Exploring Distribution of Uranium in Ukraine: Geovisualization and Spatial Statistics

Michael Govorov*, Victor Putrenko, Gennady Gienko*****

* Vancouver Island University, Canada
** Institute of Geography, National Academy of Sciences of Ukraine, Ukraine
*** University of Alaska Anchorage, USA

Abstract. Many powerful geovisualization and spatial statistical methods are available to reveal spatial patterns in distribution of various environmental indicators. This paper discusses several cartographic and geostatistical techniques to explore patterns in spatial distribution of uranium in Ukraine and identify major factors contributing in particular distributions. The study resulted in a series of maps reflecting association between uranium and several environmental indicators.

Keywords: Geovisualization, spatial statistics, environmental mapping

Introduction

The problem of drinking water is relevant for many Ukrainian regions. Quality of drinking water, determined by its chemical and biological content, depends on several anthropogenic and natural factors, including natural radioactive elements such as uranium. Uranium concentration higher than 0.08 Mg/L is potentially dangerous to human health. Multiple studies suggest that the geological structure is the main natural factor which determines the content of natural radionuclides in ground and surface waters (Skeppström & Olofsson 2007). The radioactivity of water in rivers is proportional to mineralization of water, radioactivity of rocks, and physical and chemical properties of water. It is also influenced by hydrological and climatic conditions. Major environmental and natural indicators affecting concentration of uranium in surface waters can be grouped into four categories: geological structure and geomorphology, geochemical, climate, and mineralogical (Salih et al. 2002).

Geological and geomorphological factors include relief of the territory (elevation), structure of the relief (slope, exposure), and deepness of bedrock. Geochemical factors are characterized by distribution of rare earth elements in the water, which often found in conjunction with uranium. These include arsenic, fluoride, chromium, copper, and zinc. Climatic factors are characterized by average annual rainfall, the average annual values of temperature and evaporation. Mineralogical indicators are divided into two groups: characteristics of water and soil. These indicators are the product of interaction between climate and geological structures. Mineralogical indicators of water include common indicators of mineralization, hardness, chloride, sulphate, magnesium, and calcium. Concentration of humus was used to characterize soil.

Figure 1 illustrates the distribution of uranium in Ukraine. Two different symbolization methods for categorization of the source data were used to illustrate spatial distribution of uranium. While the areas of low concentration (below 0.04 Mg/l) and transitional types of uranium have been clearly identified using quintile categorization (*Figure 1*, left), geometric interval-based categorization can be used to illustrate the distribution of high levels of uranium concentration (0.04-0.145 Mg/l, *Figure 1*, right).

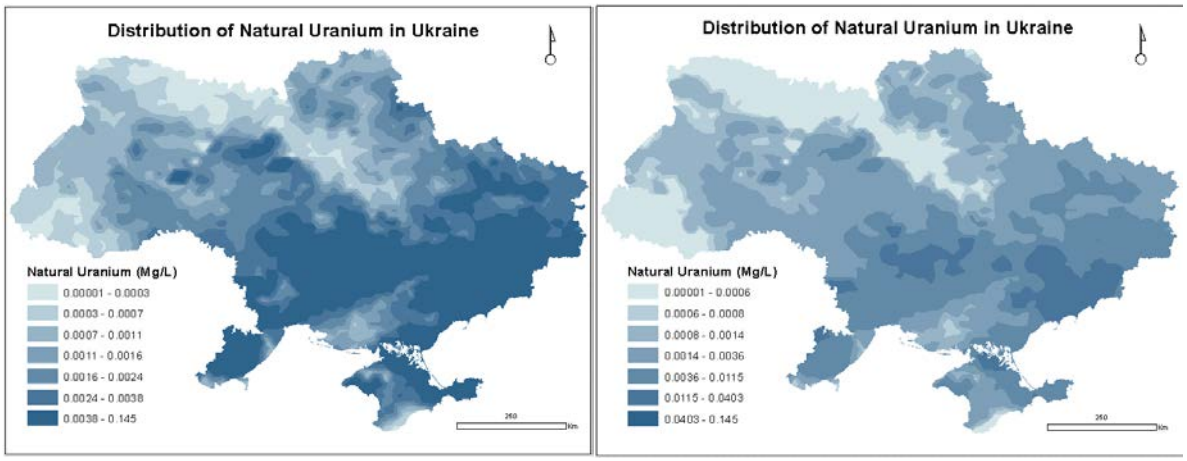


Figure 1. Distribution of natural uranium in Ukraine. Categorization parameters: quantile (left), and geometrical interval (right). Data courtesy of State Enterprise "Kirovgeologiya" (Держкомприродресурсів України 2004; Макаренко 2000).

The goal of this study is to explore various geostatistical methods to model spatial dependences between several environmental variables and distribution of uranium in surface waters. The study is focused on comparative analysis of several techniques to identify the most robust method to describe spatial distribution of uranium in surface waters Ukraine. The analysis was implemented using tools available in SPSS, ArcGIS, and GeoDa packages.

Methodology

The study is based on the results of geological surveys in Ukraine carried out by State Enterprise "Kirovgeologiya" (Держкомприродресурсів України 2004; Макаренко 2000). The database consists of 23 environmental indicators collected in 6546 points in Ukraine and neighboring territories of Russia, Belarus, and Moldova. Several spatial statistical methods were used to discover patterns of spatial distribution of natural uranium. (Note on terminology: the term 'natural variables', or 'environmental indicators', is widely used in this study to describe various natural and environmental factors, for example, mineralization, precipitation, relief, and many others. Some models use 'predictors' with the same meaning as 'variables'.)

The following spatial analyses and cartographic visualization workflow was utilized to describe the impact of several environmental variables on spatial distribution of uranium.

1. Exploratory spatial data analysis (ESDA) is used to check for statistical distribution, linearity, multicollinearity, and presence (or absence) of a pattern (both in spatial and non-spatial domains). Histogram of uranium concentration (*Figure 2*, left) shows that statistical distribution of the source data does not exhibit normality. However, logarithmic transformation brings the dataset closer to the normal (*Figure 2*, center and right).

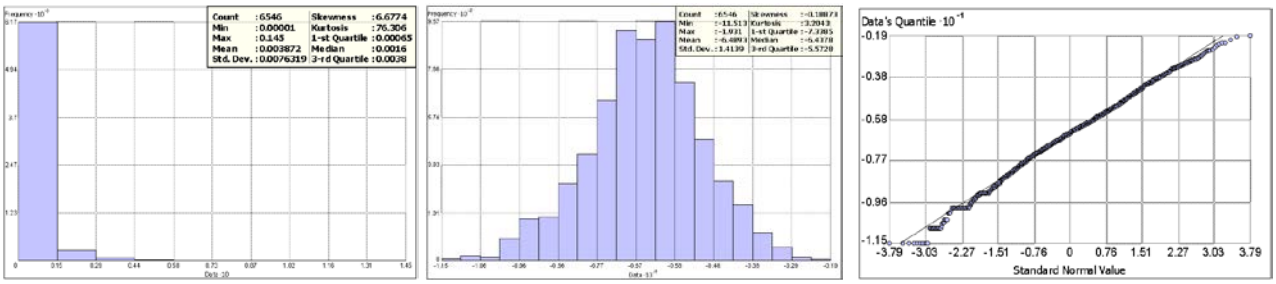


Figure 2. Histograms of the source (left), and normalized data (right), and normal QQPlot after applying logarithmic transformation (right).

For linear regression models the relationship between a dependent variable and each independent variable should be linear. For example, relationship between uranium and mineralization of water

is more exponential than linear, requiring the use of nonlinear regression (However, implementation of multiple nonlinear regression for these variables can be problematic.)

Independent variables can be significantly correlated (multicollinear) and produce unstable and unreliable multiple regression models. Several variables in the source dataset exhibit multicollinearity and require appropriate treatment.

Spatial autocorrelation methods were used to identify patterns in spatial measurements of uranium concentration. According to Moran's I and Getis-Ord analysis (Getis & Ord 1992), the distribution of uranium can be described as highly clustered with statistical significance. Moran's I index is 0.5475 (p-value = 0.0) and Observed General G = 0.00007 (p-value = 0.0). Thus, spatial patterns in observations of uranium should be taken into account.

2. In the second step, quantitative measure of global correlation is used to confirm or reject several hypotheses of relationship between the dependent variable (uranium) and independent variables (environmental indicators). The analysis should identify natural variables which define high concentration of uranium, taking into account multicollineaity and clustered nature of the data.

3. In the next step, global factor analysis is applied as an attempt to identify underlying indicators and factors that explain the pattern of correlations within a set of natural variables. Factor analysis is used to identify a smaller number of natural variables that explain most of the variance of uranium distribution. This study utilizes factor analysis based on the principal components using Varimax rotation with Kaiser normalization (Harman 1976).

4. The most significant natural variables, identified from the factor analysis, were used to demonstrate spatial non-stationarity of correlation between uranium and the natural variables. This was done by implementing local correlation analysis (or geographically weighted correlation, GWC). In this study, a local form of bilinear regression with the optimized bandwidth was used to model spatially varying relationships between uranium and natural variables.

5. Relationships between the dependent variable and all explanatory variables, identified from the factor analysis, were modeled by using stepwise linear multivariate regression analysis (or ordinary least squares regression, OLS regression). Then several most significant explanatory variables identified from the stepwise multivariate regression analysis were used to build local linear multivariate regression models (or geographically weighted regression, GWR with the optimized bandwidth (Fotheringham et al. 2002)).

6. Finally, local factor analysis (or geographically weighted factor analysis, GWFA) with the optimized bandwidth, was used to generate linear multivariate regression models for the dependent variable (uranium) and six major factors produced in the global factor analysis.

Global and Local Spatial Correlation Analysis of Distribution of Uranium

The overall distribution of uranium depends on several factors defined by geographical conditions of the area. The most significant factors can be identified using global correlation of uranium with all indicators from the four defined groups: geological, geochemical, climatic, and mineralogical.

The highest global correlation coefficients of uranium were obtained for humus (r=0.52), temperature (r=0.51), precipitation (r=-0.50), and volume of natural groundwater resources (r=-0.56). Uranium also has significant correlation with the overall water mineralization (r=0.49) and its components: SO4 (r=0.45), Cl (r=0.44), and hardness of water (r=0.49). These indicators are inter-dependent and highly correlated. Geomorphic indicators, which include relief (r=-0.24), sediment thickness (r=0.02), and slope of terrain (r=-0.08), were less associated with the distribution of uranium. Although there is a direct association between granite rocks represented in relief uplifts and uranium concentration in water, correlation analysis identifies certain relationship between uranium and mineralization of water and variations in climatic conditions.

Local correlations for different indicators can form complex spatial patterns. For example, Moran's *I* and Getis-Ord analyses indicate that the pattern of uranium is highly clustered. Thus, values of

correlation coefficients inherit high nonstationarity and should be modeled by using local methods. For example, the coefficient of global correlation shows very low relationship between uranium and isopachs (only r=-0.02), but local coefficients of correlation range from -0.71 up to 0.42 which indicates strong relationship between these two variables in Bessarabia region (*Figure 3*).

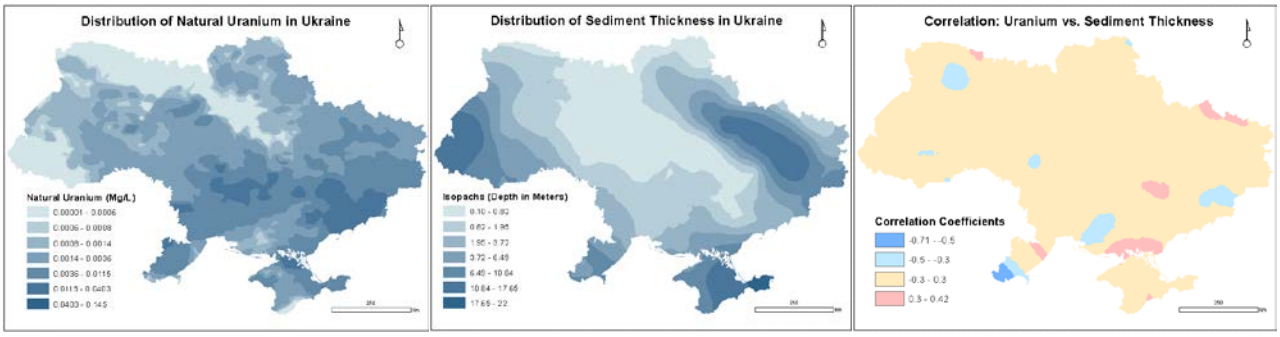


Figure 3. Uranium (left), isopachs (center), and their correlation (right). This is an example of low global correlation (r=-0.02) with high values in local variations (r=-0.70, blue, and r=-0.42, red).

Global Factor Analysis of the Relationship between Natural Variables and Uranium

Factor analysis has been used to find the input of a particular variable into distribution of uranium. Analysis of 23 different natural variables has revealed six principle components which shape the distribution of uranium in Ukraine (*Table 1*).

	Component						
	Global r	**1**	**2**	**3**	**4**	**5**	**6**
Explained variance (cumulative) %		30.2	44.2	54.1	62.3	68.0	72.5
SO_4	.448	0.952	0.139	-0.092	-0.045	0.1	0.091
Mineralization of water	.493	0.945	0.108	-0.159	-0.003	0.132	0.154
Hardness of water	.493	0.937	0.035	-0.173	0.01	0.062	0.172
Cl	.445	0.918	0.07	-0.147	0.016	0.148	0.128
NO_3	.269	0.629	-0.177	0.198	0.368	0.354	-0.061
Cu	-.067	0.027	0.886	0.101	0.074	0.043	0.038
Fe	-.148	0.012	0.851	0.286	-0.002	0.07	-0.01
Mn	.118	0.392	0.716	0.008	-0.091	0.088	-0.004
Zn	.007	-0.101	0.608	-0.417	0.103	-0.096	0.103
Precipitation	-.497	-0.328	0.243	0.762	0.16	-0.156	-0.276
Relief	-.244	-0.148	0.077	0.703	0.074	-0.014	-0.064
Slope	-.078	-0.011	-0.051	0.617	-0.139	0.235	0.091
Temperature	.507	0.434	-0.347	-0.521	-0.332	0.293	0.217
NH_4	-.117	-0.025	0.036	0.165	0.816	-0.162	-0.037
NO_2	.288	0.525	-0.009	-0.126	0.67	0.228	0.062
PO_4	.166	0.04	0.211	-0.327	0.627	0.331	0.199
Cr	-.390	-0.346	-0.228	0.25	0.589	-0.485	-0.215
Isopach	.023	0.13	0.083	0.287	0.071	0.704	-0.146

Humus	.521	0.372	0.058	-0.075	-0.021	0.648	0.306
Volume of natural groundwater resources	-.558	-0.431	0.26	0.384	0.264	-0.507	-0.29
HCO_3	.160	0.129	-0.017	0.077	0.038	-0.23	0.784
F	.353	0.207	0.048	-0.228	-0.039	0.128	0.546
As	.229	0.032	0.052	-0.044	0.023	0.298	0.473

Table 1. Principal components of environmental indicators.

The first major component is characterized by mineralization of the water with the most important being the total salinity and hardness of water. The second component largely comprises metals dissolved in water (copper, iron, magnesium, and zinc). The third component describes climatic conditions of the territory and formation of ground water. The fourth component is associated with the presence of organic compounds in water. The fifth component describes geomorphological characteristics, and the sixth component is associated with the mineral compounds and metals, often accompanying uranium ('satellites').

The analysis shows that all six identified principal components largely coincide with the four groups of natural variables, outlined in the hypothesis of uranium distribution in Ukraine (*Table 2*).

It should be mentioned, that factor analysis on all variable including uranium shows that uranium belongs to the first mineralogical group.

Local Spatial Correlation

Results of factor analysis were used to explore patterns of local spatial correlation of a particular indicator with uranium. The local correlation analysis was conducted for the most significant element from each component , for example, for Component 1 the following considerations were made.

Uranium exhibit positive correlation with total mineralization of water reaching r=0.56 in certain areas (*Figure 4*). The relationship strengthens in the north, center, and south of Ukraine, and weakens in the forest-steppe zone. Both mineralization and uranium are less exposed in the northern and western parts of Ukraine but still exhibit relatively high correlation. Amount of mineralization of water in the forest-steppe zone is increased, but the uranium does not show the same pattern, which results in wakening the correlation. High concentration of uranium in the area of the Ukrainian Crystalline Shield contributes into higher correlation with mineralization of water. In the south, the correlation is increased by improving the performance and uranium mineralization. However, correlation weakens in the western part of the Black Sea Lowland. High mineralization is associated with low values of uranium, which is found in small quantities in these sedimentary rocks.

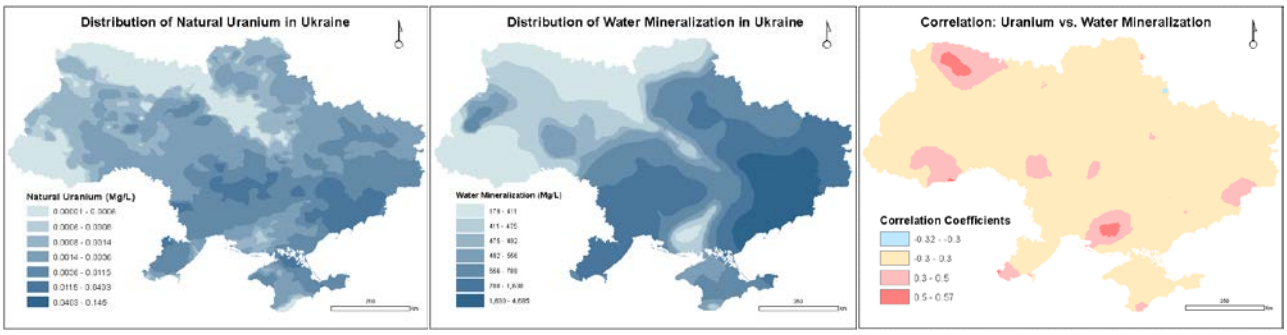


Figure 4. Uranium (left), water mineralization (center), and distribution of local correlation (right). Global correlation r=0.49.

Total mineralization of water is closely related to water hardness, SO4, and Cl. Water hardness exhibits correlation patterns very similar to that of mineralization but with slightly higher maximum values of the local correlation (r=0.59). There is also clearly defined relationship of uranium with folded structures of Donetsk range in the east part of Ukraine, which may be associated with higher natural hardness of groundwater in this area.

Similar analysis was conducted for components 2-6.

Local Spatial Multiple Regression

While the correlation and factor analyses allowed for identification of significant environmental variables which impact the distribution of uranium, the contribution of each factor can be better understood by carrying out a multiple regression analysis. The main steps for the analysis were selection of indicators for multiple regression, assessing their relevance and contribution to the distribution of uranium, and calculation of parameters of regression. *Table 2* outlines environmental variables, chosen as predictors for building several multiple linear regression models. The summary of the multiple regression models is provided in *Table 3*.

While the correlation and factor analyses allowed for identification of significant environmental variables which impact the distribution of uranium in groundwaters in Ukraine, the contribution of each factor can be better understood by carrying out a multiple regression analysis. The main steps for the analysis were selection of indicators for multiple regression, assessing their relevance and contribution to the distribution of uranium, and calculation of parameters of regression. *Table 2* outlines environmental variables, chosen as predictors for building several multiple linear regression models. The summary of the multiple regression models is provided in *Table 3*.

Component	Natural Variables	Group	Predictors
1	Mineralization and hardness of water	Mineralogical	Hardness of water, mineralization of water
2	Metals dissolved in water	Geochemical	Cu, Fe, Cl, Zn
3	Climatic conditions of territory and formation of ground water	Climatic	Precipitation, temperature, humus
4	Organic compounds in water	Mineralogical	NO_3, NH_4, PO_4
5	Geomorphological characteristics	Geological structure and geomorphology	Relief, isopach
6	Mineral compounds and satellite elements of uranium	Geochemical	Bicarbonate, fluoride, arsenic

Table 2. Environmental indicators (predictors), selected for multiple regression analysis.

Model	Predictors	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	precipitation	0.51	0.26	0.26	1.1611
2	1 + humus	0.595	0.354	0.354	1.0849
3	2 + water hardness	0.623	0.388	0.387	1.0564
4	3 + F	0.629	0.396	0.395	1.0496
5	4 + Fe	0.634	0.402	0.402	1.0438
6	5 + As	0.637	0.405	0.405	1.0414
7	6 + SO_4	0.638	0.407	0.407	1.0395
8	7 − water hardness	0.638	0.407	0.407	1.0396
9	8 + isopach	0.641	0.411	0.41	1.0368
10	9 + NH_4	0.642	0.413	0.412	1.0351
11	10 + Cl	0.644	0.415	0.414	1.0332
12	11 + temperature	0.645	0.416	0.415	1.0325

13	12 + NO_3	0.645	0.417	0.416	1.0317
14	13 + HCO_3	0.646	0.418	0.416	1.0310
15	14 + Zn	0.647	0.418	0.417	1.0305
16	15 + Cu	0.647	0.419	0.418	1.0300
17	16 + PO_4	0.648	0.419	0.418	1.0296
18	17 + mineralization	0.648	0.42	0.419	1.0288

Table 3. Multiple regression model summary.

A global linear regression model has been used to describe contribution of several environmental variables into spatial distribution of uranium in ground waters. Calculation of global regression was followed by spatial local multiple regression or GWR.

Fifteen multiple regression models were built incrementally using predictors outlined in *Table 3*. The first five most significant predictors (precipitation, humus, water hardness, F, and Fe) contribute 40.2% into the overall model. Examination of individual and cumulative contribution of environmental variables is provided below.

Analysis of environmental relationships has indicated that the most significant environmental variable is precipitation. The second predicting variable is humus content in the topsoil. Although this indicator is not a major factor in determining the distribution of uranium, it has a high degree of association to conditions of the geochemical migration of uranium. Soil is a product of interaction of all components of the environment and has direct relationship with underlying rocks, mineral composition, climate, and biota. As a result, the humus content in the soil reflects climate and geochemical characteristics of migration of chemical elements, including uranium. Local r-square coefficients for multiple regression model, built only on precipitation variable, are shown in *Figure 5*, left. Adding the humus component and then hardness of water improves the model (*Figure 5*, center and right correspondingly).

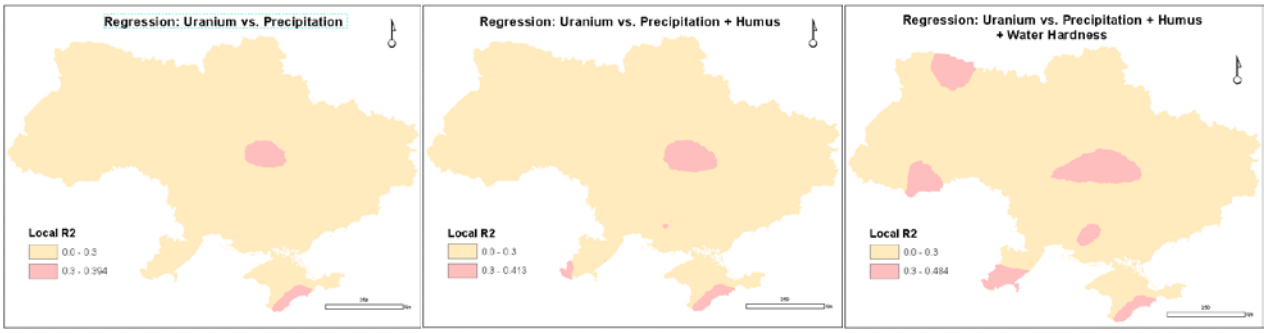


Figure 5. Local R^2-square coefficients for geographically weighted regression models: precipitation (left), precipitation + humus (center), precipitation + humus + water hardness (right).

Ftor (F), iron (Fe), and arsenium (As) add only 1.7% in variability of the data, but still contribute into cumulative regression model (*Figure 6*). None of the rest of environmental variables contributed more than 0.2% into the final model, which indicates their very low impact on the total distribution of uranium. Partially, this can be explained by the fact that many of these variables depend on each other, and thus, are highly correlated.

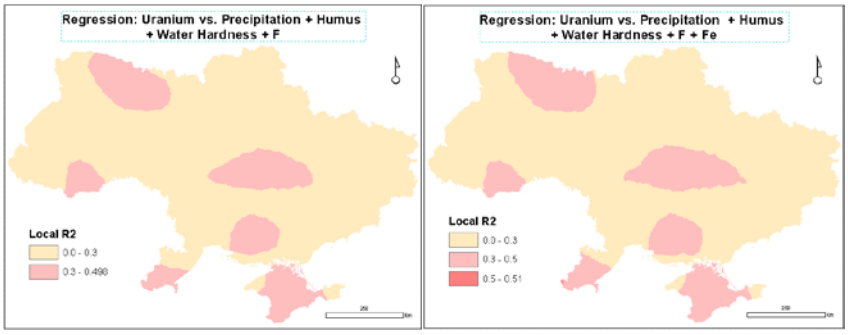
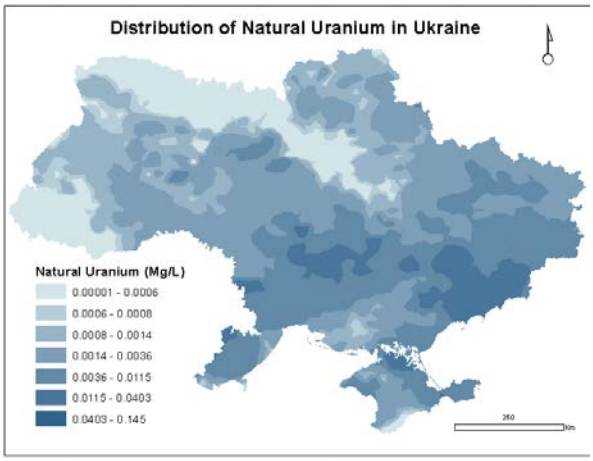


Figure 6. Local R^2-square coefficients for geographically weighted regression models: precipitation + humus + water hardness + F (left); precipitation + humus + water hardness + F + Fe (right).

Geographical analysis of results of local multiple regression indicates that the impact of precipitation and humus in distribution of uranium is mostly noticeable in the south areas of Ukraine. Inclusion of hardness of water in the model spreads the relationship in the center and north-west. Adding one more variable (F) improves the model and extends it to the central part of the Ukrainian Crystalline Shield and in the south-west. The multiple regression with all six factors strengthens the model, especially in the south and north-west. The local correlation achieved r=0.51, which confirms that the inclusion of all six indicators in the model was sensible. It is also worth mentioning that several variables have shown high local correlation in spatial multiple regression model. In particular, fluoride and magnesium have high degree of correlation with uranium in the south of Ukraine.

Multiple Regression Based on Principle Components

Multiple regression analysis in the previous section was based on selection of the most significant environmental variables (predictors), and consecutively adding those individual predictors to incrementally improve the model. An alternative approach is to build the models using principal components. In this approach, all variables constituting the first group of principle components (Component 1, *Table 1*), are used to create the first multiple regression model. The model is further improved by adding all elements forming the second groups (Component 2). The final regression model will constitute all six principal components (*Figure 7*). *Figure 8* illustrates successive steps in building the multiple regression models using principal components.



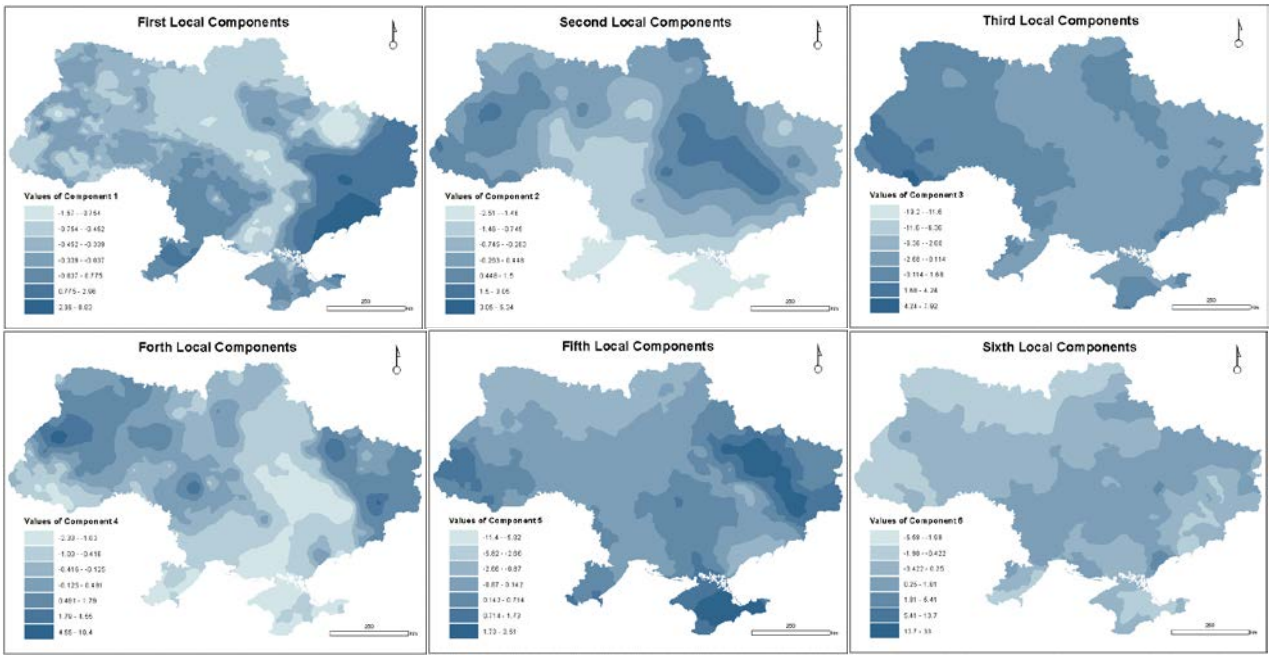


Figure 7. Uranium (top), and maps of six local principal components.

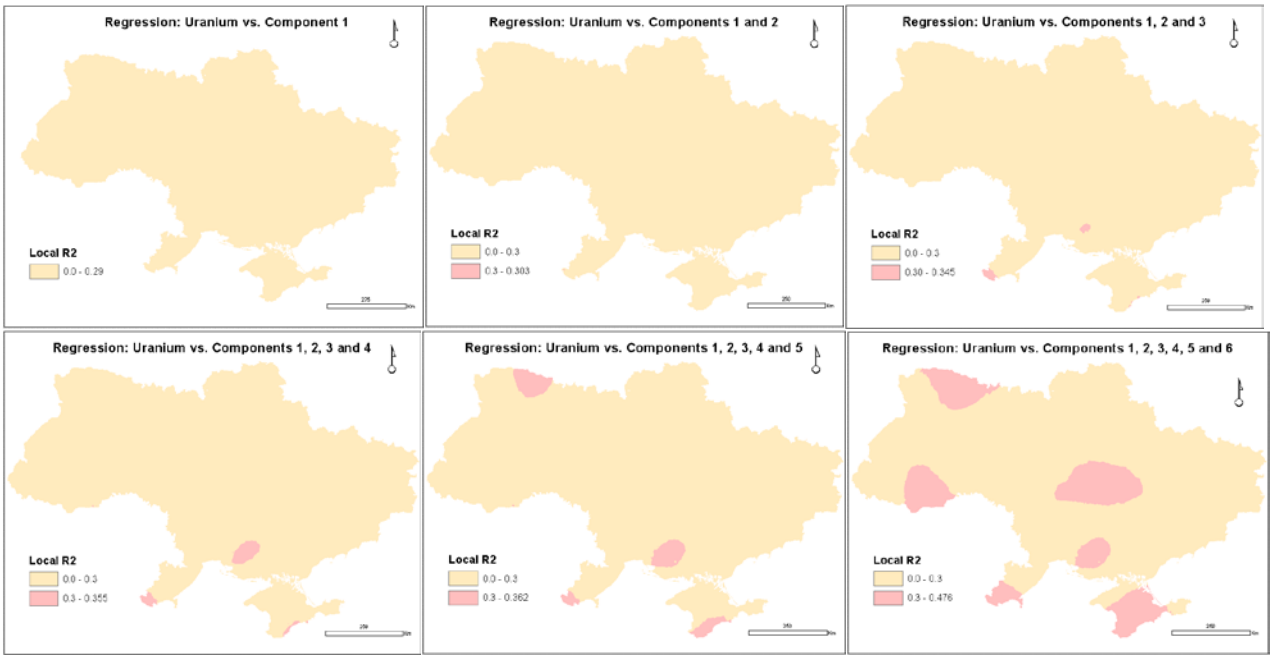


Figure 8. Building consecutive multiple regression models using principal components. First map demonstrates correlation of uranium with component 1, and the last map illustrates correlation with all six principal components, incorporated in one regression model.

Comparison of two multiple regression models based on the six principal components and the six individual environmental indicators shows that in general, both models indicate similar associations with the distribution of uranium in ground waters. However, the maximum coefficient of local correlation for the model, based on the principal components, is only 0.48, while the same correlation for the latter model is 0.51. This indicates that the zones of high regression are identified objectively, and the model of the five environmental variables shows stronger local associations comparing to the component model which takes into account all studied indicators.

Conclusion

Several spatial statistics and cartographic visualization methods were used in this study to explore spatial distribution of uranium in groundwater in Ukraine. Factor analysis and correlation matrices

revealed six principle components and respective environmental variables which explain nearly 72.5% of the variability in the original 23 independent variables. These components include total mineralization of water, climatic conditions, metals, anions nitrate and phosphate, geomorphology of terrain, and chemical elements, often accompanying uranium. Multiple regression model defines fifteen environmental variables that describe 42% of the variability of the data. The first six most significant predictors (precipitation, humus, water hardness, F, Fe, and As) contribute 40.5% into the overall model. Local spatial correlation and multiple regression analyses show significant spatial differentiation of factors influencing the distribution of uranium. The use of local correlation improves assessment of relationships between environmental variables. Local multiple regression was used to estimate contribution of individual environmental variables.

Outcomes of different models sometimes do not support each other, e.g., some explanatory variables have low correlation with the dependent variable but at the same time have high percentage of explained variance in factor analysis. Global spatial regression modeling in a large-scale spatial analysis can be unsuitable for the local inference. The modeling results including their cartographic representations remain mainly descriptive and require interpretation by application experts.

Further research is envisioned in refining relationships between the environmental indicators and improving numerical forecasts by expanding the range of applied spatial statistical methods. The study is planned on exploring econometric models and spatial-clustering techniques to improve the robustness of the developed statistical model.

References

Fotheringham SA, Brunsdon C, Charlton M (2002) *Geographically Weighted Regression: the analysis of spatially varying relationships,* John Wiley & Sons

Harman HH (1976) Modern Factor Analysis, 3rd ed. Chicago: University of Chicago Press.

Salih I, Pettersson H, Sivertun Å, Lund E (2002) Spatial correlation between radon (222Rn) in groundwater and bedrock uranium (238U): GIS and geostatistical analyses. J. of Spatial Hydrology Vol.2 No.2

Skeppström K, Olofsson B (2007) Uranium and radon in groundwater. An overview of the problem, European Water 17/18: 51-62

Держкомприродресурсів України, Департамент геологічної служби, Казенне підприємство „Кіровгеологія” (2004) Природні та антропогенні джерела формування радіоактивності природних вод України та радіаційний захист населення, Киев, 32 с.

Макаренко ММ (2000) Оцінка природних і техногенних факторів забруднень підземних і поверхневих вод природними радіонуклідами навколо уранових родовищ України. Інформаційний бюлетень про стан геологічного середовища України, Київ, С.102-111