Using GIS and Regression Analysis to Evaluate Physical Factors of Radon Concentration

Nana Yaw Owusu-Amponsah
Department of Resource Analysis, Saint Mary’s University of Minnesota, Minneapolis, MN 55404

Keywords: Radon, Soil, Geologic Unit, Slope, Elevation, Lithology, Soil Permeability, Regression Analysis

Abstract

Radon is a concern for many home owners and real estate developers in North America due to its cancer causing characteristics. Radon can move freely in soil through mediums such as water or gas and can escape from the earth into the air. Whiteside, Carroll, Lee, and Olge Counties in Illinois are known to have high radon levels. This study explores correlation between physical factors such as geology, soil permeability, elevation, and slope and the radon levels within the study area. Due to data sensitivity, radon measurements were only available summarized by zip code. A suitability map was made to predict radon concentration and tested by the use of regression analysis to evaluate the predictive nature of the physical factors. The resulting multiple regression model predicted 16% of radon concentrations, with geology being the best predictor of all the variables. Although not accounting for a large percentage of radon variability, geology proved to have a significant relationship. Elevation, soil permeability, and slope were weak predictors of radon concentration.

Introduction

According to Jelle (2012), radon is the decay product of radium, which is a member of the uranium series. It is a colorless, odorless, radioactive gas that forms from the decay of naturally occurring uranium (Frumkin and Samet, 2001). The noble gas radon can move through soil pores and migrate into the ground surface (United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR), 2000).

Because this gas has a single atom, it has a very high penetrative ability. Radon can move rather freely through porous mediums such as soil or fragmented rock and creep into dwelling places where it accumulates in homes (Moreno, Bach, Baixeras, and Font, 2013). Radon easily seeps its way into common materials such as leather, plastic bags, paper, wood paneling, concrete blocks, and tar paper (National Research Council, 1999).

Exposure to natural sources of radon has become a significant issue in terms of radiological protection. The UNSCEAR (2000) reports that nearly half of the total natural background dose received from natural sources can be attributed to inhaling radon and its progenies present in dwellings. It is estimated that nearly 1 out of every 15 homes in the United States has excessive elevated radon levels (United States
Environmental Protection Agency (USEPA, 2012).

The presence of radon in indoor air has raised concern that it may also be a cause of lung cancer (Frumkin and Samet, 2001). Long-term exposure to elevated radon concentrations has been linked to increased lung cancer risk. Inhalation of radon is considered to be the second leading cause of lung cancer in the United States (National Research Council, 1999), and probably worldwide. When radon concentrations in a home exceed 4 pCi/L, the USEPA recommends that the house be mitigated (Steck, 2008). The American Association of Radon Scientists and Technologists estimates 10 million homes in America have indoor radon in excess of 4 pCi/L (Siaway, Mose, and Metcalf, 2010). The health effects of this natural gas make it sensitive, yet necessary, to research.

It is known that prevalence and emission of soil radon is influenced by many factors, including geologic structure, atmospheric pressure, temperature, soil moisture, carbon dioxide concentration in the soil, and soil porosity (Fujiyoshi, Morimoto, and Sawamura, 2002). Building a predictive model with factors influencing radon levels has the potential to aid in decision making by stakeholders in real estate development and other related industries.

**Methods**

**Study Area**

The study area was comprised of 72 zip codes in four selected counties in northwestern Illinois: Whiteside, Carroll, Lee, and Olge (Figure 1 and 2). This area was chosen because of the availability of radon measurements summarized by zip code, which was at a relatively higher spatial resolution than other locations summarized at the county level. The total surface area covered by these four counties is approximately 2612.53 square miles according to the United States Census Bureau (2010). The health implications of the presence of radon gas in this area is of great concern since, on average, approximately 63 people live within each square mile (Table 1).

<table>
<thead>
<tr>
<th>County</th>
<th>Area (mi²)</th>
<th>Population</th>
<th>Persons per mi²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carroll</td>
<td>444.81</td>
<td>15387</td>
<td>34.6</td>
</tr>
<tr>
<td>Lee</td>
<td>724.9</td>
<td>36031</td>
<td>49.7</td>
</tr>
<tr>
<td>Olge</td>
<td>758.25</td>
<td>53497</td>
<td>70.5</td>
</tr>
<tr>
<td>Whitefield</td>
<td>684.25</td>
<td>58498</td>
<td>85.5</td>
</tr>
<tr>
<td>Total</td>
<td>2612.53</td>
<td>163413</td>
<td>62.5</td>
</tr>
</tbody>
</table>

Table 1. Study area counties, their population, and persons per square mile (United States Census Bureau, 2010).

![Figure 1. Map of Illinois showing the study area in the northwest portion of the state.](image)

**Radon**

Radon content levels are measured in picocuries per liter (pCi/L) or Becquerel
per kilogram (Bq/kg). The most common concentration measure used is picocuries per liter, which was also used for this study (Frumkin and Samet, 2001).

Physical characteristics, namely geology, soil permeability, elevation, and slope, may be associated with radon concentration and were assessed in this study to ascertain the extent of their relationship. As cited in İçhedef, Saç, Camgöz, Bolca and Harmanşah (2013), Vaupotic; 2007 made clear that in spite of many detailed investigations into meteorological factors and geological structure, only a few studies focused on soil characteristics. Thus, there is a need to investigate these less studied factors of radon concentration. Exploring soil types and other known factors affecting radon levels will help to identify the risks involved in other areas of the county.

Figure 2. Map showing zip codes within the four study area counties.

Radon measurements were gathered between 2003 and 2011 from approximately 1413 sites in the study area by the Illinois Emergency Management Agency. These measurements were summarized at the zip code level and provided by Illinois Emergency Management Agency (2011).

The average radon value of the site measurements within a zip code was attributed to the zip code’s spatial boundary for mapping. In the event that a zip code crossed the county boundary, separate radon averages were reported for only the portion of the zip code within the specified county.

Figure 3. Map of radon concentrations in the study area by zip code.

Radon levels within the study area, according to the Illinois Emergency Management Agency, are summarized in Figure 3 and 4. Of the 72 zip codes within the study area, radon data was available for 49.

Figure 4. Percentage of zip codes in the study area within in radon measurement categories (Illinois Emergency Management Agency, 2011).

Factors Affecting Radon
Geology

As reported by Oliver and Kharyat;1999 cited in Siaway et al., 2010, some radon potential maps show that very high indoor radon concentrations may be correlated with areas where uranium is found in soil, over uranium enriched crystalline rock units, or over locally fractured rocks. However, in other instances high radon homes are simply over soils that have higher permeability.

“To understand the geology of radon - where it forms, how it forms, how it moves - we have to start with its ultimate source, uranium” (Otton, 1992). Uranium deposits can be located and grouped into types based on the geological setting. The most recent geological classification was defined by the International Atomic Energy Agency in 2013 and was adopted in the Red Book in 2014. The uranium deposit types have fundamental characteristics and recognition criteria, which can be observed (World Nuclear Association, 2015).

Uranium deposits can exist in medium to coarse-grained sandstones in a continental fluvial or marginal marine environment. Impermeable shale and mudstone units are interbedded in the sedimentary sequence and often occur above and below the mineralized sandstone (International Atomic Energy Agency [IAEA], 2009). Sandstone deposits consist of about 18% of the world’s uranium and 41% of known deposits (IAEA, 2009).

Eight geological units with the study area were identified: Ancell Group, Cambrian System, Galena Group, Platteville Group, Maquoketa Shale Group, Platteville Group, Prairie du Chien Group, Silurian System, and Tradewater Fm (Figure 5). Four out of the eight units identified were known to have a percentage of sandstone: Ancell Group (60%), Cambrian System (70%), Prairie du Chien (20%), and Tradewater Formation (30%) (United States Geological Survey [USGS], 2014).

Figure 5. Map of study area showing the different types of geological units.

For this study, the geological units with sandstone were categorized as high radon potential geological units based on their uranium deposit probability. The other units were categorized as low radon potential. Classification was accomplished using Esri’s ArcGIS Raster Calculator tool (Figure 6).

Soil Types and Soil Permeability

The soils of some types of rocks are more permeable (more sandy), allowing for more rapid radon movement through soil and facilitating faster and greater entry into homes. Radon can more easily leave the rocks and soils by escaping into fractures and openings in rocks and into the pore spaces between grains of soil. The method and speed of radon’s movement through soil is controlled by the amount of water present in the pore space (the soil
moisture content), the percentage of pore space in the soil (the porosity), and the "interconnectedness" of the pore spaces that determines the soil's ability to transmit water and air (called soil permeability) (USGS, 1995) (Figure 7).

Figure 6. Map of study area showing areas of high and low radon potential geological units.

Figure 7. Map of study area showing soil types based on parent material and permeability.

For these reasons, homes in areas with drier, highly permeable soils and bedrock, such as hill slopes, mouths and bottoms of canyons, coarse glacial deposits, and fractured or cavernous bedrock, may have high levels of indoor radon. Even if the radon content of the air in the soil or fracture is in normal range (200 - 2000 pCi/L), the permeability of these areas permits radon-bearing air to move greater distances before it decays, and thus contributes to high indoor radon (USGS, 1995).

"The most rapid water and air movement is in sands and strongly aggregated soils, whose aggregates act like sand grains and pack to form many large pores" (Babar, 2005). On the other hand, clay has low permeability due to small grain sizes with large surface area. Therefore, soils with sand or silt were considered more permeable and more prone to high radon concentration than clayey soils. The map in Figure 8 was derived by the use of Esri’s Raster Calculator tool.

Figure 8. Map of study area showing areas of high and low radon potential based on soil type.

Slope and Elevation

The location of a home relative to topography may also be important. It is likely that homes constructed on hillsides
(homes on greater slopes) and hilltops (homes at higher elevations) might tend to have more indoor radon because these soils tend to be more permeable, allowing greater movement of radon in soil gas as seen in Mose and Mushrush; 1997 and Siaway, Mushrush, and Mose; 2006 (Siaway et al., 2010).

As seen in Fairfax County GIS (2006), in Northern Virginia, greater slopes tend to be more permeable because they have a higher sand content and therefore might have a higher probability of greater gas flow (Siaway et al., 2010). Areas of high elevation also tend to have more permeable and sandier soils, and gas and liquids move faster through such soils as seen in Mose, Mushrush, Chrosniak, and Morgan; 2006 and Shi, Hoftiezer, Duell, and Onega; 2006 (Siaway et al., 2010). Moreover, it is recognized that significant amounts of radon accumulate in some homes in the Appalachian Mountain System as seen in Mose, Mushrush, and Chrosniak; 1992 (Siaway et al., 2010).

In this study, since there are no published standards as to what exact high elevation contributes to high radon, the elevation ranges generated by the Jenks Natural Breaks using Esri’s ArcMap software were used. Thus, high elevation was considered to be more than 735 m, and classified as high radon potential, whilst elevations less than 735 m were categorized as having low radon potential (Figure 9).

Most of the areas in the east and north portions of the study area have high elevations, while areas west in the study area are low-lying (Figure 10). There is a river flowing from the northeast to southwest creating a low lying area across the study area.

Figure 9. Map of study area showing elevation.

Figure 10. Map of study area showing areas of high and low radon potential based on elevation.

Carson and Biscaye (2006) found most of the sites studied showed relatively low excess radon concentrations with low vertical gradients. For the sake of this study, percent rise less than 25% was considered gently sloped, whilst areas with percent rise greater than 25% were considered steep slopes. This criterion was derived from the Jenks Natural Breaks generated by Esri’s ArcMap software since there is no specific known percent rise in slope that is considered to have a higher radon potential. Figure 11 depicts the slope of the study area.
While steep slopes were considered to be high radon potential slopes, gentle slopes were classified as low radon potential slopes. As shown in Figure 12, most of the steep slopes were found to be in the northwestern part of the study area.

Regression Analysis

As seen in Savović, Djordjevich, Tse, and Nikezić; 2011 reported specialists prefer mathematical models when assessing the radon risk of territories (Nadezhda, 2014). Regression is used to evaluate relationships between two or more feature attributes. Identifying and measuring relationships allows one to better understand what is occurring in a place, predict where something is likely to occur, or begin to examine causes of why things occur where they do (Esri, 2015).

Multiple regression analysis was performed on all the variables together to test their relationship with the radon concentrations in the study area measured by the Illinois Emergency Management Agency. The equation for multiple regression is:

\[ Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4. \]

where \( Y \) is the dependent variable (radon), \( X_i \) are the independent variables, ‘a’ is the intercept and ‘b’ is the slope or beta coefficient.

The individual variables were also tested using simple linear regression to determine their individual influence on the radon levels. According to Rogers and Nicewander (1988), the strength of relationship is expressed by the Pearson's linear correlation coefficient (r). The higher the correlation coefficient value, the stronger the relationship. The sign in front of the r value indicates whether there is a positive or negative relationship. A positive relationship means that as radon levels increased, the variable value also increased. A negative sign indicates that as radon levels decrease, the variable values are larger. The R square value is widely used for assessing the fit of a regression model (Gordon, 2015). All of the statistical tests applied in this study were at the 95% confidence level.

Due to the nominal nature of the geology and soil data, there had to be some sort of transformation to numerical data in order to conduct a regression.
analysis. The Tabulate Area tool was used to find the area of each category within each zip code. The percentage of the area that represents high radon potential based on the stated criteria were deduced and used as values for the variables for the regression analysis. For consistency, the same approach (calculating the percentage of the zip code classified as high radon potential) was used to derive data for elevation and slope. Zip codes that did not have any radon measurement reported by the Illinois Emergency Management Agency were not included in the regression analysis.

Radon Potential Model

Modelling and mapping of geogenics and indoor radon potentials provides an opportunity to identify radon prone areas; therefore, a radon potential map assists to reduce the cumulative radiation risks (Szabó, Jordan, Horváth, and Szabó, 2013). Also, having knowledge of the radon potential of an area can support decisions regarding whether further local measurements are necessary in the area of planned development (Szabó, Jordan, Horváth, Szabó, 2014).

According to the USEPA (2012), because radon comes from the soil, the geology of an area can help to predict the potential for elevated indoor radon levels. Building a GIS model makes it easier to identify potential high risk areas of radon concentration. The resulting map is intended to highlight areas that meet all the physical characteristics potentially influencing increased radon exposure. Real estate developers and other related industry players could make use of this model when planning and assessing risks.

The radon suitability output rasters for the four criteria (geology, soil type, elevation, and slope) described above were used to generate a radon potential map. Each input raster was comprised of values of 1 (high probability) or 0 (low probability). The result was achieved by adding the four rasters using Esri’s Raster Calculator tool.

Results and Discussion

Figure 13 shows the radon potential map generated from the variables. According to the analysis conducted, five suitability categories were generated with zero (0) being the least potential radon concentration and four (4) being the highest potential radon concentration.

Figure 13. Map of study area showing areas of high radon potential and areas of least radon potential.

High radon potential areas can be seen in the northern and eastern part of the study area, which does not fully correlate with the existing radon data. Those places of high radon potential indeed have high radon values. However, there are still many areas within the study area that have high radon values and were not predicted accurately by the model criteria.

Simple linear regression analysis was performed on the variables. For geology, soil, elevation, and slope, the percentage of the zip code considered high
probability for radon, based on the criteria defined earlier, was used for the independent variable values. The average radon value for the zip code was the dependent variable.

Table 2 shows the relationship between geology and the radon measurements. The positive $r$ value indicates a positive relationship between the dependent and independent variables. The significance value also suggests that the model is statistically significant. Thus, the analysis indicated that geology is correlated with radon concentration at a 95% confidence level. The $R$ square value of 0.124 indicated approximately 12.4% of the variance of radon concentration in the study area was predicted by geology. Figure 14 shows the relationship between geology and radon.

Table 2. Values from regression analysis between geology and radon measurements within the study area.

<table>
<thead>
<tr>
<th>$r$</th>
<th>$R$ Square</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.352</td>
<td>0.124</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Figure 14. Scatter plot of the percentage of the zip code classified as high probability due to geology vs. the average radon measurement for the zip code.

Soil permeability had an $R$ square value of 0.069 and a significance value of 0.069, which indicated that this variable was a weak predictor of radon concentration. It accounted for 6.9% of the variance of radon concentration in the study area (Table 3). This buttressed the ascertainment by Khayrat, Oliver, and Durrani (2001) that the correlations between the raw values of radon concentration, soil particle, and size fractions were weak. Figure 15 shows the linear relationship between soil and radon.

Table 3. Values from regression analysis between soil and radon measurements within the study area.

<table>
<thead>
<tr>
<th>$r$</th>
<th>$R$ Square</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.262</td>
<td>0.069</td>
<td>0.069</td>
</tr>
</tbody>
</table>

Figure 15. Scatter plot of the percentage of the zip code classified as high risk due to soil type vs. the average radon measurement for the zip code.

Results for elevation had an $R$ square value of 0.115 and a significance value of 0.017 (Table 4). This variable was statically significant. There is 95% confidence that elevation was correlated with radon concentration and predicted 11.5% of radon concentration. Figure 16 shows the linear relationship between elevation and radon.

Table 4. Values from regression analysis between elevation and radon measurements within the study area.

<table>
<thead>
<tr>
<th>$r$</th>
<th>$R$ Square</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.339</td>
<td>0.115</td>
<td>0.017</td>
</tr>
</tbody>
</table>

Testing the correlation of slope resulted in an $R$ square value is 0.008, which indicates that approximately 0.8% of the variance of radon was accounted for by slope (Table 5). This model was not statistically significant. Hence, the slope
variable was a weak predictor of radon concentration. Figure 17 shows the linear relationship between slope and radon.

**Figure 16.** Scatter plot of the percentage of the zip code classified as high risk due to elevation vs. the average radon measurement for the zip code.

Table 5. Values from regression analysis between slope and radon measurements within study area.

<table>
<thead>
<tr>
<th>r</th>
<th>R Square</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.088</td>
<td>0.008</td>
<td>0.546</td>
</tr>
</tbody>
</table>

**Figure 17.** Scatter plot of the percentage of the zip code classified as high risk due to slope vs. the average radon measurement for the zip code.

**Multiple Regression**

All together these variables were assessed using SPSS to conduct multiple regression analysis. The result reported a positive R value, indicating a positive relationship between the independent variables and the dependent variable. The R square value and significance values were 0.165 and 0.087, respectively. This means the model explained 16.5% of radon concentration. Based on the results of this study, these physical factors were somewhat predictive; however, the significance value greater than 0.05 indicated this model was not statistically significant. The model significance was less than 0.1; however, only being able to predict approximately 17% of radon suggests this model should not be relied upon as a sole predictor of radon (Table 6).

Table 6. Values from regression analysis including geology, soil, elevation, and slope as predictors of radon measurements within the study area.

<table>
<thead>
<tr>
<th>R</th>
<th>R Square</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.406</td>
<td>0.165</td>
<td>0.087</td>
</tr>
</tbody>
</table>

According to the analysis, the beta coefficients derived are reported in Table 7. Thus, the resulting equation was:

$$ \text{Radon} = -0.05 + 0.25(\text{Geology}) + 0.11(\text{Soil}) + 0.16(\text{Elevation}) + 0.08(\text{Slope}) $$

Table 7. Coefficients from the regression analysis using geology, soil, elevation, and slope as predictors of radon measurements within the study area.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta Coefficient (Slope)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geology</td>
<td>0.25</td>
</tr>
<tr>
<td>Soil</td>
<td>0.11</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.16</td>
</tr>
<tr>
<td>Slope</td>
<td>0.08</td>
</tr>
</tbody>
</table>

The slope variable was found to be the least significant of the four (p=0.633), and was thus the weakest contributor to the model. Another multiple regression analysis without slope indicated that the model was statistically significant and still predicted 16% of radon concentration in the study area. This model included geology, soil, and elevation as explanatory variables (Table 8).

According to the analysis, the beta coefficients derived are reported in Table 9. Thus, the resulting equation was:
Radon = 1.44 + 0.23(Geology) + 0.08(Soil) + 0.18(Elevation)

Table 8. Values from regression analysis including geology, elevation, and soil as predictors of radon measurements within the study area.

<table>
<thead>
<tr>
<th>Variable</th>
<th>R</th>
<th>R Square</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.402</td>
<td>0.161</td>
<td>0.046</td>
</tr>
</tbody>
</table>

Table 9. Coefficients from the regression analysis using geology, soil, and elevation as predictors of radon measurements within the study area.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta Coefficient (Slope)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geology</td>
<td>0.23</td>
</tr>
<tr>
<td>Soil</td>
<td>0.08</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.18</td>
</tr>
</tbody>
</table>

The multiple regression model, despite being significant, failed to have strong predictive power and could not be relied upon, probably in part due to the scale at which the study was conducted. Radon data may have been too generalized due to the availability of radon data at the zip code level rather than at a higher resolution. The graphs generated from the analysis also revealed three zip codes with very high radon concentrations compared to the rest; these three values may have had a strong effect on the regression analysis. It would be interesting to see the impact of excluding these outliers from the analysis or modifying the study area to include more zip codes in an effort to better account for the range of average radon levels possible.

Also, there are other essential non-physical factors, such as soil radioactivity, which have been proven to affect radon concentration. These non-physical factors may contribute more in predicting radon concentration.

Conclusion

According to the analysis, the physical factors modeled were weak predictors of the radon concentrations. There are certainly other physical characteristics such as weather, season, temperature, and soil moisture that may improve the model. There are also other non-physical factors, such as soil gas radioactivity. All these factors together may result in a stronger model that would better predict radon concentrations.

References


United States Environmental Protection Agency (USEPA). 2012. A citizen’s guide to radon: the guide to protecting