | Fir | 0.796 | −13.586 |
| | | | Beech | 0.662 | −0.945 |
| | | | Spruce | 0.621 | −0.224 |
| LV16 | 0.038 | −0.897 | Pine | 0.844 | −0.757 |
| LV18 | < 0.001 | 2.535 | Water area | 0.904 | 2.292 |
| LV20 | 0.004 | 8.378 | Rye | 0.787 | 6.589 |
| LV21 | 0.007 | −0.151 | Forest | 0.587 | −0.089 |

To the contrary, the variables LV6, LV10, LV16 and LV 21 represent forested landscape with occurrence of tree species like ash, fir, spruce, pine, and they were negatively associated with the asthma prevalence. In areas with higher proportions of the above mentioned woods, lower asthma prevalence in children was determined. The total number of motor vehicles and $NO_2$ concentrations were involved in the latent variable LV3, which was not statistically significant (Table 5).

The results of the model 3 analysis as well as the main observable variables from statistics of significant LV are pictured in Fig. 9. Regression coefficients indicate by how many units a number of asthmatics change in case that a particular observable variable changes by one unit.

## DISCUSSION

The evaluation of asthma prevalence is based on the results of NIPH cross-sectional study, which provided detailed localization of the child's residence address, but it only covered children who have their paediatricians in several selected larger cities. Any data from rural areas were missing in this case. In the nationwide scale, we could use available IHIS data from reports of all paediatricians which made possible to cover the Czech Republic, but with localization in respect of the paediatrician's address only. The paediatricians' data were preferred even though IHIS also gathers information from allergologists, but the higher number of paediatricians than that of allergologists gives more precise localization of young patients (related to the catchment area of a doctor's surgery).

The quality of input data on asthma prevalence can be also influenced by still not clear diagnostics that is constantly developing and unifying based on new scientific knowledge. The determination or verification of asthma diagnosis should be done by an erudite pneumologist, allergologist or clinical immunologist (24). But many patients do not get to specialists and they are treated by paediatricians only. An individual approach of each paediatrician to diagnosis determination and filling up questionnaires providing base for assessment, can be a source of distortion. This risk is minimalized in the NIPH cross-sectional survey by periodical

education of involved physicians, in case of IHIS data by including all paediatricians that represents sufficiently large data set.

The information about the quality of the environment was obtained from multiple sources which should cover a wide range of characteristics of the outdoor environment, starting from the most often applied air pollution characteristics, through land cover to ways of agricultural production of the land. The vast majority of the data are available to the public or they are provided upon a justified request. Moreover, these data are updated periodically, which is always one of the prerequisites for routine utilization within similar assessments.

Using the GIS analysis, the expected integration of all source data through one spatial unit was achieved and complete data sets were compiled to be subjected to statistical analysis with the aim of evaluation of suitability of the proposed approach to the assessment. Because of the extent of the input data a combination of factor analysis with logistic/linear regression was finally chosen.

The performed analyses confirmed the associations between asthma prevalence and air pollution published in literature (6, 7, 25) only in case of Model 3. Surprisingly, there were also associations with land use/land cover, in all three models.

The factors indicating the extensive land use (areas with a high proportion of cereal crops and other agricultural crops) showed the positive association with asthma prevalence in all three models. The agricultural production results in high amounts of emissions released into the air which may be either of natural origin (pollen, plant dust, fungal spores, soil), or of anthropogenic origin (agrochemicals). All of them may affect expression and development of asthma in some way (1, 16). To the contrary, the negative association was determined in relation to the areas covered with natural forests which have a positive effect on the quality of the environment and so implicitly on the lower disease occurrence (15).

The association between asthma prevalence and air pollution was found in Model 3. There were data for the Czech Republic and so they could already reflect differences in air pollution between urban and rural areas which were eliminated in case of Models 1 and 2 (children from cities only). The higher asthma prevalence in Model 3 was proved in relation to both increasing concentrations of pollutants ($PM_{10}$, $PM_{2.5}$, benzene, benzo(a)pyren, $SO_2$, Pb, Cd) and amounts of emissions from local (household) heating (REZZO3), which contribute significantly to poor air quality situation especially in small municipalities during winter months. The association between elevated concentrations of $PM_{2.5}$ and $SO_2$ and asthma prevalence was significant for example among residents of New York State (6), in Northern Israel was childhood asthma significantly associated with $PM_{10}$ concentration (7). The influence of benzo(a)pyren and $PM_{10}$ on asthma occurrence is confirmed by studies conducted on epigenetic level (26, 27). However, there are also studies which did not confirm these associations (28). This may be related to the existence of a complicated mechanism of development and expression of asthma which involves both host (genetics, obesity, sex) and environmental factors (1). In our study no significant associations were observed between the asthma prevalence and the distance from the road, traffic intensity and $NO_2$ concentration even though these impacts are also often indicated as significant (10, 17).

Considering the above information relating to outdoor it needs to be said that children spend a great part of their day indoor which was not assessed in our study due to inaccessibility of data. The truth is that indoor and outdoor communicate together by ventilation and a concentration of substances indoor is connected to outdoor, but there are also significant indoor sources of pollution (cooking, vacuuming, smoking, presence of pet) which have not been taken into account.

The applied approach does not enable us to perform more detailed analyses and adjust the results to other equally important influences, but it provides a complex view of the quality of the outdoor environment and its potential associations with the asthma prevalence. Our study demonstrated, using a large database, the association of asthma prevalence with the extensive land use and air pollution, and brought some interesting findings about possible connections that can be used for further detailed studies.

## CONCLUSIONS

In this study, the approach of combined use of GIS analyses and statistical evaluation of large input data sets was tested. A positive association was determined between the asthma prevalence in the population of children in the Czech Republic and air pollution (concentrations of $PM_{10}$, $PM_{2.5}$, benzo(a)pyrene, benzene, $SO_2$, Pb, Cd; emissions of $SO_2$, CO, $NO_x$ and particulate matter from local heating). Furthermore, there was also a positive association between the rate of agricultural use of landscape (areas with cereal crops and other agricultural crops), conversely, there was a negative association with the occurrence of natural forests. The study has brought a broader view of potential associations between the quality of the outdoor environment and asthma prevalence.

## REFERENCES

1. Global Initiative for Asthma. Global strategy for asthma management and prevention [Internet]. Vencouver: GINA; 2014 [cited 2014 Oct 10]. Available from: http://www.ginasthma.org/local/uploads/files/GINA_Report_2014_Aug12.pdf.
2. Institute of health information and statistics of the Czech Republic. Activities of medical facilities in selected fields of medical and preventive care in 2012 [Internet]. Prague: ÚZIS; 2013 [cited 2014 Oct 5]. Available from: http://www.uzis.cz/katalog/zdravotnicka-statistika/cinnost-zdravotnickych-zarizeni-ve-vybranych-oborech. (In Czech.)
3. Institute of health information and statistics of the Czech Republic. Activities of medical facilities in selected fields of medical and preventive care in 2000 [Internet]. Prague: ÚZIS; 2001 [cited 2014 Oct 5]. Available from: http://www.uzis.cz/katalog/zdravotnicka-statistika/cinnost-zdravotnickych-zarizeni-ve-vybranych-oborech. (In Czech.)
4. Kratěnová J, Puklová V, Malý M. Monitoring of childhood allergic diseases in the Czech Republic in the period 2011/2012. Alergie. 2014;16 Suppl 1:17-21. (In Czech.)
5. Esposito S, Galeone C, Lelii M, Longhi B, Ascolese B, Senatore L, et al. Impact of air pollution on respiratory diseases in children with recurrent wheezing or asthma. BMC Pulm Med. 2014 Aug 7;14:130.
6. Gorai AK, Tuluri F, Tchounwou PB. A GIS based approach for assessing the association between air pollution and asthma in New York State, USA. Int J Environ Res Public Health. 2014 May 6;11(5):4845-69.
7. Portnov BA, Reiser B, Karkabi K, Cohen-Kastel O, Dubnov J. High prevalence of childhood asthma in Northern Israel is linked to air pol-

lution by particulate matter: evidence from GIS analysis and Bayesian Model Averaging. Int J Environ Health Res. 2012;22(3):249-69.

8. Brauer M, Hoek G, Smit HA, de Jongste JC, Gerritsen J, Postma DS, et al. Air pollution and development of asthma, allergy and infections in a birth cohort. Eur Respir J. 2007 May;29(5):879-88.

9. Jedrychowski W. Environmental respiratory health in central and eastern Europe. Cent Eur J Public Health. 2000 Feb;8(1):33-9.

10. Bernstein DI. Diesel exhaust exposure, wheezing and sneezing. Allergy Asthma Immunol Res. 2012 Jul;4(4):178-83.

11. Dell SD, Jerrett M, Beckerman B, Brook JR, Foty RG, Gilbert NL, et al. Presence of other allergic disease modifies the effect of early childhood traffic-related air pollution exposure on asthma prevalence. Environ Int. 2014 Apr;65:83-92.

12. Jacquemin B, Siroux V, Sanchez M, Carsin AE, Schikowski T, Adam M, et al. Ambient air pollution and adult asthma incidence in six European cohorts (ESCAPE). Environ Health Perspect. 2015 Jun;123(6):613-21.

13. Favarato G, Anderson HR, Atkinson R, Fuller G, Mills I, Walton H. Traffic-related pollution and asthma prevalence in children. Quantification of associations with nitrogen dioxide. Air Qual Atmos Health. 2014;7(4):459-466.

14. Svendsen ER, Gonzales M, Mukerjee S, Smith L, Ross M, Walsh D, et al. GIS-modeled indicators of traffic-related air pollutants and adverse pulmonary health among children in El Paso, Texas. Am J Epidemiol. 2012 Oct 1;176 Suppl 7:S131-41.

15. Ruokolainen L, von Hertzen L, Fyhrquist N, Laatikainen T, Lehtomäki J, Auvinen P, et al. Green areas around homes reduce atopic sensitization in children. Allergy. 2015 Feb;70(2):195-202.

16. Khan IA, Arsalan MH, Siddiqui MF, Zeeshan S, Shaukat SS. Spatial association of asthma and vegetation in Karachi: A GIS perspective. Pak J Bot. 2010;42(5):3547-54.

17. Cook AG, deVos AJ, Pereira G, Jardine A, Weinstein P. Use of a total traffic count metric to investigate the impact of roadways on asthma severity: a case-control study. Environ Health. 2011 Jun 2;10:52.

18. Nuvolone D, Della Maggiore R, Maio S, Fresco R, Baldacci S, Carrozzi L, et al. Geographical information system and environmental epidemiology: a cross-sectional spatial analysis of the effects of traffic-related air pollution on population respiratory health. Environ Health. 2011 Mar 1;10:12.

19. Lyseen AK, Nøhr C, Sørensen EM, Gudes O, Geraghty EM, Shaw NT, et al. A review and framework for categorizing current research and development in health related Geographical Information Systems (GIS) studies. Yearb Med Inform. 2014 Aug 15;9:110-24.

20. Gribov A, Krivoruchko K. New flexible non-parametric data transformation for trans-Gaussian kriging. Quant Geol Geostat. 2012;17(1):51-65.

21. Krivoruchko K. Spatial statistical data analysis for GIS users. Redlands (CA): ESRI Press, 2011.

22. Czech Hydrometeorological Institute [Internet]. Prague: CHMI; 2014 [cited 2014 Oct 10]. Five-year averages in 2008-2012. Available from: http://portal.chmi.cz/files/portal/docs/uoco/isko/ozko/12petileti/12petiletzip.html. (In Czech.)

23. Hebák P, Hustopecký J, Jarošová E, Pecáková I. Multivariate statistical methods 1. Prague: Informatorium; 2004. (In Czech.)

24. Kašák V. Asthma in 2013: under extraordinary pressure. Alergie. 2013;15 Suppl 1:23-30. (In Czech.)

25. Dostál M, Pastorková A, Švecová V, Šrám RJ. The follow up of children health in Ostrava 2001-2009. Alergie. 2013;15(1):16-28. (In Czech.)

26. Kabesch M, Michel S, Tost J. Epigenetic mechanisms and the relationship to childhood asthma. Eur Respir J. 2010 Oct;36(4):950-61.

27. Ho SM. Environmental epigenetics of asthma: an update. J Allergy Clin Immunol. 2010 Sep;126(3):453-65.

28. Mölter A, Simpson A, Berdel D, Brunekreef B, Custovic A, Cyrys J, et al. A multicentre study of air pollution exposure and childhood asthma prevalence: the ESCAPE project. Eur Respir J. 2015 Mar;45(3):610-24.