The Design of Geographic Information Systems for the Storage and Analysis of Public Health and Environmental Data

Randolph B. Flay
University of California, Berkeley

The effective collection and usage of data about environmental conditions, such as air and drinking water quality, and their influence on public health are necessary to guide government decision-making and target environmental remediation efforts to the areas of greatest need. This paper explores the combination of air and drinking water quality data with morbidity data in Vilnius, Lithuania (population 580,000) in a Geographic Information Systems (GIS) format for mapping and statistical analysis. Correlations documented in previous epidemiological work between indicators of air/water quality and instances of certain respiratory/gastrointestinal illnesses, respectively, were also found in Vilnius. For example, the average rate of chronic bronchitis (ICD-9 Code 491) in the 0-19 year age group calculated for city districts with an Air Pollution Index (API) of "very polluted" was 76% higher than districts with an API of "unpolluted." Additionally, drinking water supply districts above the median level of turbidity had a 124% higher rate of disorders of stomach function (ICD-9 Code 536) in the 0-19 year age group. It is hoped that this approach to understanding the spatial distributions of disease and their relationships to environmental conditions will prove useful in guiding government decision-making, encouraging information exchange, and educating the public.

Keywords: Air Quality, Drinking Water Quality, Epidemiology, Geographic Information Systems, Lithuania, Morbidity, Vilnius.

Address: Department of Environmental Science, Policy & Management; 151 Hilgard Hall # 3110; Berkeley, CA 94720-3110.
Telephone: (510) 643-7837; E-mail: randy@waterscape.org

I. Introduction and Background:

The legacy of Soviet industrialization, a nascent economy, and recently restructured environmental and health protection regime pose a great challenge to air and drinking water quality in the Republic of Lithuania (Feshbach 1995; Kadunas 1997). Financing for the protection and remediation of the environment is limited, and systems to display areas where environmental conditions influence public health are of need to target programming to the areas of greatest need. In this paper we discuss the conversion of existing morbidity and environmental data to Microsoft Access database software for storage and querying; the statistical analysis of relationships between morbidity rates and environmental conditions in Microsoft Excel and Systat; and its display in ESRI's ArcView, a form of Geographic Information Systems (GIS) software. In particular, we spatially map the distribution of diseases in Vilnius, the capital of Lithuania, and see if these distributions of disease correlated to air and drinking water quality indicators (where correlations have been previously documented in epidemiological work).

Although air pollution from stationary sources has declined in the past few years, an increase in pollution from mobile sources yielded an air pollution total of 107,900 tons in Vilnius in 1996 (Zickus 1999). Carbon monoxide (CO) accounted for about 77% of this total by weight, while nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and other pollutants comprised the remainder. Currently mobile sources, mostly automobiles, account for 88% of air pollution in Vilnius. In addition to an aesthetic influence on air quality in the commonly known forms of smog and smell, these air quality indicators have an influence on public health (Van Leeuwen 1997; World Resources Institute 1999). Epidemiological studies have demonstrated that increased levels in CO, NO_2, and SO_2, increase the instances of several types of respiratory infections, heart disease, and other respiratory-related diseases (Pershagen 1995; Brunekreef 1997; Wojtyniak 1997; Jedrychowski 1999). Given this conditions there is a need to set air quality standards, monitor air quality, and meet air quality standards in order to protect public health. Further people should be educated about the influence of air quality on public health and how they can help reduce their exposure to unhealthy conditions and reduce their contribution to air pollution (Ministry of Health Protection 1998b). Vilnius city districts are shown in **Map 1** (appended).

In the air quality-morbidity rate portion of this study, we compared levels of CO, NO_2, SO_2, and the Air Pollution Index (API) to rates of respiratory-related diseases and other diseases that are known to be influenced by air quality. Our results showed a positive correlation between air quality indicators and several diseases. Hopefully this will highlight the need to meet standards, seek ways to improve quality, and develop other ways of minimizing the health risk from air pollution. The diseases evaluated in relation to air quality are shown in **Figure 1**.

Additionally in Vilnius, 19 well fields supply approximately 155,000 cubic meters per day to the 580,000 residents, comprising the bulk of drinking water consumed (Klimas 1998). These 19 well fields feed nine distinct municipal water supply districts shown in **Map 2** (appended). Many of these well fields are located in urbanized areas where the threat of anthropogenic contamination is high, raising questions about water quality and the influence on public health (Ministry of Health Protection 1998a). The influence of nickel and lead on kidney disorders and birth defects is also documented in epidemiological work (Agency for Toxic Substances and Disease Registry 1993; Agency for Toxic Substances and Disease Registry 1996). Additionally, it is known that turbidity can compromise disinfection and serve as a medium for microbiological growth, which can lead to gastrointestinal illness (Schwartz 1997).

For the drinking water quality-morbidity rate portion of this study, we examined public health data from years 1991 to 1995 in conjunction with sets of water quality data from water supply districts to see if perhaps any relationships exist in Vilnius. Of interest were the indirect influence that turbidity might have on rates of intestinal infectious diseases (International Classification of Diseases-9th Revision (ICD-9) codes 001-009), gastritis and duodenitis (535), disorders of stomach function (536), and diseases of the esophagus, stomach, and duodenum (530-537). We also studied the influence of nickel levels on kidney infections (590) and the influence of lead on congenital anomalies (740-759). The diseases evaluated in relation to water quality are shown in **Figure 2**.

Code	INTERNATIONAL CLASSIFICATION OF DISEASES, 9th Revision
162	Malignant neoplasm of trachea, bronchus, and lung
410	Acute myocardial infarction
460	Acute nasopharyngitis [common cold]
461	Acute sinusitis
462	Acute pharyngitis
464	Acute laryngitis and tracheitis
465	Acute upper respiratory infections of multiple or unspecified sites
466	Acute bronchitis and bronchiolitis
472	Chronic pharyngitis and nasopharyngitis
473	Chronic sinusitis
476	Chronic laryngitis and laryngotracheitis
477	Allergic rhinitis
480	Viral pneumonia
481	Pneumococcal pneumonia
482	Other bacterial pneumonia
485	Bronchopneumonia, organism unspecified
486	Pneumonia, organism unspecified
487	Influenza
490	Bronchitis, not specified as acute or chronic
491	Chronic bronchitis
493	Asthma

Figure 1: Diseases examined for relationships to air quality indicators.

We formed the following hypotheses about relationships between air quality and morbidity data for age groups 0-19 and 0-100 years:

$H_{O(1)}$: The Air Pollution Index (API) in city districts does not display a correlation to ICD-9 diseases listed in Figure 1. (Each disease was tested separately).
$H_{A(1)}$: The Air Pollution Index (API) in city districts displays a correlation to ICD-9 diseases listed in Figure 1.
$H_{O(2)}$: CO in city districts does not display a correlation to ICD-9 diseases listed in Figure 1.
$H_{A(2)}$: CO in city districts displays a correlation to ICD-9 diseases listed in Figure 1.
$H_{O(3)}$: NO_2 in city districts does not display a correlation to ICD-9 diseases listed in Figure 1.
$H_{A(3)}$: NO_2 in city districts displays a correlation to ICD-9 diseases listed in Figure 1.
$H_{O(4)}$: SO_2 in city districts does not display a correlation to ICD-9 diseases listed in Figure 1.
$H_{A(4)}$: SO_2 in city districts displays a correlation to ICD-9 diseases listed in Figure 1.

Code(s)	INTERNATIONAL CLASSIFICATION OF DISEASES, 9th Revision
(001-009)	INTESTINAL INFECTIOUS DISEASES
(530-537)	DISEASES OF ESOPHAGUS, STOMACH, AND DUODENUM
535	GASTRITIS AND DUODENITIS
536	DISORDERS OF FUNCTION OF STOMACH
590	INFECTIONS OF KIDNEY
(740-759)	CONGENITAL ANOMALIES

Figure 2: Diseases examined for relationships to drinking water quality indicators.

We formed the following hypotheses about relationships between water quality and morbidity data:

$H_{O(1)}$: Water supply districts with high turbidity as measured at the tap do not have higher instances of ICD-9 diseases 001-009 (grouped), 530-537 (grouped), 535, and 536. (Each disease or group of diseases was tested separately).
$H_{A(1)}$: Water supply districts with high turbidity as measured at the tap do have higher instances of ICD-9 diseases 001-009 (grouped), 530-537 (grouped), 535, and 536.
$H_{O(2)}$: Water supply districts with high turbidity as measured at the treatment facility, i.e. the source, do not have higher instances of ICD-9 diseases 001-009 (grouped), 530-537 (grouped), 535, and 536.
$H_{A(2)}$: Water supply districts with high turbidity as measured at the treatment facility do have higher instances of ICD-9 diseases 001-009 (grouped), 530-537 (grouped), 535, and 536.

$H_{O(3)}$: Water supply districts with higher nickel concentrations as measured at the treatment facility do not have higher instances of ICD-9 590.
$H_{A(3)}$: Water supply districts with higher nickel concentrations at the treatment facility do have higher instances of ICD-9 590.
$H_{O(4)}$: Water supply districts with higher lead concentrations as measured at the treatment facility do not have higher instances of ICD-9 740-759 (grouped).
$H_{A(4)}$: Water supply districts with higher lead concentrations as measured at the treatment facility do have higher instances of ICD-9 740-759 (grouped).

II. Data

General Data Structure

A main purpose of this paper is to show how information that is frequently collected by separate government agencies can be effectively used to guide decision-making when brought together and echo the need to provide data in formats readily suitable for use by other agencies, universities, and the public. Three sources provided the data for this analysis in various formats ranging from handwritten water quality measurements to GIS layers of the political features of the city. Morbidity data from Vilnius and some drinking water quality data was provided by Vilnius Public Health Center under the Ministry of Health Protection, *Vilniaus visuomenės sveikatos centras* (VVSC). Air quality data was provided by Ministry of Environment, *Aplinkos ministerija* (AM). The bulk of the drinking water quality data was provided by Vilnius Waterworks, *Vilniaus vandenys* (VV). Additionally, basic political features of Vilnius city were provided in GIS format by the City Development Department of Vilnius Municipality, *Miesto plėtros departamentas* (MPD).

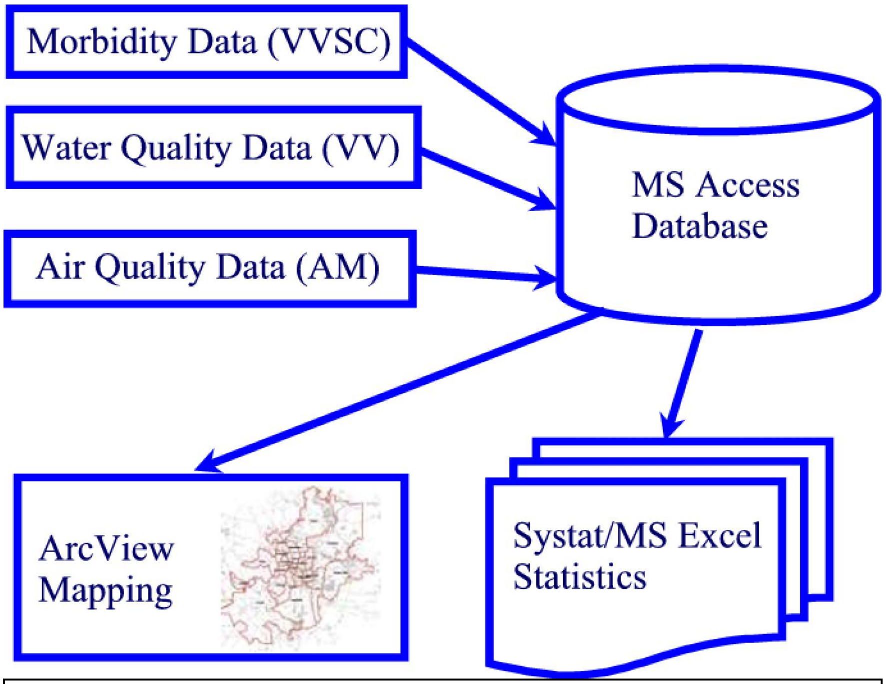

Figure 3: A general scheme showing the organization of information in our study.

A database was designed in Microsoft Access to accept the data; and using various geographic parameters for each data type, relationships were created in the database to relate the spatial distributions of air and drinking water quality to morbidity rates. Once this database was completed, sets of data were extracted to map the morbidity rates in ArcView and perform statistical tests for relationships between air/drinking water quality and morbidity rates of selected diseases in Systat and Microsoft Excel. **Figure 3** shows the general structure of the data.

Morbidity Data

Morbidity data were obtained from the public health database maintained by VVSC. Each time a patient visits a polyclinic in Vilnius, a record is sent to the VVSC recording the diagnosis, the patient’s age, residence, the date, and other useful information. The data spanned the years 1991 to 1995 for ages 0-19 and years 1991 and 1992 for all ages.

After 1992, only records for those under age 20 were kept. The data set contains over 1 million records and diseases are recorded utilizing the *International Classification of Diseases, 9th Revision* (ICD-9) codes.

Raw data from the Vilniaus VSC was collected in a DB4 database format. In order to analyze the information, a Microsoft Access database was designed and the information imported. Also developed in the database was a feature to analyze the data geographically by city district and water supply district. To accomplish this, each street in the public health database was assigned to the city district and the water supply district in which it is located. This allowed for the creation of queries based on the districts; i.e. one could then calculate the number of diagnoses of a particular illness by city district or water supply district. Selected data is currently available on the Internet and by request (http://www.is.lt/vilniaus_vsc/).

Air Quality Data

In 1995, the "Air Quality Management in Vilnius City" project was initiated by AM and the Swedish government, and a computerized air pollution management system know as Airviro was implemented (Air Quality Management Group 1999). This system consists of three automatic pollutant monitoring stations in Vilnius which monitor concentrations of CO, NO_2, NO, SO_2, and O_3. These stations are located in Senamiestis, Žirmūnai, and Žverynas. Measurement data is sent hourly to a central database and is available via the Internet (http://vilnair.gamta.lt/).

	CO*	SO_2	O_3	NO_2	NO
Hourly Average (DLK)	5	500	160	85	400
24 Hour Average (DLK)	3	50	30	40	60

CO* - concentration, mg/m3.

Note: O_3 and NO where not evaluated in this study and are only presented for reference.

Figure 4: Maximum Permissible Concentrations (HN 33-1998) µg/m^3.

City District	API	CO (mg/m^3)	NO_2 ($\mu g/m^3$)	SO_2 ($\mu g/m^3$)
ANTAKALNIO	1	2.5	40	12
FABIJONISKIU	1	1.5	35	12
JUSTINISKIU	1	2	40	10
KAROLINISKIU	1	1.5	37	13.5
LAZDYNU	1	1.5	35	15
N.VILNIOS	1	1	35	10
NAUJAMIESCIO	2.5	4.5	60	17
NAUJININKU	1.5	2.5	45	15
PASILAICIU	1	1.5	35	10
RASU	1	1.5	40	13
SENAMIESCIO	2.5	3.5	50	15
SESKINES	2	3.5	50	15
SNIPISKIU	2.5	4	55	17
VERKIU	1	1.5	35	12
VILKPEDES	2.5	4	55	15
VIRSULISKIU	1.5	3	45	11
ZIRMUNU	2	3	45	15
ZVERYNO	2.5	3.5	50	20

Source: Data was extrapolated from maps of Vilnius produced by the Air Quality Management Group at the Ministry of Environmental Protection in 1996.

Figure 5: Estimated average concentrations of selected air pollutants in 1996.

The Air Quality Management Group at AM constructed models based on these measurements, taking into account meteorological conditions in order to simulate pollutant cover throughout Vilnius city districts. CO, NO_2, and SO_2, concentrations are plotted on a map of Vilnius, allowing for the pollutant level in each city district to be estimated. Also, an API is calculated in Vilnius taking into account government standards for CO, NO_2, and SO_2. This is the formula for calculating the API:

$$API = C(_{CO})/LV(_{CO}) + C(_{NO2})/LV(_{NO2}) + C(_{SO2})/LV(_{SO2})$$

Where, $C(_{CO})$ = CO concentration
$LV(_{CO})$ = Hourly Limit Value for CO
$C(_{NO2})$ = NO_2 concentration
$LV(_{NO2})$ = Hourly Limit Value for NO_2
$C(_{SO2})$ = SO_2 concentration
$LV(_{SO2})$ = Hourly Limit Value for SO_2

Based on the API, each Vilnius city district was classified as unpolluted, semi-polluted, and very polluted, and then set to numbers 1,2, and 3 respectively. In this way "1" denotes clean air and "3" denotes very polluted air. **Figure 4** shows Lithuania Hygienic Norm 33-1998 standards for selected air pollutants (Ministry of Health Protection 1998b).

Figure 5 shows the estimated average concentrations of selected air pollution indicators for 1996. Perhaps it would have been more conducive to use data from 1995 or slightly earlier, but the data were not available given the 1996 start date of the air quality monitoring program. However, the air quality conditions in Vilnius have not drastically changed over the time period of the study and it is believed that air quality data among districts should still be representative of a chronic burden to respiratory systems.

Based on the API, the worst air conditions are found in the city districts of Žverynas, Vilkpedės, Snipiškes, Senamiestis, and Naujamiestis. The best air conditions are found in Antakalnis, Fabijoniškės, Justiniškės, Karoliniškes, Lazdynai, N. Vilnia, Pašilaičiai, Rasos, and Verkiai.

Drinking Water Quality Data

Drinking water quality data for years 1996 to 1999 were obtained from VV for nickel, lead, and turbidity concentrations as measured at the nine separate water supply district reservoirs. Additional water quality data for years 1992 to 1994 also came from the VVSC that performs turbidity testing in the municipal water supply system at the tap. General, city-wide descriptive statistics are found in **Figure 6**. Also found in Figure 6 is the Shapiro-Wilk test showing the probability of a normal distributions (results below 0.05 are typically considered non-normal or skewed).

	Turbidity (mg/L) at Tap	Turbidity (mg/L) at Source	Ni (mg/L)	Pb (mg/L)
75th Quartile	2.175	1.425	0.03	0.02
Median	1.45	0.9	0.0185	0.016
25th Quartile	1	0.3	0.014	0.012
Shapiro-Wilk W Test	0.0001	0.0001	0.0001	0.0246
N	28	78	70	70
Mean	1.825	0.919	0.0218	0.0169
Standard Error of Mean	0.275	0.073	0.00147	0.0008
Upper 95% CI	2.388	1.065	0.0248	0.0185
Lower 95% CI	1.262	0.077	0.0189	0.0153

Figure 6: Descriptive statistics for Turbidity, Nickel (Ni), and Lead (Pb).

Average turbidity measured at the tap was highest in the water supply system 7-A.Paneriai and lowest in 2-Trinapolis. Average concentration in 7-A.Paneriai was nearly twice the Lithuania Hygienic Norm 24-1998 standard of 1.5 mg/L (Ministry of Health Protection 1998a). Average turbidity measured at the source were highest in 5-Tupatiškes and lowest in 3-Vingis. Average nickel concentrations were highest in 3-Vingio at 0.038 mg/L, exceeding the norm of 0.020 mg/L, and lowest in 6-Kirtimai at 0.014 mg/L. Average lead concentrations were highest in 3-Vingio at 0.025 mg/L, at the norm of 0.025 mg/L, and lowest in 6-Kirtimai at 0.010 mg/L.

III. Methods, Analysis & Results:

Morbidity Rate Calculations

The lack of air quality data for Trakų Vokė and A. Paneriai districts, along with the absence of a residential population in Pilaitė, excluded them from the study. Thus only results for 18 of the 21 city districts are presented. Utilizing a data table from VVSC that also included the population living on each street in Vilnius, we calculated annual disease rates per 100,000 residents. In total, we composed two general morbidity data sets; one contained all ages for years 1991 to 1992 and the second contained ages 0-19 but spanned from 1991 to 1995. Health data after 1992 were only collected for children. Since to analyze the data for years 1991 to 1995 only included patients aged 0-19, we had to multiply by the portion of the population represented by the age-group 0-19. This was roughly calculated as 0.261 (Statistical Office of Vilnius 1998). We further delimited this set by two geographic units: city districts for our air quality analyses and drinking water supply districts for our drinking water quality analyses.

Disease	Acute nasopharyngitis [common cold]	Acute laryngitis and tracheitis	Chronic pharyngitis and nasopharyngitis	Chronic sinusitis	Chronic laryngitis and laryngotracheitis	Viral pneumonia	"Bronchitis, not specified as acute or chronic"	Chronic bronchitis
ICD-9 CODE	460	464	472	473	476	480	490	491
Total Number of Diagnoses	59973.00	13716.00	214.00	185.00	35.00	301.00	715.00	942.00
ANTAKALNIO	11858.92	3701.94	15.91	58.34	5.30	5.30	291.70	10.61
FABIJONISKIU	6910.55	2694.40	6.68	3.34	3.34	10.02	141.99	71.83
JUSTINISKIU	4911.57	2483.20	12.38	0.00	0.00	14.15	91.97	109.66
KAROLINISKIU	5422.82	1279.64	11.93	5.97	0.00	98.43	26.85	47.73
LAZDYNU	6201.50	1428.11	13.04	6.52	0.00	254.32	39.13	123.90
N.VILNIOS	1734.13	459.59	5.34	0.00	2.67	8.02	5.34	0.00
NAUJAMIESCIO	8585.47	680.04	21.25	13.28	15.94	10.63	0.00	154.07
NAUJININKU	5501.56	1042.37	21.66	5.41	0.00	8.12	113.71	316.77
PASILAICIU	7752.56	3091.53	20.77	0.00	5.93	32.64	121.64	195.82
RASU	5780.84	1069.35	5.29	84.70	10.59	5.29	280.57	100.58
SENAMIESCIO	8526.35	877.25	10.52	33.66	0.00	27.35	71.53	159.88
SESKINES	6102.73	1713.00	2.65	2.65	2.65	21.18	58.25	103.26
SNIPISKIU	8274.54	2104.51	4.78	33.48	11.96	16.74	78.92	133.92
VERKIU	7903.62	1006.51	4.07	46.77	2.03	20.33	18.30	79.30
VILKPEDES	11111.44	1102.53	20.80	50.52	8.92	5.94	0.00	356.61
VIRSULISKIU	5818.24	755.30	9.21	0.00	3.07	171.94	30.70	61.41
ZIRMUNU	10418.10	2255.89	14.55	51.60	0.00	6.62	224.93	54.25
ZVERYNO	6114.93	1617.74	12.49	74.95	0.00	68.71	0.00	274.83

Source: The raw data for these geographical calculations were obtained from the Vilniaus Visuomenes Sveikatos Centras database.

Figure 7: Average annual number of diagnoses per 100,000 for years 1991-1995 for the age group 0-19.

Two figures refer to the air quality-morbidity rate analyses. **Figure 7** shows the average annual number of diagnoses per 100,000 people averaged over years 1991 and 1992 and including all ages, 0-100 years. **Figure 8** shows the average annual number of diagnoses per 100,000 people as well. However, this table is averaged over years 1991 to 1995 and includes only persons aged 0-19 years. The total number of diagnoses are also listed in the tables.

Disease	Acute myocardial infarction	Acute nasopharyngitis [common cold]	Acute laryngitis and tracheitis	Chronic pharyngitis and nasopharyngitis	Chronic sinusitis	Chronic laryngitis and laryngotracheitis	Viral pneumonia	"Bronchitis, not specified as acute or chronic"	Chronic bronchitis
ICD-9 CODE	410	460	464	472	473	476	480	490	491
Total Number of Diagnoses	343.00	28160.00	10536.00	826.00	238.00	475.00	351.00	603.00	948.00
ANTAKALNIO	46.72	2978.02	1233.78	58.83	48.45	6.92	1.73	112.48	41.53
FABIJONISKIU	4.36	1664.41	740.10	13.08	1.09	4.36	4.36	105.73	28.34
JUSTINISKIU	10.39	1307.56	1512.98	39.24	1.15	8.08	8.08	65.78	78.48
KAROLINISKIU	17.52	1202.76	474.87	60.33	13.62	5.84	17.52	15.57	54.49
LAZDYNU	29.78	1463.71	604.20	51.06	12.76	17.02	17.02	34.04	110.63
N.VILNIOS	1.74	1115.76	299.86	6.97	0.00	1.74	5.23	3.49	0.00
NAUJAMIESCIO	53.73	2282.65	383.04	65.86	10.40	135.19	74.53	6.93	116.13
NAUJININKU	22.96	1406.16	408.07	28.26	3.53	14.13	70.66	70.66	88.33
PASILAICIU	11.62	2791.54	1839.09	38.72	0.00	1.94	11.62	94.86	94.86
RASU	51.81	1861.57	524.97	58.71	20.72	44.90	27.63	151.97	69.08
SENAMIESCIO	42.55	2689.16	483.20	60.40	24.71	65.89	72.75	35.69	116.68
SESKINES	0.00	1679.28	666.87	5.18	3.46	1.73	10.37	39.74	38.01
SNIPISKIU	34.33	2474.88	861.37	87.39	35.89	18.73	20.29	17.16	102.99
VERKIU	22.55	2668.03	608.96	82.26	27.86	9.29	27.86	14.59	61.03
VILKPEDES	40.72	3833.48	465.37	71.74	21.33	238.50	141.55	7.76	139.61
VIRSULISKIU	42.07	1324.17	1406.31	52.09	8.01	12.02	16.03	6.01	50.08
ZIRMUNU	56.98	3040.56	1049.78	76.83	64.75	18.13	11.22	80.29	84.60
ZVERYNO	28.53	1785.28	570.64	118.20	36.68	20.38	12.23	4.08	142.66

Source: The raw data for these geographical calculations were obtained from the Vilniaus Visuomenes Sveikatos Centras database.

Figure 8: Average annual number of diagnoses per 100,000 for years 1991-1992 for the age group 0-100.

The vast majority of diseases chosen for examination in relation to air quality were respiratory-related infections of various locations. Also chosen was Acute Myocardial infarction (410) due to a documented relationship to CO levels. For the sake of brevity, only data for diseases which showed a relationship to air pollutants are presented in the tables.

The tables can be referred to for the identification of city districts having the highest and lowest disease rates. As a highlight of the findings, Naujamiestis, Rasos, and Žirmūnai, have the highest rates of Acute Myocardial infarction

(410). Chronic bronchitis (491) is most prevalent in the city districts of Žverynas, Vilkpedės, Senamiestis, and Naujamiestis for the 0-100 year age group.

For the drinking water quality-morbidity rate analyses, we calculated the morbidity rates as a monthly diagnosis rate in 100,000. General city-wide descriptive statistics are presented in **Figure 9**. Only the cohort aged 0-19 for years 1991 to 1995 was included.

Disease ICD-9 Code	001-009	530-537	535	536	590	740-759
75th Quartile	3.31	90.44	98.48	105.48	9.87	0
Median	0	43.59	55.42	61.23	0	0
25th Quartile	0	0	19.81	19.81	0	0
Shapiro-Wilk W Test	0.001	0	0	0	0	0.001
N	2358	1314	540	540	540	4824
Mean	7.436	61.816	66.974	83.136	7.295	1.894
Standard Error of Mean	0.443	2.279	2.742	4.402	0.939	0.106
Upper 95% CI	8.305	66.287	72.361	91.782	9.14	2.103
Lower 95% CI	6.567	57.345	61.588	74.49	5.45	1.686

Figure 9: Monthly Diagnosis Rates for Age Group 0-19 for years 1991-1995.

Mapping Morbidity in ArcView

Mapping of diseases began on the level of city district since there are only 21 city districts and about 800 streets in Vilnius. Recently we have obtained a GIS layer of Vilnius streets, enabling very detailed maps of disease distributions to be created. Although they do not directly relate to our studies relating to the influence of air and drinking water quality on public health, we think they are a valuable product of this research.

In order to map the disease distributions, a look-up table in the Microsoft Access database relating streets of the morbidity data to the attribute table of the streets in the GIS layer in ArcView was created. Although ArcView supports live, Open Database Connectivity (ODBC) connections to Microsoft Access, we found it more efficient to export query results from Microsoft Access of disease rates and disease counts to DB4 format for addition to ArcView projects as tables. Then we used ArcView's *join* function (an SQL "JOIN" statement) to link the street codes of the GIS attribute table to the street codes of the corresponding, imported morbidity rate or count table.

Since most streets are stored as vectors in ArcView, converting the streets to polygons helps give a somewhat more impressive display of the distribution of disease. A useful feature found in ArcView's Spatial Analyst extension is the *proximity* function. This function created a grid of Vilnius of different cells or polygons and each cell was assigned to the nearest street. This allowed us to map the disease rates or disease counts to polygons closest to the street.

As a sample of the mapping capabilities now possible, **Map 3** (appended) shows the count of cases and **Map 4** (appended) shows the rates of acute nasopharyngitis (common cold) (ICD-10 code J00) in 100,000 children aged 0-19 for years 1997-1999. The *International Classification of Diseases-10th Revision* was implemented in Lithuania in 1997.

Air Quality and Morbidity Rates

Based on the results for air quality indicators and public health diagnosis rates calculated in each of the 18 city districts, regression analysis was performed to look for relationships utilizing Systat 7.0. In order to use linear regression we made two basic assumptions: (1) the data used in fitting the model are representative of the population, and (2) the true underlying relationship between air quality and the disease rate is linear. Other assumptions necessary to calculate the standard error of the predicted values or construct confidence intervals for the regression slope, such as constant variance (homoscedasticity) and normal distribution of the residuals, were not met. Therefore we have only presented results where we think a relationship exists and requires further study, without giving information regarding the magnitude of the relationship (such as the regression slope).

Air pollution indicators were held as the independent and the disease rates were held as the dependent. The results of these linear regression analyses are presented in **Figure 10**. The regression analysis is a unique tool because it tells us how much of the variability in our dependent variable is explained by our independent variable. This is enumerated in the r^2 value in **Figure 10**. The r^2 value is equal to 1.0 minus the ratio of the residual variability of the Y variable to the original variance (StatSoft Inc. 1998). If X and Y are perfectly related the residual variance will be equal to 0.0 and thus the r^2 value equal to 1.0. For example, if we have an r^2 value of 0.55, then we know that the relationship explains 55% of the variability in the data, and we have 45% residual variability. In short, the closer the r^2 value is to 1.0, the more of the variability we have explained.

Also included is the p-value of the relationship. The p-value is the probability of randomly finding a correlation between the independent and dependent variables. A p-value of 0.093 means that there is a 9.3% chance that the relationship between the variables is a complete “fluke.” For the purpose of this study a p-value < 0.05 has been chosen for values that we will accept, which corresponds to "confidence level" of $\alpha > 95\%$. Although the r^2 value and the p-value are related, one often refers to the r^2 value as a measure of strength or magnitude and the p-value as a measure of significance.

Figure 10 shows the results of regression analyses comparing air quality indicators to these disease rates. Where there appear numbers in a particular comparison, the p-value of the regression was less than 0.05 and the α or confidence level was greater than 95%. This essentially means that there is only a 5% chance that the relationship we found was a statistical "fluke." These relationships were accepted as being statistically significant. Presented adjacent to the p-value is the r^2 value, the measure of strength. Again, the r^2 value tells us how much of the variation in a disease rate can be explained by the air pollution indicator.

Where there appear only an "X" in the table the p-value of the relationship was greater than 0.05 but less than 0.10. Although these relationships were not considered significant in this study, there does seem to be some grounds for a relationship that might be clarified with further study. Where there is simply a blank spot in the chart, p-values were greater than 0.10 and there seems to be no significant statistical relationship.

		INTERNATIONAL CLASSIFICATION OF DISEASES	Acute myocardial infarction	Acute nasopharyngitis [common cold]	Chronic pharyngitis and nasopharyngitis	Chronic sinusitis	Chronic laryngitis and laryngotracheitis	Viral pneumonia	Bronchitis, not specified as acute or chronic	Chronic bronchitis
Air Quality Indicator	**Age Group**	**ICD-9 CODE**	**410**	**460**	**472**	**473**	**476**	**480**	**490**	**491**
API	0-19 Years	p-value								0.03
		r^2								0.27
	0-100 Years	p-value	x	x	0.05		0.01	0.02	0.05	0.00
		r^2	x	x	0.23		0.32	0.30	0.22	0.43
CO	0-19 Years	p-value		0.03			x			x
		r^2		0.25			x			x
	0-100 Years	p-value	0.04	x	x			0.03		0.01
		r^2	0.23	x	x			0.27		0.34
NO_2	0-19 Years	p-value								0.04
		r^2								0.24
	0-100 Years	p-value	0.05					0.01		0.01
		r^2	0.22					0.34		0.34
SO_2	0-19 Years	p-value								0.03
		r^2								0.26
	0-100 Years	p-value			0.01	x				0.00
		r^2			0.35	x				0.46

Notes: Where the p-values where less than 0.05 we accepted the relationship. This is a statistical standard equal to the conficence level (alpha) = 95%.
x = p-values were between 0.10 and 0.05

Figure 10: Results of regression analysis between air quality indicators and selected diseases in Vilnius city districts.

In this study, socioeconomic and other factors may also influence public health in addition to environmental conditions and be confounders in the analysis (Anderson 1997). Age distribution throughout city districts may not be uniform and results may be affected if a given disease afflicts a certain age group to a greater extent than other age groups. For this analysis, age distribution was assumed to be equal among districts, variances in diagnosis rates were assumed to be reasonably equal, and the distributions of diagnosis rates were assumed to be relatively normal. The confounding effects of socioeconomic status, temperature, and other factors that influence public health were assumed to be negligible (Anderson 1997). We are currently developing techniques to perform time-series analysis on relationships between air/drinking water quality indicators and public health. This will allow us to hold certain confounders in the analysis constant that vary spatially, such as socio-economic conditions.

Drinking Water Quality and Morbidity Rates

The first step in analysis involved the coding of each district as high or low based on median values of turbidity at the tap, turbidity at the treatment facility, nickel, and lead. We selected the median as the threshold values for turbidity as measured at the tap at 1.4 mg/L, turbidity measured at the treatment facility at 0.75 mg/L, nickel at 0.018 mg/L, and lead at 0.0165 mg/L. Coding the data allowed me to overcome the problems we had with non-normal distributions of water quality measurements and a very large amount of variability. The coding is shown in **Figure 11**.

District	**Turbidity (mg/L) at Tap**	**Turbidity (mg/L) at Source**	**Ni (mg/L)**	**Pb (mg/L)**
1-Antaviliai	Low	High	High	Low
2-Trinapolis	Low	High	High	High
3-Vingis Parkas	High	Low	High	High
4-Bukciai	High	Low	High	High
5-Tupatiskes	High	High	Low	High
6-Kirtimai	Low	Low	Low	Low
7-A.Paneriai	High	High	Low	Low
8-N.Vilnia	Low	Low	Low	Low
9-Sereikises	Low	Low	Low	Low
Threshhold Median Value (mg/L)	>1.4	>0.75	>0.018	>0.0165

Figure 11: Coding of water supply districts as high or low for concentrations of selected water quality indicators.

0-19 Year Age Group		**ANOVA Test**									**Wilcoxon/Kruskal-Wallis Rank-Sum Test*****	
Indicator	**ICD-9 Code**	**Mean of High**	**Mean of Low**	**Difference**	**Upper 95%**	**Lower 95%**	**Std Error**	**DF**	**Prob>t**	**Power**	**of High from mean**	**Probability>ChiSq**
Turbidity at Tap	001-009	8.385	6.676	1.71	3.457	0	0.891	2356	0.0552	0.4831	-5.005	<.0001
Turbidity at Tap	530-537	83.346	44.653	38.693	47.452	29.935	4.465	1311	<.0001	1	6.826	<.0001
Turbidity at Tap	535	77.885	58.245	19.64	30.361	8.918	5.458	538	0.0003	0.9487	2.803	0.0051
Turbidity at Tap	536	124.812	49.795	75.017	91.233	58.801	8.255	538	<.0001	1	9.018	<.0001
Turbidity at Source	001-009	9.615	5.692	3.923	5.665	2.181	0.888	2356	<.0001	0.993	11.561	<.0001
Turbidity at Source	530-537	89.869	39.428	50.441	59.022	41.859	4.374	1311	<.0001	1	11.094	<.0001
Turbidity at Source	535	98.444	41.798	56.65	66.377	46.915	4.954	538	<.0001	1	10.418	<.0001
Turbidity at Source	536	120	53.645	66.3542	82.839	49.869	8.392	538	<.0001	1	8.392	<.0001
Nickel	590	7.715	6.96	0.756	4.472	0	1.892	538	<.69	0.068	9.848	<.0001
Lead	740-759	2.266	1.429	0.838	1.257	0.418	0.214	4821	<.0001	0.975	8.223	<.0001

0-19 Year Age Group		**Accept or Reject Null Hypothesis based on an alpha-level of 0.05 and a beta-level of 0.10 in the ANOVA Test, and an alpha-level of 0.05 in the Wilcoxon/Kruskal-Wallis Rank-Sum Test**		**Percentage Difference in Diagnosis Rates in Districts with Higher Concentrations of Contaminants***		
Indicator	**ICD-9 Code**	**Accept/Reject Null Hypothesis (that diagnosis rates are not higher)**	**Hypothesis Number****	**Percentage Higher**	**Upper 95% CI of %**	**Lower 95% CI for %**
Turbidity at Tap	001-009	Accept	1	**NA**		
Turbidity at Tap	530-537	Reject	1	**87%**	**106%**	**67%**
Turbidity at Tap	535	Reject	1	**34%**	**52%**	**15%**
Turbidity at Tap	536	Reject	1	**151%**	**183%**	**118%**
Turbidity at Source	001-009	Reject	2	**69%**	**100%**	**38%**
Turbidity at Source	530-537	Reject	2	**128%**	**150%**	**106%**
Turbidity at Source	535	Reject	2	**136%**	**159%**	**112%**
Turbidity at Source	536	Reject	2	**124%**	**154%**	**93%**
Nickel	590	Accept	3	**NA**		
Lead	740-759	Reject	4	**59%**	**88%**	**29%**

*Most interesting in this table is the percentage difference in diagnosis rates (plus 95% confidence intervals) between the districts with lower concentrations of contaminants and the districts with higher concentrations of contaminants.

**Specifics of the hypotheses are described in the text.

***As mentioned in the text, the Wilcoxon/Kruskal-Wallis Rank-Sum Test was an extra measure of security to avoid coming to an incorrect decision about rejecting the null hypothesis, given the potential of violating assumptions of normality in the distributions and equal variances. The null hypothesis had to be rejected for both tests in order for me to accept the alternative hypothesis.

Figure 12: Results of ANOVA and Wilcoxon/Kruskal-Wallis Rank-Sum tests for correlations between water quality indicators and morbidity rates.

Second, based on districts grouped high and low for each water quality indicator, an ANOVA test was performed to determine if there was a significant difference between the disease rates of the high group and disease rates of the low group. Additionally, a Wilcoxon/Kruskal-Wallis Rank-Sum test was performed to test for differences between the high and low groups. This test was chosen given that the public health data are not normally distributed based on the Shapiro-Wilk test and possess significantly unequal variances. The results of these analyses are shown in **Figure 12**.

For a good description of statistical techniques, please see (Wassertheil-Smoller 1995; Zar 1999).

IV. Discussion, Findings & Recommendations:

Air Quality and Morbidity Rates

CO, NO_2, and SO_2 act in a variety of ways to weaken the body's ability to function properly. CO is known to lead to these health effects: "aggravation of angina pectoris and other aspects of coronary heart disease, decrease exercise tolerance in persons with peripheral vascular disease and lung disease, impairment of central nervous system functions, and possible increase risk to fetuses (South Coast Air Quality Management District 1999)." NO_2 is known to lead to these health and other effects: "potential to aggravate chronic respiratory disease and respiratory symptoms in sensitive groups, risk to public health implied by pulmonary and extra-pulmonary biochemical and cellular changes and pulmonary structural changes, and contribution to atmospheric discoloration." SO_2 is known to lead to these health effects: "bronchoconstriction accompanied by symptoms which may include wheezing, shortness of breath and chest tightness, during exercise or physical activity in persons with asthma."

The results of the study showed several interesting relationships, none perhaps stronger than the relationship of air pollution indicators to several chronic respiratory infections. In particular, Chronic pharyngitis and nasopharyngitis (472), Chronic laryngitis and laryngotracheitis (476), Bronchitis, not specified as acute or chronic (490), Chronic bronchitis (491) showed low p-values and high r^2 values.

Specifically, we rejected the API null hypothesis $H_{O(1)}$ for diseases 472, 476, 480, 490, and 491 for the 0-100 year age group, and 491 in the 0-19 year age group. In doing so, we accepted our alternative hypothesis that increases in the API correlate to increases in morbidity rates. Additionally we rejected the CO null hypothesis $H_{O(2)}$ and accepted the alternative hypothesis for diseases 410, 480, and 491 for the 0-100 year age group, and for disease 460 in the 0-19 year age group. The NO_2 null hypothesis $H_{O(3)}$ for diseases 410, 480, and 491 was rejected for patients aged 0-100, and disease 491 was rejected for those aged 0-19. The SO_2 null hypothesis $H_{O(4)}$ was rejected for the 0-100 year age group for diseases 472 and 491; in the 0-19 year age group, the null hypothesis for disease 491 was rejected.

Although this finding is documented in other cities, this study helps highlight the issue in Vilnius. CO, NO_2, and SO_2 lead to an increase in disease, especially when they are present in concentrations above hygienic norms. To illustrate the extent of the influence, an average rate of chronic bronchitis (491) was calculated for city districts with an API of "unpolluted" and for those with an API of "very polluted." The average rate for the unpolluted districts was 125 per 100,000 for the 0-19 year age group, and in the polluted districts, the rate was 216. That means that residents of the very polluted city districts had a 72% greater rate of infection of chronic bronchitis than residents of the unpolluted districts. A similar percentage increase was also found for the 0-100 year age group.

Another interesting finding is the apparent relationship between Acute myocardial infarction (410) and levels of CO ($r^2 = 0.23$) and NO_2 ($r^2 = 0.22$) in the 0-100 year age group. This finding holds consistent with prior epidemiological studies.

Although this is a preliminary study, the relationships found show that a strong influence is being exhibited by air quality on public health. These findings merit further examination and bring attention to the need for cities such as Vilnius to take strong steps to bring air quality into compliance with government norms. It is further hoped that other cities in Lithuania will begin to take the same steps toward more effective monitoring and management of air quality.

Drinking Water Quality and Morbidity Rates

In order to reject the null hypothesis, the districts with high levels of a given water quality indicator had to test significantly higher for a given disease rate at an α-level of 0.05 and a β-level of 0.10 in the ANOVA test. β is the chance that, if the null hypothesis is actually false, and the true distribution differs from it by some amount δ, you will mistakenly accept it. In **Figure 12** you will also find the Power of the analysis which is 1-β (therefore the Power of the test must be greater than 0.9 in order to reject the null hypothesis). Power is the chance of detecting a difference of size δ. These districts also had to test higher and be significant in the Wilcoxon/Kruskal-Wallis Rank-Sum test at an α-level of 0.05. This was an extra built-in measure of security to reduce the risk of coming to an incorrect conclusion, given the non-normal nature of the distributions.

Based on these standards, we rejected $H_{O(1)}$ and accepted $H_{A(1)}$ that districts with high turbidity as measured at the tap do have higher instances of gastritis and duodenitis (535), disorders of stomach function (536), and diseases of the esophagus, stomach, and duodenum (530-537). We further rejected $H_{O(2)}$, and accepted $H_{A(2)}$ that districts with high turbidity as measured at the treatment facility do have higher instances of ICD-9 diseases 001-009 (grouped), 530-537 (grouped), 535, and 536. The ANOVA test showed a sizable difference in the means of the high and low groups in several instances. For example, for diseases of the esophagus, stomach, and duodenum (530-537) there were 83 diagnoses per 100,000 per month in the high turbidity districts, versus 47 per month in the low.

With respect to nickel, we found no significant difference in both the ANOVA test and the Wilcoxon/Kruskal-Wallis test, thereby not refuting $H_{O(3)}$. For lead, we did find variability among the groups that was more than one would expect to find by chance. Rejecting $H_{O(4)}$, we accepted $H_{A(4)}$ that districts with higher lead concentrations as measured at the treatment facility do have higher instances of congenital anomalies (740-759 grouped).

In summary, districts with higher turbidity concentrations have 34%-151% higher instances of gastrointestinal disorders and districts with higher lead concentrations seemed to have 59% higher instances of congenital anomalies for the 0-19 year age-group. It is hoped that these preliminary findings will encourage Vilnius Municipality to seek ways to protect drinking water quality and educate its citizens about the influence of water quality on public health.

Final Remarks & Recommendations

The purpose of this paper was twofold: (1) to show how public health and environmental data can be effectively combined in a GIS framework to aid government decision-making, and (2) to demonstrate the general application of this system to air and drinking water quality in Lithuania. Vilnius and other cities in the Baltic region have great environmental and public health challenges but also great opportunities for improvement. We believe fulfillment of the following recommendations will help to remedy the situation:

- A plan for complying with air and water pollutant standards at the municipal level should be adopted (Van Leeuwen 1997; Ministry of Health Protection 1998a; Ministry of Health Protection 1998b).
- Monitoring of both air and drinking water quality should be continued in Vilnius and expanded to other cities in Lithuania with potential air and drinking water pollution problems.
- Residents should be supplied with information about air and drinking water quality conditions, how to minimize their health risk from poor air quality, and how to reduce their contribution to pollutant levels (Hallo 1997; Regional Environmental Center for Central and Eastern Europe 1999).
- Data exchange and dissemination to government agencies, the public, and research institutions should be a priority of agencies collecting data.

V. References:

Agency for Toxic Substances and Disease Registry (1993). Toxicological profile for lead. Department of Health and Human Services, Public Health Service, Atlanta, 3.

Agency for Toxic Substances and Disease Registry (1996). Toxicological profile for nickel (update). Department of Health and Human Services, Public Health Service, Atlanta, 3.

Air Quality Management Group (1999). Vilnius Air Quality Management. Internet: Ministry of Environmental Protection, Lithuania, 1999.

Anderson, R, P Sorlie, E Backlund, N Johnson & G Kaplan (1997). “Mortality Effects of Community Socioeconomic Status.” Epidemiology **8**(1): 42-47.

Brunekreef, B, N Janssen, J Hartog, H Harssema, M Knape & P Vliet (1997). “Air Pollution from Truck Traffic and Lung Function in Children Living Near Motorways.” Epidemiology **8**(3): 298-303.

Feshbach, M (1995). Ecological Disaster: Cleaning up the hidden legacy of the Soviet Regime. New York, The Twentieth Century Fund Press.

Hallo, RE (1997). Experts Corner 1997/1: Public Access to Environmental Information. European Environmental Agency, Copenhagen, 1-46.

Jedrychowski, W, U Maugeri, E Flak , E Mroz & I Bianchi (1999). “Predisposition to Acute Respiratory Infections Among Overweight Preadolescent Children: An Epidemiologic Study in Poland.” European EpiMarker **3**(2): 1-7.

Kadunas, K, A Domesevicius, A Daubaras & D Yeskis (1997). A Preliminary Study of Ground-water Quality of the City of Siauliai. U.S. Environmental Protection Agency, Lithuanian Geological Survey, Lithuanian Ministry of Environmental Protection, Chicago, 1-50.

Klimas, A (1998). Vilniaus Miesto Vandenvieciu Eksploatacijos ir Vandens Tiekmo Sistemos Optimizacija. Vilniaus Hidrogeologija, Vilnius, 1-124.

Ministry of Health Protection (1998a). HN 24-1998: Hygienic Norm 24-1998, Drinking Water: Quality Requirements and Monitoring. Ministry of Health Protection, Vilnius, 1-28.

Ministry of Health Protection (1998b). HN 33-1998: Hygienic Norm 33-1998, Air Quality Standards. Ministry of Health Protection, Vilnius, 1-33.

Pershagen, G, E Rylander, S Norberg, M Eriksson & SL Nordvall (1995). “Air Pollution Involving Nitrogen Dioxide Exposure and Wheezing Bronchitis in Children.” International Journal of Epidemiology **24**: 1147-1153.

Regional Environmental Center for Central and Eastern Europe (1999). Access to Information, Public Participation in Decision-Making and Access to Justice in Environment and Health Matters. Third European Ministerial Conference on Environment and Health, London.

Schwartz, J, & others (1997). “Drinking Water Turbidity and Pediatric Hospital Use for Gastrointestinal Illness in Philadelphia.” Epidemiology **8**(6): 9.

South Coast Air Quality Management District (1999). 1997 Air Quality Management Plan. Internet: State of California, 1999.

Statistical Office of Vilnius (1998). Vilnius in Figures, 1996. Vilniaus Apskrities Statistikos Valdyba, Vilnius, 1-103.

StatSoft Inc. (1998). Electronic Textbook. Internet: StatSoft, Inc, 1999.

Van Leeuwen, J, & F Rolaf (1997). “WHO Air Quality Guidelines for Europe.” European EpiMarker **April 1997**: 1-3.

Wassertheil-Smoller, S (1995). Biostatistics and Epidemiology: A Primer for Health Professionals. New York, Springer-Verlag.

Wojtyniak, B, D Rabczenko & T Piekarski, (1997). “Short-term Effect of Air Pollution on Mortality in Poland.” European EpiMarker **July 1997**: 1-3.

World Resources Institute (1999). Why the Increase in Asthma? World Resources Institute, 1999.

Zar, JH (1999). Biostatistics. Upper Saddle River, New Jersey, Prentice Hall.

Zickus, M (1999). Influence Of Meteorological Parameters On The Urban Air Pollution And Its Forecast. Vilnius, Vilnius University.