Modelling and Mapping Regional Indoor Radon Risk in British Columbia, Canada

by

Michael C. Branion-Calles
B.Sc., University of Victoria, 2013

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Geography

SUPERVISORY COMMITTEE

Modelling and Mapping Regional Indoor Radon Risk in British Columbia, Canada

by

Michael C. Branion-Calles
B.Sc., University of Victoria, 2013

Supervisory Committee

Dr. Trisalyn A. Nelson, Supervisor
(Department of Geography, University of Victoria)

Dr. Sarah B. Henderson, Co-Supervisor
(School of Population and Public Health, University of British Columbia; Environmental Health Services, BC Centre for Disease Control)

Dr. Aleck Ostry, Departmental Member
(Department of Geography, University of Victoria)

ABSTRACT

Supervisory Committee

Dr. Trisalyn A. Nelson, Supervisor
(Department of Geography, University of Victoria)

Dr. Sarah B. Henderson, Co-Supervisor
(School of Population and Public Health, University of British Columbia; Environmental Health Services, BC Centre for Disease Control)

Dr. Aleck Ostry, Departmental Member
(Department of Geography, University of Victoria)

Monitoring and mapping the presence and/or intensity of an environmental hazard through space, is an essential part of public health surveillance. Radon, a naturally occurring radioactive carcinogenic gas, is an environmental hazard that is both the greatest source of natural radiation exposure in human populations and the second leading cause of lung cancer worldwide. Concentrations of radon can accumulate in an indoor setting, and, though there is no safe concentration, various guideline values from different countries, organizations and regions provide differing threshold concentrations that are often used to delineate geographic areas at higher risk. Radon maps demarcate geographic areas more prone to higher concentrations but can underestimate or overestimate indoor radon risk depending on the concentration threshold used. The goals of this thesis are to map indoor radon risk in the province of British Columbia, identify areas more prone to higher concentrations and their associations with different radon concentration thresholds and lung cancer mortality trends.

The first analysis was concerned with developing a data-driven method to predict and map ordinal classes of indoor radon vulnerability at aggregated spatial units. Spatially referenced indoor radon concentration data were used to define low, medium and high classes of radon vulnerability, which were then linked to regional environmental and housing data derived from existing geospatial datasets. A balanced random forests algorithm was used to model environmental predictors of indoor radon vulnerability and predict values for un-sampled locations. A model was generated and evaluated using accuracy, precision, and kappa statistics. We investigated the influence of predictor variables through variable importance and partial dependence plots. The model performed 34% better than a random classifier. Increased probabilities of high vulnerability were found to be associated with cold and dry winters, close proximity to major river systems, and fluvioglacial and colluvial soil parent materials. The Kootenays and Columbia-Shuswap regions were most at risk.

We built upon the first analysis by assessing the difference between temporal trends in lung cancer mortality associated with areas of differing predicted radon risk. We assessed multiple scenarios of risk by using eight different radon concentration thresholds, ranging from 50 to 600 Bq m^{-3}, to define low and high radon vulnerability. We then examined how the following parameters changed with the use of a different concentration threshold: the classification accuracy of each radon vulnerability model, the geographic characterizations of high risk, the population within high risk areas and the differences in lung cancer mortality trends between high and low vulnerability stratified by sex and smoking prevalence. We found the classification accuracy of the model improved as the threshold concentrations decreased and the area classified as high vulnerability increased. The majority of the population were found to live in areas of lower vulnerability regardless of the threshold value. Thresholds as low as 50 Bq m^{-3} were associated with higher lung cancer mortality trends, even in areas with relatively low smoking prevalence. Lung cancer mortality trends were increasing through time for women, while decreasing for men. We suggest a reference level as low as 50 Bq m^{-3} is justified for the province.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

There are many people who have had a great impact on my life, each of whom I admire and owe a great deal of thanks for any successes I have had thus far, and may have in the future. I would like to start by thanking my supervisor Dr. Trisalyn Nelson and my co-supervisor Dr. Sarah Henderson for their expertise, guidance, and support, without which I would undoubtedly still be working on my first manuscript. I would like to thank Trisalyn additionally for her infectious positivity, encouragement and dedication, attributes I will aspire to replicate in any academic and/or professional capacity I may have in the future. I would also like to thank Jessica Fitterer for her patience and understanding as my TA several years ago, without which I would not be in the position that I am today. Additionally, thank you to all my lab mates for the advice, words of encouragement, and friendship I have received over the past two years. To my parents, Carlos and Christine, thank you for your unwavering and constant support throughout the entirety of my life. Thank you to my fantastic, large, and boisterous extended family for further contributing to the already infinite supply of support from which I can draw. Finally I would like to thank my wonderfully understanding girlfriend, Justina, for all of her years of dedication, support, and encouragement, the value of which is impossible to convey in words.

CO-AUTHORSHIP STATEMENT

This thesis is the combination of two scientific manuscripts for which I am the lead author. The project structure was developed by Dr. Trisalyn Nelson and Dr. Sarah Henderson, where modelling and mapping of regional radon risk in British Columbia was identified as a key research opportunity. For these two scientific manuscripts I led all research, data preparation, data analysis, initial interpretation of results and the final manuscript preparation. Dr. Trisalyn Nelson and Dr. Sarah Henderson provided guidance in the initial development of research questions, as well as contextualization and interpretation of results. Dr. Trisalyn Nelson and Dr. Sarah Henderson supplied editorial comments and suggestions that were incorporated into the final manuscript.

1.0 INTRODUCTION

1.1 Research Context

The interactions between human populations and environmental hazards have important implications for global population health. It is estimated that nearly a quarter of the global burden of disease can be attributed to human exposure to environmental hazards (Prüss-Üstün et al. 2006). Regional disparities in disease burden to specific environmental hazards arise in part as a result of the differing presence or intensity of a given environmental hazard through space and their proximity to human populations (Prüss-Üstün et al. 2006). In order to mitigate the negative effects of environmental hazards it is of utmost importance to understand the hazards physical properties, generating processes, and biological mechanisms by which it induces negative health effects (Maantay & Mclafferty 2011). Once the health effects of an environmental hazard are understood, a central component of strategies to reduce human exposure is to map its variation in magnitude or presence through space, making spatial perspectives essential (Maantay & Mclafferty 2011). Specific interventions to mitigate the effects of environmental hazards can then be put into place to reduce the burden of disease and increase population-level health of an affected region.

When adverse health outcomes associated with exposure to a specific and measurable environmental hazard has been established, the surveillance of the spatial distribution of that hazard represents the most effective means for intervention in reducing human exposure (Thacker et al. 1996). Hazard surveillance refers to simply measuring the intensity or presence of a specific hazard responsible for negative health outcomes in a population within a given geographic region (Thacker et al. 1996). Often environmental hazards are spatially continuous and therefore surveys of measured observations will only represent a sample of the spatial distribution of the phenomenon of interest. Therefore, applied spatial analysis methods are suited for predicting values in unmeasured areas of a jurisdiction (Zhu et al. 2001; Miles & Appleton 2005; Kemski et al. 2008).

Geographic Information Science (GIS) approaches and techniques are appropriate for studying environmental hazards as GIS technologies can effectively store, manipulate, analyze and visualize spatial data, such as measurements of the intensity of an environmental hazard. Using applied spatial analysis methods and the data acquired from directly or indirectly monitoring a given hazard, researchers can determine where a hazard poses the greatest threat and visualize the results (Maantay & Mclafferty 2011; Kemski et al. 2008; Miles & Appleton 2005; Zhu et al. 2001; Ielsch et al. 2010; Sainz-Fernandez et al. 2014). When these datasets are overlaid with other relevant geospatial datasets that describe the conditions known, or theorized to affect the intensity or presence of a hazard, it can result in the discovery of relationships between spatial-variables associated with the higher intensities of the hazard through space, a model of the hazards spatial distribution and an assessment of its subsequent impact on human populations (Cromley 2003). There exists a growing range of studies on different environmental hazards, from the modeling of airborne toxic chemicals to the mapping of the spatial distribution of biological agents of disease (Cromley 2003). The use of GIS technologies and techniques for the analysis, modeling and visualization of the spatial distributions of various causative disease agents is a critical component to hazard surveillance and a vital precursor to effectively implementing interventions to reduce negative health effects in local populations.

1.2 Research Focus

The focus of this thesis is concerned with the environmental hazard radon, a naturally occurring radioactive carcinogenic gas. Radon is not only the greatest source of natural radiation exposure in human populations, but also the second leading cause of lung cancer worldwide (Charles 2001; World Health Organization 2009). Radon is produced naturally by the earth’s surface through the radioactive decay of uranium and is diluted to low concentrations when exhaled into outdoor air. Uranium and its daughter products are present in varying amounts in all terrestrial substances, meaning some concentration of radon is present in both outdoor and indoor air (Bissett & McLaughlin 2010; Appleton 2007). Radon concentrations can, however, accumulate within enclosed structures such as residential homes to levels several orders of magnitude higher than a typical outdoor concentration. There is no safe concentration of radon , and the risk of lung cancer increases linearly with increasing concentrations (Darby et al. 2005). In order to reduce population level exposure to indoor radon, the hazard must first be monitored. Surveillance of indoor radon involves testing individual homes within jurisdictions, which consists of placing a radon detector in a home for a specified period of time, typically at least three months during the heating season, which will record the average concentration during that period. Indoor radon is a spatially variable environmental hazard that can be readily monitored, and, as a result, can be studied using GIS approaches to analyze, model and map regions at greater risk to higher concentrations.

Radon maps that identify areas more prone to higher indoor radon concentrations are an important component of any radon reduction strategy that can help to guide radon policy, future radon surveys and communicate risk (Chen 2009; Long & Fenton 2011; Miles & Appleton 2005). The methods used to create radon risk maps vary based on the availability of existing relevant data sources, but can be delineated into two broad areas based on which data sources they use to infer radon risk: indoor radon data or geologic proxy data (Chen 2009; Appleton & Ball 2002). Maps produced by the former generally will either visualize the variability in radon risk through the mean observed concentration across mapping units or estimates of the proportion of homes expected to exceed a threshold concentration (Dubois 2005; Miles & Appleton 2005; Sainz-Fernandez et al. 2014). The latter method infers indoor radon risk through the use of proxy data such as uranium and/or radium concentrations in rocks and soils, radon concentrations in soil gas, or soil permeability, among others, which all serve to estimate a regions capacity for delivering radon to the surface (Kemski et al. 2001; Kemski et al. 2008; Appleton & Ball 2002; Ielsch et al. 2010). In order to produce spatially continuous maps using observed measurements at a fine level of geographic detail, a large number of measurements that are uniformly distributed throughout the jurisdiction are required (Miles & Appleton 2005). If the region is sparsely sampled and/or populated, the resulting map will either contain many blank areas or make use of much larger mapping units (Chen 2009; Sainz-Fernandez et al. 2014). Though the use of geologic proxy data can provide a means for predicting radon risk in sparsely measured or populated areas, they can be unreliable for inferring indoor radon risk due to the importance of housing characteristics on individual concentrations (Appleton & Ball 2002; Rauch & Henderson 2013).

Additional uncertainty is introduced for maps of indoor radon risk that make direct use of indoor radon data, due to the fact that generally, a specific concentration threshold is used either directly or indirectly to delineate different classes of radon risk for mapping units (Miles & Appleton 2005; Dubois 2005; Friedmann 2005; Sainz-Fernandez et al. 2014). There are a variety of differing radon concentration guidelines provided throughout the world that are generally intended for homeowners to decide if they need to implement remediation measures to reduce the concentration in their home (World Health Organization 2007), but are also often used as a threshold concentration for delineating classes of risk in radon mapping. A recommended concentration threshold within a given jurisdiction can be used to define regional risk, and, due to the arbitrary nature of its recommendation, can potentially over or underestimate risk depending on the concentration selected.

British Columbia(BC) has many radon-prone communities and indoor radon has been identified as an important contributor to lung cancer incidence and mortality (Henderson et al. 2014; Henderson et al. 2012). A rich dataset of spatially referenced observed indoor radon concentrations from several sampling campaigns that took place in the province between 1991 and 2014 are archived at the BC Centre for Disease Control. Due to the fact large regions of the province are sparsely populated, the indoor radon dataset is not uniformly distributed throughout the province, resulting in current radon risk maps making use of large mapping units (Henderson et al. 2012) or having blank spaces in unmeasured areas (BC Centre for Disease Control 2009). The Radon Potential Map of Canada (Radon Environmental Management Corp. 2011) is available and can provide a spatially continuous estimate of radon risk for the province, but its predictions are inconsistent with radon observations in BC (Rauch & Henderson 2013). The availability of indoor radon data, combined with the lack of spatially continuous maps of *indoor* radon risk at fine spatial resolutions, provide opportunity to develop methods for mapping indoor radon risk in the province using GIS approaches and techniques.

1.3 Research Goals and Objectives

The goals of this thesis are to map indoor radon risk in the province of British Columbia, identify areas more prone to higher concentrations of indoor radon and their associations with different concentration thresholds and lung cancer mortality trends. Using applied spatial modeling techniques and methods we base our approach on combining observed indoor radon concentrations with various related environmental geospatial datasets to predict ordinal classes of regional vulnerability to indoor radon, and assess the sensitivity of geographic characterizations of risk to different parameters, specifically, the use of different concentration thresholds to delineate areas of high and low radon risk. In order to accomplish these goals the following objectives will be met:

1) The first objective consists of developing a data-driven method to predict classes of indoor radon risk and assess the relationships between predictors and classes of radon risk that we term radon *vulnerability*. The results can then be mapped and used to identify regions most at risk in the province.
2) The second objective is to assess the difference in temporal trends in lung cancer mortality associated with areas of differing predicted radon vulnerability. We test different geographic characterizations of radon vulnerability associated with changes in radon concentration thresholds and observe the subsequent changes in

populations within high vulnerability areas. We then compare lung cancer mortality trends across them.

References

Appleton, J.D., 2007. Radon: sources, health risks, and hazard mapping. *Ambio*, 36(1), pp.85–89. Available at: http://www.jstor.org/stable/4315791.

Appleton, J.D. & Ball, T.., 2002. Geological radon potential mapping. In P. T. Bobrowsky, ed. *Geoenvironmental Mapping: Methods, Theory and Practice*. Exton, PA: A.A. Balkema Publishers, pp. 577–613.

BC Centre for Disease Control, 2009. Radon Terrestrial Maps of BC. Available at: http://www.bccdc.ca/resourcematerials/guidelinesandforms/guidelinesandmanuals/EH_Sum_Radon_Maps_BC.htm [Accessed September 15, 2013].

Bissett, R.J. & McLaughlin, J.R., 2010. Radon. *Chronic Diseases in Canada*, 29. Available at: http://search.proquest.com.ezproxy.library.uvic.ca/docview/1115551026?accountid=14846.

Charles, M., 2001. UNSCEAR Report 2000: Sources and Effects of Ionizing Radiation. *Journal of Radiological Protection*, 21(1), p.83.

Chen, J., 2009. A preliminary design of a radon potential map for Canada: a multi-tier approach. *Environmental Earth Sciences*, 59(4), pp.775–782. Available at: http://link.springer.com/10.1007/s12665-009-0073-x [Accessed September 25, 2013].

Cromley, E.K., 2003. GIS and Disease. *Annual review of public health*, 24, pp.7–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12668753 [Accessed November 14, 2013].

Darby, S. et al., 2005. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ*, 330(7485), pp.223–226. Available at: http://www.bmj.com/cgi/doi/10.1136/bmj.38308.477650.63 [Accessed September 25, 2013].

Dubois, G., 2005. *An Overview of Radon Surveys in Europe*,

Friedmann, H., 2005. Final results of the Austrian Radon Project. *Health physics*, 89(4), pp.339–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16155455.

Henderson, S.B. et al., 2014. Differences in lung cancer mortality trends from 1986-2012 by radon risk areas in British Columbia, Canada. *Health Physics*, 106(5), pp.608–613. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24670910 [Accessed October 16, 2014].

Henderson, S.B., Kosatsky, T. & Barn, P., 2012. How to Ensure That National Radon Survey Results Are Useful for Public Health Practice. *Can J Public Health*, 103(3), pp.231–234.

Ielsch, G. et al., 2010. Mapping of the geogenic radon potential in France to improve radon risk management: methodology and first application to region Bourgogne. *Journal of environmental radioactivity*, 101(10), pp.813–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20471142 [Accessed September 25, 2013].

Kemski, J. et al., 2008. From radon hazard to risk prediction-based on geological maps, soil gas and indoor measurements in Germany. *Environmental Geology*, 56(7), pp.1269–1279. Available at: http://link.springer.com/10.1007/s00254-008-1226-z [Accessed November 5, 2013].

Kemski, J. et al., 2001. Mapping the geogenic radon potential in Germany. *The Science of the total environment*, 272(1-3), pp.217–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11379913.

Long, S. & Fenton, D., 2011. An overview of Ireland's National Radon Policy. *Radiation protection dosimetry*, 145(2-3), pp.96–100.

Maantay, J.A. & Mclafferty, S., 2011. Environmental Health and Geospatial Analysis: An Overview. In J. A. Maantay & S. McLafferty, eds. *Geospatial Analysis of Environmental Health*. Dordrecht: Springer Netherlands, pp. 3–37. Available at: http://link.springer.com/10.1007/978-94-007-0329-2 [Accessed November 27, 2013].

Miles, J.C.H. & Appleton, J.D., 2005. Mapping variation in radon potential both between and within geological units. *Journal of Radiological Protection*, 25(3), pp.257–276. Available at: http://iopscience.iop.org/0952-4746/25/3/003/ [Accessed September 24, 2013].

Prüss-Üstün, A., Corvalán, C. & World Health Organization, 2006. *Preventing disease through healthy environments: towards an estimate of the environmental burden of disease*, Geneva: World Health Organization. Available at: http://uvic.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwfV27CsIwFL34WAQHn6hVyA-oNbdNm1ksDi6Cu6Qxcevu33tDotYiLoFkyANCTs4J5wQA-SZeN86E240bIXfaCp3blAqdapkYtInZOdT-dpPBKxWimZwYjCd_3meI7gjENrSJfDlDx_n0VlwIrUQmU6Jm0sGWRMJGH8HzrmMt-zNATDGAjrMdDKFlqhH0vZrGvE.

Radon Environmental Management Corp., 2011. Radon Potential Map of Canada. Available at: http://www.radoncorp.com/pdf/presentationMappingPublic.pdf [Accessed November 14, 2013].

Rauch, S.A. & Henderson, S.B., 2013. A comparison of two methods for ecologic classification of radon exposure in British Columbia: residential observations and the radon potential map of Canada. *Canadian journal of public health*, 104(3), pp.e240–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23823889.

Sainz-Fernandez, C. et al., 2014. The Spanish Indoor Radon Mapping Strategy. *Radiation Protection Dosimetry*, 162(1-2), pp.58–62.

Thacker, S.B. et al., 1996. Surveillance in environmental public health: Issues, systems, and sources. *American Journal of Public Health*, 86(5), pp.633–638.

World Health Organization, 2007. *International Radon Project Survey on Radon Guidelines, Programmes and Acvitivites*, Geneva. Available at: http://www.who.int/ionizing_radiation/env/radon/IRP_Survey_on_Radon.pdf.

World Health Organization, 2009. *WHO Handbook on Indoor Radon: A Public Health Perspective* H. Zeeb & F. Shannoun, eds., Geneva.

Zhu, H.C., Charlet, J.M. & Poffijn, a., 2001. Radon risk mapping in southern Belgium: an application of geostatistical and GIS techniques. *Science of the Total Environment*, 272(1-3), pp.203–210.

2.0 A GEOSPATIAL APPROACH TO THE PREDICTION OF INDOOR RADON VULNERABILITY IN BRITISH COLUMBIA, CANADA

2.1 Abstract

Radon is a carcinogenic radioactive gas produced by the decay of uranium. Accumulation of radon in residential structures contributes to lung cancer mortality. The goal of this research is to predict residential radon vulnerability classes for the province of British Columbia (BC) at aggregated spatial units. Spatially referenced indoor radon concentration data were partitioned into low, medium, and high classes of radon vulnerability. Radon vulnerability classes were then linked to environmental and housing data derived from existing geospatial datasets. A balanced random forests algorithm was used to model environmental predictors of indoor radon vulnerability and values at un-sampled locations across BC. A model was generated and evaluated using accuracy, precision, and kappa statistics. The influence of predictor variables was investigated through variable importance and partial dependence plots. The model performed 34% better than a random classifier. Increased probabilities of high vulnerability were associated with cold and dry winters, close proximity to major river systems, and fluvioglacial and colluvial soil parent materials. The Kootenays and Columbia-Shuswap regions were most at risk. Here we present a novel method for predictive radon mapping that is broadly applicable to regions throughout the world.

2.2 Introduction

Indoor radon is the second-leading cause of global lung cancer, and puts those who smoke at elevated risk (World Health Organization 2009; Saccomanno et al. 1988).

In Canada, radon is estimated to be a factor in more than 3,000 lung cancer deaths annually (Chen et al. 2012). Radon-222 is an odourless, and colourless radioactive noble gas that results from the decay sequence of uranium-238. Uranium-238 occurs naturally in bedrock and soil so its daughter products are present in varying amounts in all terrestrial substances (Bissett & McLaughlin 2010). Because radon is a gas with a half-life of 3.8 days, it can migrate from its source through permeable soils or cracks in rocks and into the atmosphere where it can interact with humans. Radon exposure accounts for an estimated 50% of the worldwide average human radiation dose from natural sources (Charles 2001). Although radon quickly disperses in outdoor air, it can enter buildings through cracks in their foundations and concentrations can accumulate (Bissett & McLaughlin 2010).

Indoor radon concentrations depend on complex interactions between environmental factors and housing characteristics, making them highly variable both locally and regionally. Variation in surficial radon is influenced by the quantity and distribution of uranium in the grains, as well as the characteristics of the substrates through which radon atoms move (Michel 1987). Radon is ejected into the pore space of rock and soils from a radium atom embedded in the grains, and is transported to the surface through diffusive or advective transport (Nazaroff 1992; Arnold 2006). Diffusive transport is the dominant process, which is affected by moisture content, porosity, and tortuosity of the substrate (Nazaroff 1992; Arnold 2006). Advective transport is controlled by permeability, moisture content, and the pressure gradient dictating the flow of soil gas from high to low concentrations. Two factors that affect permeability of certain soils are grain size and moisture content (Nazaroff 1992). The larger the grain, the larger the pore spaces, the more space through which soil gas can flow. Higher moisture contents generally reduce air permeability of a soil, as more moisture in the pore spaces reduces the amount of space through which soil gas can flow (Nazaroff 1992). Factors affecting soil moisture and pressure gradients will also affect the diffusive and advective movement of radon in the subsurface (Washington & Rose 1990; Schumann et al. 1988). Additionally, radon transport can be increased by movement through crevices in the earth such as faults, or anthropogenic openings such as mining tunnels (Appleton 2007).

While geologic properties influence surficial radon levels, indoor radon levels can be primarily attributed to the permeability of a building, especially the parts of the foundation that are in contact with the ground. Most indoor radon can be attributed to the flow of soil gas into a building through permeable entry points (Appleton 2007). This occurs because of the "stack effect" (Vasilyev & Zhukovsky 2013; Al-Ahmady & Hintenlang 1994; Kitto 2005) whereby temperature differences create an area of low pressure within the building compared with outside, causing soil gas to be drawn indoors (Wang & Ward 2002; Garbesi et al. 1993). However, radon concentrations in soil gas are weakly correlated to corresponding indoor radon concentrations (Varley & Flowers 1998). The complexities introduced by differing foundation types, construction methods, and ventilation characteristics of homes can result in variable rates of radon entry and accumulation, even within homes that have equal concentrations of radon in the underlying soil gas (Appleton 2007). Similarly, homes with the same construction may have different concentration measurements due to differing underlying geologic conditions, causing different rates of geogenic production and transport of radon into the home (Appleton & Miles 2010; Appleton 2007). Increased rates of geogenic production will not necessarily translate into high indoor radon concentrations, just as low geogenic production will not necessarily translate into low indoor radon concentrations.

The province of British Columbia (BC) in Canada has areas with an abundance of uranium (Jones 1990), and many small and large radon-prone communities. Indoor radon concentrations in BC have been measured in five disparate sampling campaigns from 1991-2013, and the data are archived at the BC Centre for Disease Control (BCCDC). The provenance of these datasets is inconsistent, but few other resources are available to gauge the regional variations in indoor radon in BC. Some provinces such as Quebec and Nova Scotia have independently developed radon potential maps in order to provide a spatial indication of regions with more or less capacity to exhale radon at the surface (Drolet et al. 2013; Drolet et al. 2014; O'Reilly et al. 2013). In British Columbia, an ambient radon potential map is available only as a part of the broader Radon Potential Map of Canada (Radon Environmental Management Corp. 2011). Radon potential maps are based on an assessment of geologic conditions that contribute to the relative difference between the natural capacities for geologic formations to deliver radon to the atmosphere. As such, they do not necessarily reflect *indoor* radon concentrations (Appleton & Ball 2002; Ielsch et al. 2010; Gruber et al. 2013). This uncertainty is reflected by the fact that the Radon Potential Map of Canada is known to be inconsistent with residential radon observations (Rauch & Henderson 2013) in many areas of BC. Therefore, an *indoor* radon vulnerability map of BC would be complementary. The significant health risks associated with radon provide great motivation to identify and map areas of BC most at risk. The creation of an indoor radon vulnerability map could inform radon mitigation policy as well as be a means to generate increased radon awareness.

The goal of this research is to create an indoor radon vulnerability map for the province of BC by addressing the following objectives: 1) pre-process spatially referenced indoor radon concentration data and relevant overlapping environmental geospatial datasets, and conflate each into a common zonal system to create an indoor radon vulnerability database; 2) using the database, develop a model for the prediction of indoor radon vulnerability for unmeasured areas of the province and assess the relationships between the predictors and radon vulnerability; 3) classify the unmeasured areas of the province, identify regions and population centres most at risk and those most in need of further sampling, and map the results.

2.2.1 Study Area

The study area is the province of BC, on the west coast of Canada (Figure 2.1). BC is a large, mountainous province, whose spatial extent covers over 940,000 km^2 and encompasses a wide variety of landscapes, geologic conditions, and surficial materials. The province has a complex tectonic and glacial history, so its uranium content, geology, climate, and soil characteristics are highly variable on local and regional scales.

2.3 Materials and Methods

2.3.1 Indoor Radon Concentration Observations

The five available datasets for residential radon concentrations were provided in tabular form by the BCCDC. They included surveys conducted by the BCCDC, the

Northern Health Authority, the BC Lung Association, The Donna Schmidt Foundation, and one private contractor. The BCCDC tested 1,552 homes between 1991-1992 and 2004-2006. The first survey was designed to oversample areas with high ambient radiation levels, and the second survey oversampled areas with moderate ambient radiation levels. The Northern Health Authority, the BC Lung Association, the Donna Schmidt Foundation, and a private contractor all have collected volunteer samples between 1997 to the present time. The Northern Health Authority collected samples from 541 homes in Northern BC, the Donna Schmidt Foundation tested 1,136 homes within the Kootenay Region, and the BC Lung Association collected samples from 1,277 homes throughout the province. A further 292 samples were collected by the private contractor primarily within the Thompson-Okanagan Region including cities such as Kelowna and Kamloops. A combined total of 4,798 homes were tested in British Columbia from 1997-2013.

Each survey had the common intent of recording indoor radon concentrations, but was executed with different objectives and over different time periods, resulting in each having varying geographic extents, sampling designs, spatial resolutions, and relevant attributes recorded. Only three common attributes are available between the surveys: a six digit postal code, the date of the test period, and a radon concentration value. Each observation was assigned a geographic coordinate (latitude and longitude) based on its associated postal code using the BCCDC geocoder. Approximately 90.7% of homes tested were successfully geocoded, which resulted in a dataset of 4,352 indoor radon observations distributed throughout the province (Figure 2.1).

2.3.2 Predictor Variables

Geospatial datasets representing environmental and housing predictors were compiled (Table 2.1). Based on the available data the following variables were assessed at each radon measurement location: (1) simplified bedrock lithological class; (2) geologic fault presence; (3) dominant soil parent material; (4) dominant soil drainage class; (5) dominant rooting depth class; (6) dominant soil coarse fragment content; (7) dominant kind of surface material; (8) average winter temperature; (9) average winter precipitation; (10) distance to nearest major river; (11) dominant age of home; and (12) proportion of homes in need of major repairs. Each of these variables was selected based on its potential to affect an indoor radon concentration.

2.3.3 Data Pre-processing

To enable modelling and prediction we integrated all data into similar spatial units that we defined by intersecting geologic units and census areas (Miles & Appleton 2005). We labelled each unit as a "Bedrock Dissemination Area" (BDA) and assumed that each had relatively homogenous environmental and social conditions.

For BDAs with observed radon concentrations, the distribution of all measurements was summarized with a single value for the purposes of modelling. Because the distribution of our indoor radon dataset approximates log-normality the mean concentration would generally underestimate indoor radon vulnerability. Instead, we summarized the distribution in each BDA using the 95th percentile.

The Health Canada guidelines for radon exposure were used to classify the 95th percentile values (Health Canada 2009) as low, moderate, or high. Health Canada suggests that homes with concentrations < 200 Bq m^{-3} do not require remediation, that homes >= 200 Bq m^{-3} and < 600 Bq m^{-3} should be remediated within the next few years, and that homes >= 600 Bq m^{-3} should be remediated within the next year.

The last step was to associate each spatial unit of prediction with relevant predictor variables derived from overlapping geospatial datasets in order to create both a training dataset and a prediction dataset (Table 2.2). The assignment of predictor variable values to each BDA geometry was based on spatial location.

2.3.4 Modelling and Predicting Indoor Radon Vulnerability Using Balanced Random Forest

To map radon vulnerability for the province we created a model using the statistical classifier random forests (Breiman 2001). The complexity of the radon data required a modelling technique that was able to describe multifaceted environmental phenomenon. Random forests were selected as they are a robust, non-parametric ensemble classifier with a high predictive ability that can accommodate mixed variable types, non-linear relationships, and high order interaction effects between predictor variables (Cutler et al. 2007; Prasad et al. 2006). Classification trees work by recursively partitioning a dataset into increasingly smaller subsets based on a value of a particular predictor variable (Breiman et al. 1984). Each binary split maximizes the homogeneity of the response variable within the resulting subsets, thereby maximizing the heterogeneity between subsets.

The random forest algorithm works by combining hundreds to thousands of maximally grown classification trees, each of which is constructed from bootstrapped samples (Breiman 2001). Balanced random forests are a variant that improves the ability to classify a minority class in an imbalanced dataset (Chen et al. 2004). In a traditional random forest the bootstrapped sample taken from an imbalanced dataset will likely be comprised almost entirely of observations that belong to a majority class, resulting in the construction of classification trees which will be incapable of effectively predicting for the minority class (Chen et al. 2004). The balanced approach modifies the sampling method for the training data. The balanced random forest model will classify the minority class more effectively than the traditional random forest, though the overall accuracy will decrease (Chen et al. 2004).

The predictive accuracy of a model can be obtained in a random forest using "out-of-bag" (OOB) data. This refers to the observations that were not used to construct an individual classification tree (Breiman 2001). Unbiased estimates of the predictive accuracy can then be derived from the summation of the predicted classifications of OOB data over all trees in the forest. Specifically, for every tree, the OOB data are dropped down and their predicted classes are recorded. The final predictions of an observation class are made by selecting the class that was most probable when it was OOB.

2.3.5 Evaluating Model Accuracy

The model was evaluated through hold-out validation (HOV) and metrics derived from OOB predictions, including class accuracy, precision, and kappa scores. HOV was computed by training the model on a stratified random sample consisting of

90% of the training data and testing on the remaining 10%. Results of the HOV may have high variance, as they are subset dependent, and therefore we used the average results from 100 runs.

Because our aim was to use the model for prediction, we also trained the model using the entire data set. When the complete data were used the model was validated using OOB comparison. Metrics derived from the OOB confusion matrix also have the advantage of giving accurate and unbiased estimate of the predictive ability of the model (Liaw & Wiener 2002).

The performances of each model were investigated though an evaluation of the accuracy and precision with which each individual class were predicted. Class accuracy describes classification accuracy associated with each individual class and indicates the proportion of the true population of a given class that will be correctly predicted for future instances. The class precision complements class accuracy by estimating the proportion of those observations predicted to be a given class that are correct.

The kappa statistic was used as a measure of overall performance of a model as it is a more robust evaluation of a models overall performance than the overall accuracy in an imbalanced dataset (Fatourechi et al. 2008). The kappa statistic quantifies the degree to which a models overall predictive accuracy (the rate at which it correctly classifies OOB data) are due to more than random chance alone (Cohen 1960).

2.3.6 Evaluating Predictors

The strongest predictor variables were selected based on the variable importance plots derived from the model, and partial dependence plots were created for the four strongest predictors. Variable importance plots reveal the relative importance of variables in the classification (Archer & Kimes 2008; Liaw & Wiener 2002). Partial dependence plots can then provide insight into the directionality of the effect for a given predictor (Berk 2008; Cutler et al. 2007).

Two measures of variable importance can be derived from a random forest algorithm: the mean decrease in the Gini Index (Gini Importance) and the mean decrease in predictive accuracy (Predictive Importance). Though each measure can be unreliable in models that use mixed variable types with different scales of measurement, we chose to use the Predictive Importance because it is less biased than the Gini Importance (Strobl et al. 2007).

The Predictive Importance of a variable reflects the average decrease in OOB estimates of predictive accuracy when the values of a given variable are randomly permuted (Archer & Kimes 2008). The variables causing the greatest decrease are considered the most important. If the decrease in predictive accuracy is zero for a variable, we can infer that it contributes no explanatory power to the model.

Partial dependence plots are a visual representation of the directionality of a relationship between a single class probability and a response variable while holding the values of the remaining predictor variables constant (Cutler et al. 2007; De’ath 2007;

Berk 2008). The units of the vertical axis are the difference between the logarithm of the class probability and the logarithm of the average class probability. Probabilities are derived from the predicted number of observations belonging to a class when the predictor variable is fixed on a single value, divided by the total number of observations (Berk 2008). The units of the horizontal axis are the units of the predictor. The resulting plot can be interpreted as the change in class probability in relation to the range of possible values for the predictor.

2.4 Results

2.4.1 Indoor Radon Vulnerability Database

The Indoor Radon Vulnerability database created in data pre-processing consisted of 36,061 total BDAs, 1054 of which were assigned an indoor radon vulnerability classification based on the 95th percentile. The 1054 BDAs containing radon concentrations made up the entirety of the training dataset, where each BDA was associated with 12 predictor variables and 3 dependent variables. The dataset for prediction consisted of the remaining BDAs with the same 12 predictor variables and no values for the dependent variables. Approximately 23% of BDAs within the province had a value for at least one predictor variable that was not present in the training data, thereby excluding them from the prediction dataset. A total of 26,719 out of the 34,972 BDAs without a response variable made up the prediction dataset.

The class distribution of indoor radon vulnerability in the training data was highly imbalanced (Figure 2.2). Low vulnerabilities made up 75.5% of the sampled BDAs. This is consistent with the fact that radon concentrations are log-normally distributed and, therefore, most areas are characterized by low concentrations, even within areas more prone to high concentrations.

2.4.2 Evaluating Model Performance

The models accuracy and precision varied between low, moderate, and high vulnerability classes based on both OOB and HOV estimates of error (Table 2.3). According to OOB estimates the model predicted low vulnerabilities 75% accurately, moderate vulnerabilities 44% accurately and high vulnerabilities 54% accurately. Precision estimates according to OOB were 92%, 29%, and 30% for low, moderate, and high vulnerabilities, respectively. A kappa score of 0.34 indicates that the model performed 34% better than a random classifier. The HOV estimates corroborated the OOB estimates within a few percentage points for all measures with the exception of the accuracy with which it predicted high vulnerabilities. The HOV estimated the class accuracy of high vulnerabilities to be 48% compared with the OOB estimation of 54%. Overall, 32% of BDAs were misclassified, the majority of which were the result of overestimation (Table 2.4). Of the 32% of misclassified BDAs, 76% could be attributed to overestimations of risk.

2.4.3 Evaluating Predictors

The four most important predictors in decreasing order were: (1) average winter temperature; (2) dominant soil parent material; (3) average winter precipitation; and (4) distance to nearest major river (Figure 2.3).

In general, BDAs with colder winter temperatures were more susceptible to moderate or high vulnerability classifications than areas with warmer winter temperatures (Figures 2.4a, b and c). The odds of a low vulnerability increased rapidly for BDAs with average winter temperatures above -2°C (Figure 2.4a). Similar observations were made by Kropat et al. (2014) where warmer ambient temperatures were associated with lower indoor radon concentrations in Switzerland (Kropat et al. 2014).

Increased rainfall was not clearly associated with radon vulnerability for any of the classes (Figure 2.4d, e and f). The odds of the highest vulnerability classification were generally lower with increasing precipitation (Figure 2.4f).

Closer proximity to major rivers was associated with increased odds of a high radon vulnerability, and decreased odds of low and moderate vulnerability (Figure 2.4g, h and i). There was a steep rise in the odds of a low vulnerability with increasing distance from 0 m to roughly 13,000 m (Figure 2.4g). At distances up to 6500 m the odds of a high vulnerability were increased (Figure 2.4i). For distances greater than 6500 m but less than 13,000 m there was greatest odds of moderate classification (Figure 2.4h). For distances greater than 13,000 there was no change in the partial dependence of any radon vulnerability class. Finally, the partial dependence of radon vulnerability on dominant soil parent material showed that fluvioglacial and colluvial material were associated with the highest probability of moderate and high vulnerability classification and a decreased probability of a low classification (Figure 2.5).

2.4.4 Mapping and Assessing Regional and Local Radon Vulnerability

The radon vulnerability map showed that the interior region of the province had a greater prevalence of moderate and high radon vulnerability than the west coast, which was comprised mostly of low vulnerabilities (Figure 2.6). The specific regions identified to be at most risk were primarily in the south-east portion of the province and include the Central Kootenay, and Kootenay Boundary census divisions (Table 2.5). Regions least at risk were those on the west coast, including the Greater Vancouver area (Table 2.5). The population centres identified to be most vulnerable were generally within the Central Kootenay and Kootenay boundary census divisions and included Grand Forks, Salmo, Rossland, and Castlegar (Table 2.6). The population centres that are both high risk and under-sampled included Lillooet, Mackenzie, Sicamous, and Tumbler Ridge (Table 2.7).

2.5 Discussion

Interpretation of the final predictive map should take into account that both moderate and high indoor radon vulnerabilities represent areas where the 95th percentile radon concentration is estimated to be greater than the threshold set by Health Canada for delineating long term risk because the vulnerability classes are based on the 200 and 600 Bq m^{-3} guidelines. There is always the potential for high individual radon concentrations within areas deemed to have a low vulnerability. Despite the fragmented appearance of the map as a result of 23% of the province being excluded from prediction, there are predictions for 99% of BDAs within population centre boundaries.

The choice of the 95th percentile radon concentration to classify indoor radon vulnerability resulted from testing multiple models, comparing their performance, and selecting the model that performed most adequately based on class accuracy, class precision and a kappa score. We tested and compared models that used classifications based on the 50th, 75th and 95th percentile concentrations. Fundamental to the evaluation was the notion that the importance of accurate classification was not equal between the classes in the context of cancer prevention. Each class represented an increasing vulnerability to high indoor radon concentrations, and therefore potentially an increasing vulnerability to higher radon induced lung cancer rates. As a result, accurately classifying high indoor radon vulnerability carried more weight than accurately classifying moderate indoor radon vulnerability. Similarly, accurately classifying moderate vulnerability was more important than accurately classifying low vulnerability. The 95th percentile model was found to have the best high vulnerability class predictions, as measured by the class accuracy and precision, as well as the highest kappa score.

The relatively low precision with which the model predicts moderate and high vulnerabilities resulted in a predictive map that overestimates their overall prevalence (Table 2.4). However, given that one of the aims of the study was to reduce radon induced lung cancer through identification of radon prone regions, overestimations of radon vulnerability were considered preferable to underestimations.

The main strength of the final model is that it depicts areas of lower and higher radon risk with accuracy. If we consider the results with no distinction between the moderate and high categories, the accuracy of areas delineated as lower radon risk (low) or higher radon risk (moderate or high) would be 75% and 81%, respectively. The precision with which the amalgamated class is predicted is also considerably improved at 51%. As such, we have confidence that radon in those low BDA is likely to be low.

Increased probabilities of high vulnerabilities (moderate and high) were generally associated with colder winters, drier winters, close proximity to major river systems, and fluvioglacial and colluvial soil parent materials. Increased probabilities of high vulnerabilities associated with colder winters is consistent with the assumption that elevated concentrations are due to decreased ventilation and greater temperature difference between outdoor and indoor air (Nazaroff 1992; Al-Ahmady & Hintenlang 1994; Wang & Ward 2002; Kropat et al. 2014). Low probabilities of high vulnerabilities associated with winter precipitation totals over 780 mm suggest that the "capping effect" (Mose et al. 1991; Schumann et al. 1988) is not a major contributor to elevated indoor radon concentration provincially. It could still be a significant contributor at regional or individual scales. Increasing soil moisture reduces the distance with which radon can be transported and can reduce the availability of radon in the subsurface to be advected into homes, which may be the cause of this provincial trend (Schumann et al. 1988; Nazaroff 1992).

Increased probabilities of high radon vulnerabilities associated with closer distances to major river systems suggest that fluvial deposition of uranium enriched sediment could be contributing to elevated concentrations. The random forest algorithm does not allow us to specifically identify which river systems may be driving this trend, but we can infer that major river systems in the interior of the province are the most plausible candidates given that coastal regions of the province are associated with greater prevalence of low radon vulnerabilities. Our data include measurements taken in close proximity to large river systems such as the Nechako, North Thompson and Kootenay. The parent material of a soil is only one of many factors influencing the characteristics that affect radon transport in the subsurface such as porosity, permeability, or drainage (Schaetzl & Anderson 2005; Nazaroff 1992). Fluvioglacial and colluvial soil parent materials encompass an extensive and varied range of different conditions (Schaetzl & Anderson 2005), making it difficult to infer any general characteristics that would enhance radon transport processes. Unfortunately, the relationships derived from partial dependence plots do not capture interaction effects and, as a result, are likely an oversimplification of the main factors.

Although partial dependence plots can help elucidate the directionality of relationships between predictor variables and response variables, they are also limited when the predictor variables are highly generalized. Many of the ancillary datasets used were highly generalized, resulting in large areas of land being characterized by a few general features. Soil and bedrock predictor variables were highly generalized due the fact they were derived from simplified soil landscape polygons and simplified bedrock geology polygons, respectively. Furthermore, random error will be present in each model due to the fact they were derived from the conflation of disparate data sources, digitized at different spatial resolutions, with different zonal systems. The results of the partial dependence plots are better conceptualized as a baseline for further and more in-depth investigation of the specific variables associated with higher radon vulnerabilities.

The accuracy of the model would be improved if more detailed attribution were available for both soil and housing characteristics. The National Soil Landscapes data were simplified in data pre-processing by taking the dominant value for each variable for each soil landscape polygon. As a result, the soil conditions in each BDA were described by a set of highly generalized variables. Similarly, the housing characteristic data were not detailed enough to detect regional differences in housing construction that may increase or decrease radon concentrations (Appleton & Ball 2002). More detailed local housing information regarding characteristics of the home that may directly affect the influx of radon into the home such as the substructure type (basement, crawl-space, or slab on grade) are needed (Nazaroff & Nero 1984). Dominant age of home and proportion of homes in need of major repair did not capture these complexities.

The inclusion of a direct estimate for the quantity of parent material in the surficial material would likely improve the results. Though the British Columbia Drainage Geochemical Atlas is available and can provide an estimate of the uranium content of a drainage catchment (Lett et al. 2008), its measurements do not cover the north-eastern part of the province. Because the geochemical data do not cover the entirety of the province the dataset could not be included in the model. Though the model attempts to differentiate uranium content of surficial material by including bedrock type as a predictor variable, the simplified categories we used for rock types were likely too broad to capture meaningful differences in uranium content between them. Moreover, local variations in uranium content of overlying soil may be unrelated to the underlying bedrock based on the fact that majority of soils in the province are derived from materials that have been transported by either air, water or ice (Heung et al. 2014). Therefore the uranium content of soils whose parent materials are characterized by transportation will be controlled by their original source material (Gundersen & Schumann 1996).

Many of these limitations could be addressed by reducing the size of the study area. Our model requires that each dataset cover the full spatial extent of the province with consistent attribution. If the study area was reduced, more datasets with detailed attribution would be available for use. For example, the detailed soil surveys are digitized at much finer spatial resolutions than the Soil Landscapes of Canada and, depending on the survey, the soil polygons can be linked to quantitative estimates of their respective soil textures and porosities, which are key predictors of indoor radon concentrations (Hauri et al. 2012). Data availability will vary from region to region, however, and different models with unique input predictors would need to be developed under such a scenario.

The final map provides a method for delineating areas more susceptible to high indoor radon concentrations, and this can be used to support further epidemiologic inquiry. The geographic delineation of ordinal categories of radon risk can be a means of estimating relative radon exposure levels in epidemiological research (Hystad et al. 2014). Exposure estimates are made by grouping spatially referenced radon concentrations by administrative units that are large enough to provide seamless coverage of the study area (Hystad et al. 2014; Henderson et al. 2014). The size of the administrative units will hide the within-unit variation, increasing the uncertainty of results. By being able to estimate the expected relative exposure for unmeasured spatial units, this research can provide a means for using finer resolution spatial units to estimate geographic differences in radon exposure. Further research is needed to specifically investigate the effect of our indoor radon vulnerability classes on lung cancer in BC.

The results of this study can also be used to more efficiently allocate resources towards increasing radon awareness in the province. Currently, 58% of households in BC are unaware of the existence of radon (Statistics Canada 2012). Targeting resources for the purposes of increasing radon awareness and monitoring can be a more cost-effective means of reducing radon induced lung cancer (Appleton & Ball 2002). We have identified jurisdictions that could be prioritized for increasing radon awareness (Tables 2.5 and 2.6). Furthermore, the populations that are largely untested but are predicted to be at risk (Table 2.7) should be targeted for sampling campaigns to gauge the validity of these predictions.

2.6 Conclusions

We have presented a novel method for the creation of a predictive indoor radon vulnerability map. Increased probabilities of high radon vulnerabilities were generally found to be associated with colder winters, drier winters, close proximity to major river systems, and fluvioglacial and colluvial soil parent materials. The methods are broadly applicable to different regions throughout Canada and the world, and they provide a promising conceptual model for the creation of indoor radon vulnerability maps using existing geospatial data sources.

Acknowledgements

We would like to thank Paul Schiarizza of BC Ministry of Energy and Mines for consultation on geological categorization and Dr. Chuck Bulmer of the BC Ministry of Forests and Range for consultation on available soil datasets. We would also like to thank the BC Lung Association, Northern Health Authority, Donna Schmidt Foundation and Peter Chataway for sharing their data. This work has been supported by the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada.

References

Agriculture and Agri-Food Canada, 2013. Soil Landscapes of Canada (SLC). *Government of Canada*. Available at: http://sis.agr.gc.ca/cansis/nsdb/slc/index.html [Accessed May 15, 2014].

Al-Ahmady, K.K. & Hintenlang, D.E., 1994. Assessment of temperature-driven pressure differences with regard to radon entry and indoor radon concentration. In *AARST*. Atlantic City: The American Association of Radon Scientists and Technologists.

Appleton, J.D., 2007. Radon: sources, health risks, and hazard mapping. *Ambio*, 36(1), pp.85–89. Available at: http://www.jstor.org/stable/4315791.

Appleton, J.D. & Ball, T.., 2002. Geological radon potential mapping. In P. T. Bobrowsky, ed. *Geoenvironmental Mapping: Methods, Theory and Practice*. Exton, PA: A.A. Balkema Publishers, pp. 577–613.

Appleton, J.D. & Miles, J.C.H., 2010. A statistical evaluation of the geogenic controls on indoor radon concentrations and radon risk. *Journal of environmental radioactivity*, 101(10), pp.799–803. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19577346 [Accessed September 25, 2013].

Archer, K.J. & Kimes, R. V., 2008. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, 52(4), pp.2249–2260. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167947307003076 [Accessed September 25, 2013].

Arnold, B.W., 2006. Radon Transport. In C. K. Ho & S. W. Webb, eds. *Gas Transport in Porous Media*. Springer Netherlands, pp. 333–338.

Berk, R.A., 2008. *Statistical Learning from a Regression Perspective*, New York, NY: Springer New York. Available at: http://link.springer.com/book/10.1007%2F978-0-387-77501-2 [Accessed March 3, 2014].

Bissett, R.J. & McLaughlin, J.R., 2010. Radon. *Chronic diseases in Canada*, 29, Supple.

Breiman, L. et al., 1984. *Classification and Regression Trees*, Belmont, California: Wadsworth International Group.

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp.5–32. Available at: http://link.springer.com/article/10.1023/A:1010933404324 [Accessed September 25, 2013].

Charles, M., 2001. UNSCEAR Report 2000: Sources and Effects of Ionizing Radiation. *Journal of Radiological Protection*, 21(1), p.83.

Chen, C., Liaw, A. & Breiman, L., 2004. *Using random forest to learn imbalanced data*, University of California, Berkeley. Available at: http://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf [Accessed August 7, 2014].

Chen, J., Moir, D. & Whyte, J., 2012. Canadian population risk of radon induced lung cancer: a re-assessment based on the recent cross-Canada radon survey. *Radiation protection dosimetry*, 152(1-3), pp.9–13.

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), pp.37–46. Available at: http://epm.sagepub.com/cgi/doi/10.1177/001316446002000104 [Accessed July 11, 2014].

Cosma, C. et al., 2013. Radon and remediation measures near Băiţa-Ştei old uranium mine (Romania). *Acta Geophysica*, 61(4), pp.859–875. Available at: http://link.springer.com/10.2478/s11600-013-0110-8 [Accessed November 13, 2014].

Cui, Y. et al., 2013. British Columbia Digital Geology: BCGS Open File 2013-04. *BC Geological Survey*. Available at: http://www.empr.gov.bc.ca/MINING/GEOSCIENCE/PUBLICATIONSCATALOGUE/OPENFILES/2013/Pages/2013-4.aspx [Accessed April 15, 2014].

Cutler, D.R. et al., 2007. Random forests for classification in ecology. *Ecology*, 88(11), pp.2783–2792. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18051647.

De'ath, G., 2007. Boosted trees for ecological modeling and prediction. *Ecology*, 88(1), pp.243–251. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17489472.

Drolet, J.-P. et al., 2013. An approach to define potential radon emission level maps using indoor radon concentration measurements and radiogeochemical data positive proportion relationships. *Journal of environmental radioactivity*, 124, pp.57–67. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23660346 [Accessed April 23, 2014].

Drolet, J.-P. et al., 2014. Methodology developed to make the Quebec indoor radon potential map. *The Science of the total environment*, 473-474, pp.372–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24378928 [Accessed April 23, 2014].

Fatourechi, M. et al., 2008. Comparison of Evaluation Metrics in Classification Applications with Imbalanced Datasets. In *Machine Learning and Applications, 2008. ICMLA '08. Seventh International Conference on*. San Diego: IEEE, pp. 777–782. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4725065 [Accessed August 16, 2014].

Garbesi, K. et al., 1993. Soil-Gas Entry into an Experimental Basement : Model Measurement Comparisons and Seasonal Effects. *Environmental science & technology*, 27(3), pp.466–473.

GeoBC, 2014. Freshwater Atlas. Available at: http://geobc.gov.bc.ca/base-mapping/atlas/fwa/ [Accessed April 15, 2014].

Gerken, M. et al., 2000. Models for retrospective quantification of indoor radon exposure in case-control studies. *Health Physics*, 78(3), pp.268–278.

Gruber, V. et al., 2013. The European map of the geogenic radon potential. *Journal of radiological protection : official journal of the Society for Radiological Protection*, 33(1), pp.51–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23295644 [Accessed October 21, 2013].

Gunby, J.A. et al., 1993. Factors affecting indoor radon concentrations in the United Kingdom. *Health Physics*, 64(1), pp.2–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8416211.

Gundersen, L.C.S. & Schumann, R.R., 1996. Mapping the radon potential of the United States: Examples from the Appalachians. *Environment International*, 22, Supple(0), pp.829–837. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0160412096001900.

Hauri, D.D. et al., 2012. A prediction model for assessing residential radon concentration in Switzerland. *Journal of environmental radioactivity*, 112(0), pp.83–89. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22683900 [Accessed September 25, 2013].

Health Canada, 2009. Government of Canada Radon Guideline. Available at: http://www.hc-sc.gc.ca/ewh-semt/radiation/radon/guidelines_lignes_directrice-eng.php [Accessed March 24, 2015].

Henderson, S.B. et al., 2014. Differences in lung cancer mortality trends from 1986-2012 by radon risk areas in British Columbia, Canada. *Health Physics*, 106(5), pp.608–613. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24670910 [Accessed October 16, 2014].

Heung, B., Bulmer, C.E. & Schmidt, M.G., 2014. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. *Geoderma*, 214-215, pp.141–154. Available at: http://www.sciencedirect.com/science/article/pii/S0016706113003443 [Accessed October 31, 2014].

Hystad, P. et al., 2014. Geographic variation in radon and associated lung cancer risk in Canada. *Canadian journal of public health = Revue canadienne de santé publique*, 105(1), pp.e4–e10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24735695.

Ielsch, G. et al., 2010. Mapping of the geogenic radon potential in France to improve radon risk management: methodology and first application to region Bourgogne. *Journal of environmental radioactivity*, 101(10), pp.813–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20471142 [Accessed September 25, 2013].

Ioannides, K. et al., 2003. Soil gas radon: a tool for exploring active fault zones. *Applied Radiation and Isotopes*, 59(2-3), pp.205–213. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0969804303001647 [Accessed November 13, 2014].

Jones, L.D., 1990. *Uranium and Thorium Occurences in British Columbia*, Available at: http://www.empr.gov.bc.ca/Mining/Geoscience/PublicationsCatalogue/OpenFiles/1990/Documents/OF1990-32.pdf.

Kitto, M.E., 2005. Interrelationship of indoor radon concentrations, soil-gas flux, and meteorological parameters. *Journal of Radioanalytical and Nuclear Chemistry*, 264(2), pp.381–385. Available at: http://link.springer.com/10.1007/s10967-005-0725-6.

Kropat, G. et al., 2014. Major influencing factors of indoor radon concentrations in Switzerland. *Journal of environmental radioactivity*, 129, pp.7–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24333637 [Accessed April 23, 2014].

Lesack, P., 2012. Combined Dissemination Block Digital Cartographic File and Geographic Attribute File, 2011 [2012] [Census of Canada, 2011]. *Census of Canada. 2011*. Available at: http://hdl.handle.net/10573/42747.

Lett, R.E. et al., 2008. *GeoFile 2008-1: A Drainage Geochemical Atlas for British Columbia*, Available at: http://www.empr.gov.bc.ca/Mining/Geoscience/PublicationsCatalogue/GeoFiles/Pages/2008-1.aspx.

Liaw, A. & Wiener, M., 2002. Classification and Regression by randomForest. *R News*, 2(3), pp.18–22. Available at: http://cran.r-project.org/doc/Rnews/ [Accessed July 22, 2014].

Malmqvist, L., Isaksson, M. & Kristiansson, K., 1989. Radon migration through soil and bedrock. *Geoexploration*, 26(2), pp.135–144. Available at: http://www.sciencedirect.com/science/article/pii/0016714289900586 [Accessed November 13, 2014].

Michel, J., 1987. Sources. In C. R. Cothern & J. E. Smith, eds. *Environmental Radon*. New York: Plenum Press, pp. 81–130.

Miles, J.C.H. & Appleton, J.D., 2005. Mapping variation in radon potential both between and within geological units. *Journal of Radiological Protection*, 25(3), pp.257–276. Available at: http://iopscience.iop.org/0952-4746/25/3/003/ [Accessed September 24, 2013].

Mose, D., Mushrush, G. & Chrosniak, C., 1991. Seasonal Indoor Radon Variations Related to Precipitation. *Environmental and Molecular Mutagenesis*, 17(4), pp.223–230.

Nazaroff, W.W., 1992. Radon transport from soil to air. *Reviews of Geophysics*, 30(2), pp.137–160. Available at: http://dx.doi.org/10.1029/92RG00055 [Accessed November 14, 2013].

Nazaroff, W.W. & Nero, A.V., 1984. Transport of Radon From Soil Into Residences. In *Third International Conference on Indoor Air Quality and Climate*. Stoockholm, Sweden.

O'Reilly, G.A. et al., 2013. Map showing the potential for radon in indoor air in Nova Scotia. *Nova Scotia Department of Natural Resources, Mineral Resources Branch, Open File Map ME 2013-028*. Available at: http://novascotia.ca/natr/meb/data/mg/ofm/pdf/ofm_2013-028_d486_dp.pdf.

Prasad, A.M., Iverson, L.R. & Liaw, A., 2006. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*, 9(2), pp.181–199.

Radon Environmental Management Corp., 2011. Radon Potential Map of Canada. Available at: http://www.radoncorp.com/pdf/presentationMappingPublic.pdf [Accessed November 14, 2013].

Rauch, S.A. & Henderson, S.B., 2013. A comparison of two methods for ecologic classification of radon exposure in British Columbia: residential observations and the radon potential map of Canada. *Canadian journal of public health*, 104(3), pp.e240–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23823889.

Saccomanno, G. et al., 1988. Relationship of radioactive radon daughters and cigarette smoking in the genesis of lung cancer in uranium miners. *Cancer*, 62(7), pp.1402–1408.

Schaetzl, R.J. & Anderson, S., 2005. *Soils : genesis and geomorphology*, New York: Cambridge University Press.

Scheib, C. et al., 2013. Geological controls on radon potential in England. *Proceedings of the Geologists' Association*, 124(6), pp.910–928. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0016787813000291 [Accessed April 23, 2014].

Schumann, R.R., Owen, D.. & Asher-Bolinder, S., 1988. Weather Factors Affecting Soil-Gas Radon Concentrations at a Single Site in the Semiarid Western U.S. In *Proceedings of the 1988 International Radon Symposium*. Denver, CO: The American Association of Radon Scientists and Technologists.

Shweikani, R., Giaddui, T.G. & Durrani, S. a., 1995. The effect of soil parameters on the radon concentration values in the environment. *Radiation Measurements*, 25(1-4), pp.581–584. Available at: http://linkinghub.elsevier.com/retrieve/pii/135044879500188K [Accessed November 13, 2014].

Statistics Canada, 2011. Census Dictionary 2011. *Statistics Canada Catalogue no. 98-301-X2011001*. Available at: http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/98-301-X2011001-eng.pdf.

Statistics Canada, 2013. National Household Survey (NHS) Profile, 2011. [2013]. *Statistics Canada*. Available at: http://hdl.handle.net/10573/42928.

Statistics Canada, 2012. Table153-0098 - Households and the environment survey, knowledge of radon and testing, Canada and provinces, every 2 years (percent). *CANSIM*. Available at: http://www5.statcan.gc.ca/cansim/pick-choisir?lang=eng&p2=33&id=1530098 [Accessed November 7, 2013].

Strobl, C. et al., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC bioinformatics*, 8, p.25. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1796903&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Varley, N.R. & Flowers, A.G., 1998. Indoor radon prediction from soil gas measurements. *Health Physics*, 74(6), pp.714–718. Available at: http://journals.lww.com/health-physics/Abstract/1998/06000/Indoor_Radon_Prediction_from_Soil_Gas.9.aspx [Accessed September 3, 2014].

Vasilyev, A.V. & Zhukovsky, M.V., 2013. Determination of mechanisms and parameters which affect radon entry into a room. *Journal of environmental radioactivity*, 124, pp.185–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23811128 [Accessed November 14, 2013].

Verger, P. et al., 1994. Use of field measurements in radon mapping in France. *Radiation protection dosimetry*, 56(1-4), pp.225–229. Available at: http://rpd.oxfordjournals.org/content/56/1-4/225.short [Accessed October 16, 2013].

Wang, F. & Ward, I.C., 2002. Radon entry, migration and reduction in houses with cellars. *Building and Environment*, 37(11), pp.1153–1165. Available at: http://linkinghub.elsevier.com/retrieve/pii/S036013230100097X.

Wang, T. et al., 2012. ClimateWNA—High-Resolution Spatial Climate Data for Western North America. *Journal of Applied Meteorology and Climatology*, 51(1), pp.16–29. Available at: http://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-11-043.1 [Accessed October 27, 2014].

Washington, J.W. & Rose, A.W., 1990. Regional and temporal relations of radon in soil gas to soil temperature and moisture. *Geophysical Research Letters*, 17(6), pp.829–832. Available at: http://dx.doi.org/10.1029/GL017i006p00829 [Accessed May 15, 2014].

World Health Organization, 2009. *WHO Handbook on Indoor Radon: A Public Health Perspective* H. Zeeb & F. Shannoun, eds., Geneva.

Table 2.1 - The geologic, pedologic, climate and housing predictor variables used to predict indoor radon vulnerability class

Variable	Rationale	References	Source
Simplified Bedrock Lithology Class	Bedrock geology can be major determinant in regional indoor radon vulnerability; can control quantity and spatial distribution of uranium; emanation of radon from bedrock can contribute to subsurface radon concentrations dependent upon overburden characteristics.	(Malmqvist et al. 1989; Appleton & Miles 2010; Scheib et al. 2013)	BC Digital Geology: Open File 2013-4 (Cui et al. 2013)
Geologic Fault Presence	Geologic faults can potentially increase rate of radon transport towards the surface by providing pathways for upward movement of radon containing soil gas	(Ioannides et al. 2003; Appleton 2007; Ielsch et al. 2010)	
Dominant Soil Parent Material Mode of Deposition	Soil characteristics that affect radon emanation and transport (grain size, porosity, permeability etc.) are partly controlled by the soils parent material; uranium content and distribution in soils that were deposited through transportation likely more related to soil parent material than underlying bedrock	(Nazaroff 1992; Gundersen & Schumann 1996; Schaetzl & Anderson 2005; Arnold 2006)	Soil Landscapes of Canada Version 2.2 (Agriculture and Agri-Food Canada 2013)
Dominant Soil Drainage Class	The drainage class of a soil will affect radon emanation rates and transport processes by affecting the soil moisture.	(Nazaroff 1992; Arnold 2006; Scheib et al. 2013; Shweikani et al. 1995)	
Dominant Soil Rooting Depth Class	Rooting depth gives an approximation of the depth to bedrock, or impermeable layer; the rooting depth class will combine with other soil characteristics to affect soil moisture and the influence of bedrock geology on radon concentrations; depth to the bedrock can influence the proportion of indoor radon contributed from underlying bedrock.	(Malmqvist et al. 1989; Nazaroff 1992; Schaetzl & Anderson 2005; Arnold 2006; Shweikani et al. 1995)	
Dominant Soil Coarse Fragment Content Class	Coarse fragment content of a soil will affect its porosity, permeability and drainage characteristics and interact with other soil variables to affect radon emanation and transport processes.	(Schaetzl & Anderson 2005; Nazaroff 1992; Arnold 2006)	
Dominant Soil Kind of Surface Material	The dominant surface material, whether organic soil or rock for example, could affect radon emanation and transport properties.		
Average Winter Temperature	Average winter temperature and average winter precipitation approximate the winter climate of a region and majority of radon concentration data were recorded in winter months; prevailing climate will affect radon emanation as well as its transport towards the subsurface; ambient temperatures can also serve as a proxy for average air exchange rates between outdoor air and a home as it will influence the "stack effect" in addition to ventilation characteristics of a home.	(Nazaroff 1992; Washington & Rose 1990; Schumann et al. 1988; Al-Ahmady & Hintenlang 1994; Vasilyev & Zhukovsky 2013; Kitto 2005; Garbesi et al. 1993)	Climate Western North America (Wang et al. 2012)
Average Winter Precipitation			
Distance to Nearest River	Hydrological systems can influence amount and distribution of uranium in river deposits if downstream from Uranium rich geologic formations; BC contains known uranium deposits upstream from urban areas built on river deposits.	(Cosma et al. 2013; Jones 1990)	Freshwater Atlas of BC (GeoBC 2014)
Dominant Age of Home	Housing characteristics such as design, age type of construction material and ventilation will have an impact on an indoor radon concentration; Dominant age of home and proportion of major repairs needed, can serve as a coarse estimate of regional household characteristics.	(Appleton 2007; Gunby et al. 1993; Verger et al. 1994; Gerken et al. 2000; Hauri et al. 2012)	NHS of Canada 2011/ CDBDCF and GAF 2011(Statistics Canada 2013; Lesack 2012)
Proportion of Homes in Need of Major Repairs			

BC = British Columbia, NHS= National Household Survey, CDBDCF = Combined Dissemination Block Digital Cartographic File, GAF = Geographic Attribute File.

Table 2.2 - Summary of pre-processing and conflation details required for each predictor variable selected.

Data Source	**Description**	**Pre-processing details**	**Conflation details**	**Predictor Variable**
BC Digital Geology: Open File 2013-4	Bedrock Geology polygons	Bedrock geology collapsed from 187 rock types into 12 broad lithological categories.	N/A	Simplified Bedrock Lithology Class
BC Digital Geology: Open File 2013-4	Geologic fault polylines	N/A	Presence/absence of overlapping fault line assigned to each BDA.	Geologic Fault Presence
Soil Landscapes of Canada Version 2.2	Soil Landscapes of Canada for BC as polygons	Each landscape polygon defined by several “components” and therefore contained multiple values for each variable. Each variable were simplified in order to derive one value for each polygon that represented the dominant value for that variable. Dominance was defined through the summation of a values percent area across all components.	SLC polygons were first intersected by BDA polygons and total area for each resulting fragment calculated. Each BDA assigned the value for each SLC variable based on which value took up the largest area within that BDA.	Soil Parent Material Mode of Deposition
				Soil Drainage Class
				Soil Rooting Depth Class
				Soil Coarse Fragment Content Class
				Soil Kind of Surface Material
Climate Western North America	Climate normal temperature and precipitation point data; 1 km resolution	N/A	Mean winter temperature and mean total winter precipitation was assigned to each BDA based on the average of all temperature points falling within 4km of a BDA. 4km selected as it was the distance at which 99% of BDA’s would have at least one associated temperature and precipitation value.	Average Winter Temperature
		N/A		Average Winter Precipitation
Freshwater Atlas of BC	Stream Network polylines	River polygons whose geometry intersected stream network polylines with a Strahler order of 7 or greater were selected as major river systems in BC.	Euclidean distance calculation from the geometry of the nearest major river system used to calculate distance.	Distance to Nearest River
	River polygons			
NHS of Canada 2011/ CDBDCF and GAF 2011	Household Characteristic data at Census Dissemination Area	Data not available for 5.4% of Dissemination Areas. Variable values estimated by multiply census dwelling counts by provincial average for available Dissemination areas.	N/A	Dominant Age of Home
	Dwelling counts at Census Dissemination Area			Proportion of Homes in Need of Major Repairs

BC= British Columbia, SLC= Soil Landscapes of Canada, BDA = Bedrock Dissemination Area, NHS= National Household Survey, CDBDCF = Combined Dissemination Block Digital Cartographic File, GAF = Geographic Attribute File.

Table 2.3 - Out-of-bag (OOB) estimates of classifier performance compared to hold-out validation (HOV).

	Metric	OOB	HOV
	Kappa Score	0.34	0.32
Class Accuracy	Low	0.75	0.74
	Moderate	0.44	0.44
	High	0.54	0.48
	Average	0.58	0.55
Class Precision	Low	0.92	0.92
	Moderate	0.29	0.28
	High	0.30	0.26
	Average	0.50	0.49

Table 2.4 - Confusion matrix for the balanced random forest model based on out-of-bag predictions.

		Predicted Class			
		Low	Moderate	High	Class Accuracy
Actual Class	Low	594	160	42	0.75
	Moderate	44	79	57	0.44
	High	5	31	42	0.54
	Precision	0.92	0.29	0.30	Overall Accuracy = 0.68

Table 2.5 -Regional indoor radon vulnerability by Census Division.

Census Division	% Low	% Moderate	% High	% Moderate or High
Central Kootenay	9	62	29	91
Kootenay Boundary	10	82	8	90
Columbia-Shuswap	31	68	1	69
Fraser-Fort George	33	57	10	67
North Okanagan	35	63	2	65
East Kootenay	37	56	7	63
Okanagan-Similkameen	46	49	5	54
Squamish-Lillooet	51	44	5	49
Central Okanagan	52	43	5	48
Thompson-Nicola	55	39	6	45
Peace River	66	33	1	34
Kitimat-Stikine	68	29	3	32
Stikine	69	29	2	31
Cariboo	70	24	6	30
Bulkley-Nechako	73	24	3	27
Fraser Valley	88	11	1	12
Northern Rockies	91	3	6	9
Central Coast	94	4	2	6
Sunshine Coast	99	1	0	1
Mount Waddington	99	1	0	1
Cowichan Valley	99	0	1	1
Strathcona	99	1	0	1
Skeena-Queen Charlotte	99	1	0	1
Capital	100	0	0	0
Greater Vancouver	100	0	0	0
Alberni-Clayoquot	100	0	0	0
Comox Valley	100	0	0	0
Nanaimo	100	0	0	0
Powell River	100	0	0	0

Table 2.6 - Local indoor radon vulnerability. The 30 most vulnerable population centres by proportion of Bedrock Dissemination Areas classified as moderate or high.

Population Centre	% Low	% Moderate	% High	% Moderate or High
Grand Forks	0	50	50	100
Salmo	0	50	50	100
Rossland	0	83	17	100
Mackenzie	0	89	11	100
Lillooet	0	92	8	100
Fruitvale	0	93	7	100
Sicamous	0	100	0	100
Tumbler Ridge	0	100	0	100
Revelstoke	4	96	0	96
Castlegar	5	30	65	95
Golden	12	88	0	88
Sparwood	17	66	17	83
Nakusp	20	60	20	80
Ashcroft	20	80	0	80
Nelson	35	52	13	65
Prince George	37	42	21	63
Vernon	43	54	3	57
Duck Lake	43	50	7	57
Trail	44	39	17	56
Summerland	46	54	0	54
Kimberley	50	50	0	50
Oliver	52	18	30	48
Enderby	56	44	0	44
Creston	60	40	0	40
Cache Creek	62	38	0	38
Elkford	62	25	13	38
Penticton	63	35	2	37
Osoyoos	67	0	33	33
Atlin	67	33	0	33
Chase	67	33	0	33

Table 2.7 - Population centres predicted to be high risk and are in need of further sampling

Population Centre	% BDA's Sampled	% Moderate or High
Lillooet	0	100
Mackenzie	0	100
Sicamous	0	100
Tumbler Ridge	0	100
Sparwood	0	83
Cache Creek	0	38
Revelstoke	4	96
Summerland	4	54
Oliver	4	48
Rossland	6	100
Enderby	6	44
Osoyoos	6	33
Chase	9	33
Princeton	11	25
Elkford	12	38
Fruitvale	13	100
Duck Lake	14	57
Fernie	17	25
Ashcroft	20	80
Penticton	23	37
Salmo	25	100
Golden	25	88
Cowichan Bay	25	25
Creston	27	40
Grand Forks	30	100
Kelowna	31	33
Vernon	38	57
Nakusp	40	80

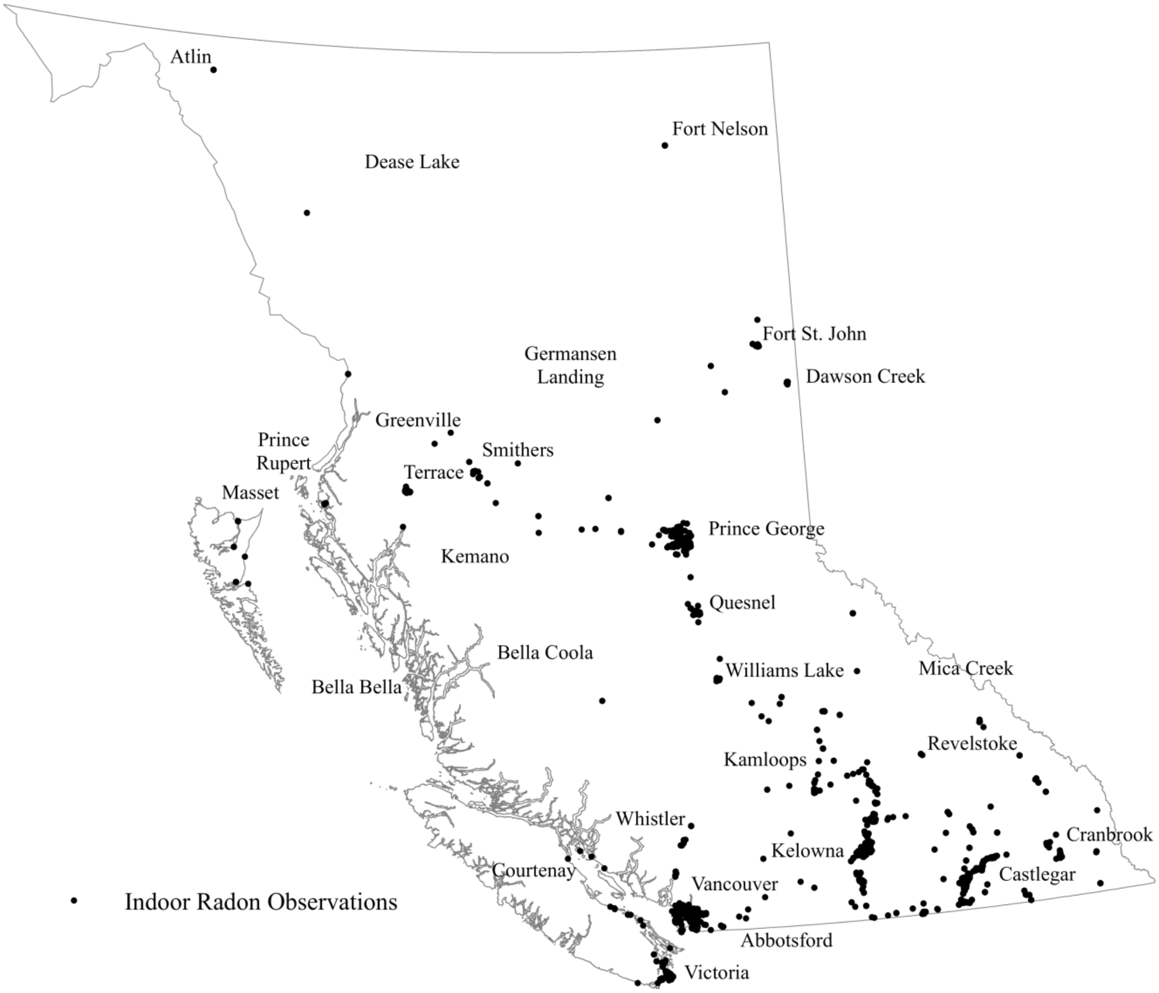

Figure 2.1 - Study area, British Columbia, Canada. The spatial distribution of all 4352 successfully geocoded indoor radon concentration measurements is also shown.

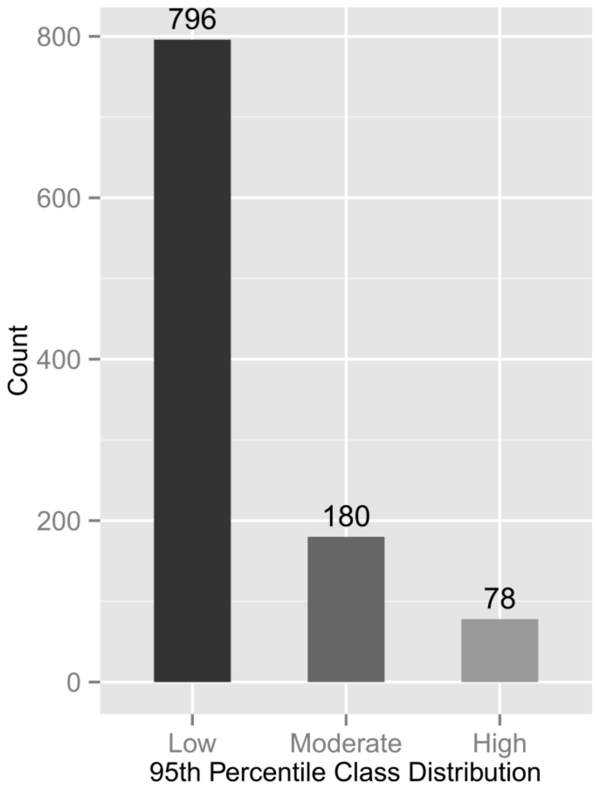

Figure 2.2 - The resulting indoor radon vulnerability class distribution by 95th percentile radon concentration of each spatial unit.

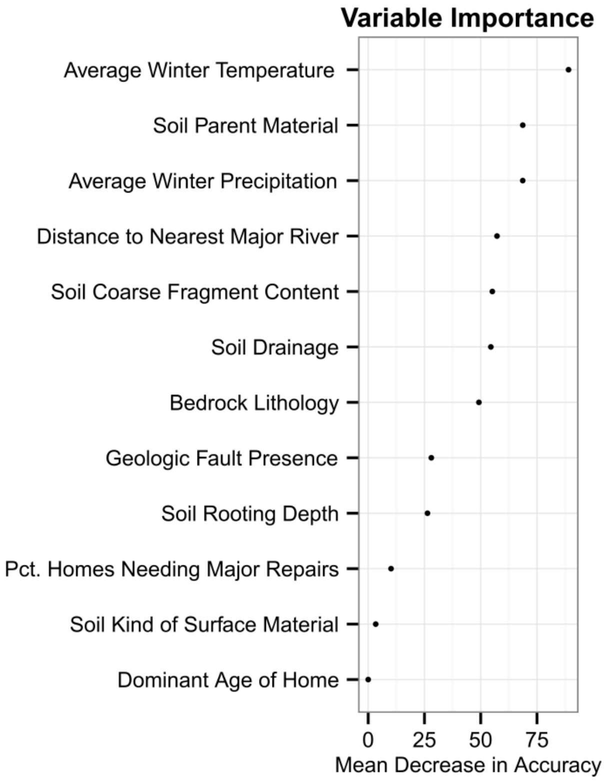

Figure 2.3 - Variable importance plots. Variable importance is measured by the mean decrease in predictive accuracy.

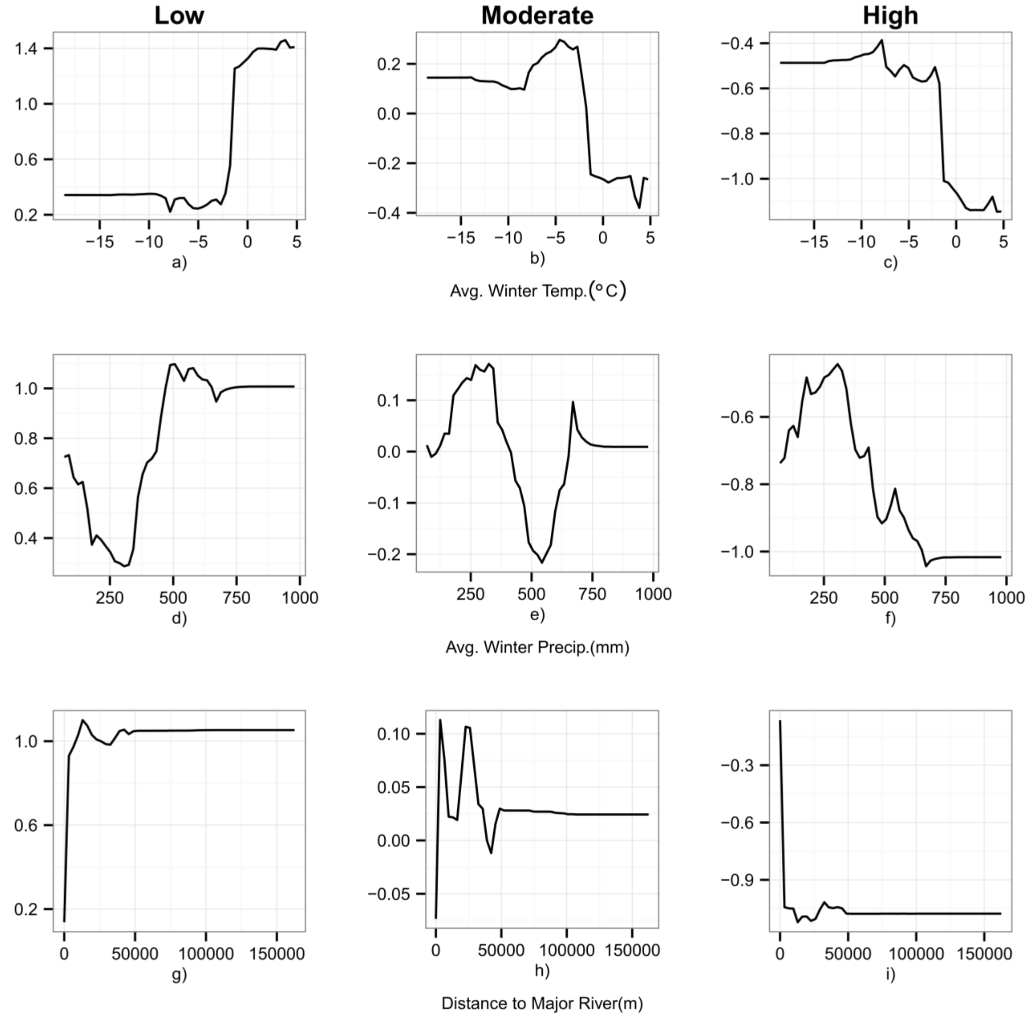

Figure 2.4 - Partial dependence plots: important numeric predictors. Partial dependence plots for average winter temperature (a–c), average total winter precipitation (d–f), and distance to nearest major river (g–i). The plotted functions are interpreted as the increasing or decreasing probability of a classification for the values of the variable of interest, holding all other variables constant. For example, in (a), the probability of a low vulnerability rating is constant and low for average winter temperature values from approximately −18 °C to approximately −2 °C, at which point the probability of a low vulnerability rating starts to increase rapidly. This plot therefore indicates that for a theoretical BDA defined by the average value for all other predictor variables, the probability that it is a low vulnerability rating is lower if it had a colder average winter temperature and higher for average winter temperatures greater than −2 °C.

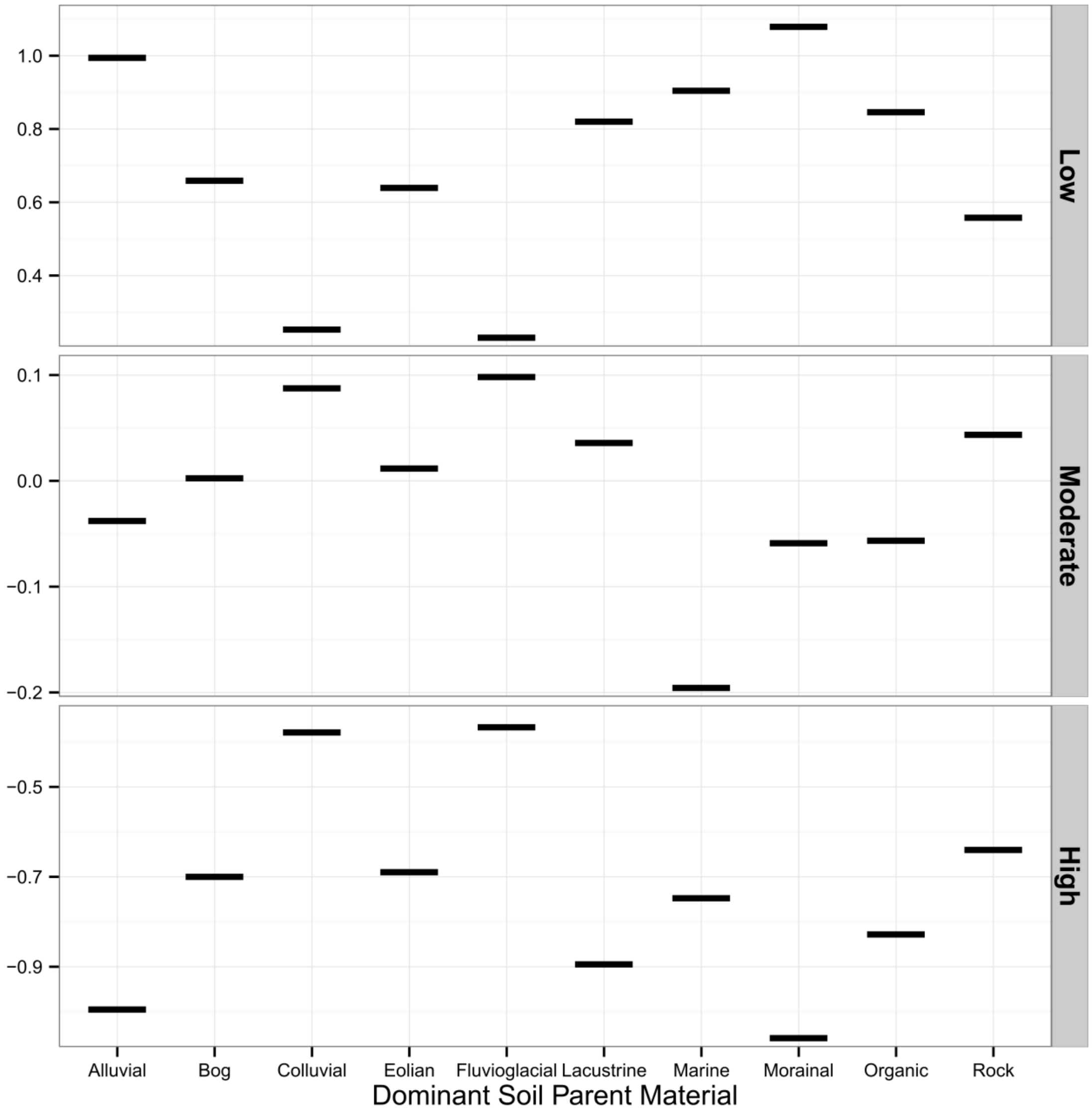

Figure 2.5 - Partial dependence plots: soil parent material. The plotted functions are interpreted as the increasing or decreasing probability of a certain classification for the values of the variable of interest, holding all other variables constant. For example, given a theoretical BDA that is defined by the average value of all predictor variables with the exception of dominant soil parent material, the probability that it has a low indoor radon vulnerability is lowest if its dominant soil parent material is fluvioglacial or colluvial, and the probability of a low vulnerability is highest if its dominant soil parent material is morainal or alluvial.

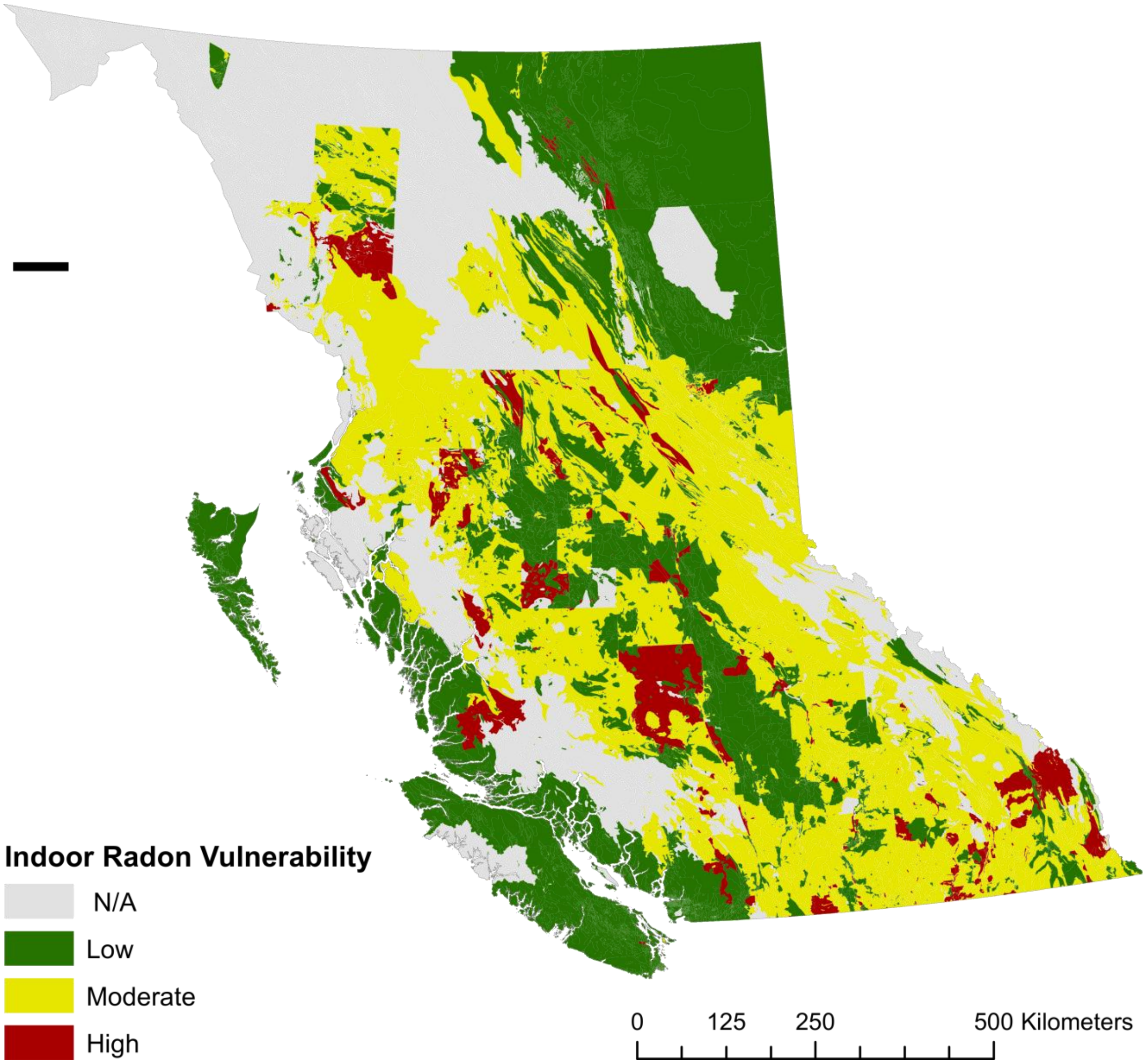

Figure 2.6 - Indoor radon vulnerability map. Indoor radon vulnerability map derived from predictions made using a balanced random forest algorithm. Only 1% of Bedrock Dissemination Areas within population centres could not be predicted for.

3.0 DIFFERENT RADON THRESHOLDS AND THEIR ASSOCIATIONS WITH GEOGRAPHIC RISK CHARACTERIZATION AND LUNG CANCER MORTALITY TRENDS IN BRITISH COLUMBIA, CANADA

3.1 Abstract

There is no safe concentration of radon gas, but guideline values provide threshold concentrations that are often used to delineate geographic areas at higher risk. These values vary between different regions, countries, and organizations, which can lead to differential classification of risk. For example the World Health Organization suggests a value of 100 Bq/m^3 while Health Canada recommends 200 Bq/m^3. Our objective was to examine how different thresholds were associated with geographic risk and lung cancer mortality trends in British Columbia, Canada. Eight threshold values between 50 and 600 Bq/m^3 were identified, and classes of regional radon vulnerability were defined based on whether the observed 95th percentile radon concentration was above or below each value. A balanced random forest algorithm was used to model vulnerability, and the results were mapped. We compared high vulnerability areas, their estimated populations, and differences in lung cancer mortality trends stratified by sex and smoking prevalence. Classification accuracy improved as the threshold concentrations decreased and the spatial area classified as high vulnerability increased. The majority of the population lived within areas of lower vulnerability regardless of the threshold value. Thresholds as low as 50 Bq m^{-3} were associated with higher lung cancer mortality, even in areas with relatively low smoking prevalence. Lung cancer mortality trends were increasing through

time for women, while decreasing for men. Radon contributes to lung cancer in British Columbia. The majority of the population is exposed to concentrations below the Canadian radon guideline, and the authors suggest a reference level as low as 50 Bq m^{-3} is justified for the province.

3.2 Introduction

Radon is a colourless, odourless, radioactive noble gas produced by the breakdown of naturally occurring uranium within the surface of the Earth. Radon is estimated to be a factor in over 3,000 lung cancer deaths in Canada per year (Chen et al. 2012). Radon atoms can be transported from their source and into homes where concentrations can accumulate. Though there is no radon concentration at which there is no risk of developing lung cancer, the probability of developing lung cancer increases with exposures to higher concentrations (Darby et al. 2005). Individuals who smoke are at an even greater risk due to the synergistic effects of radon and cigarette smoke (Saccomanno et al. 1988).

In light of the public health threat posed by residential radon, varying concentration thresholds have been set by different regions, countries, and organizations throughout the world. Here we define a threshold value as the concentration above which remedial action to reduce radon is recommended. These thresholds do not imply a level of safety, but rather a concentration below which the risk of developing radon-induced lung cancer is considered acceptably small. Threshold values are chosen to maximize the overall reduction in lung cancer mortality while considering what is practical to achieve in a majority of homes in a given jurisdiction (Chen et al. 2012). Though the World

Health Organization (WHO) recommends a concentration threshold of 100 Bq m^{-3}, other established thresholds are typically higher. For example, the USA uses a threshold of 148 Bq m^{-3}, Canada uses a threshold of 200 Bq m^{-3}, and the European Union uses thresholds ranging between 200 and 400 Bq m^{-3} (World Health Organization 2007; Synnott 2005).

Radon concentration thresholds are used to inform policy and to enable risk communication. For example, radon risk maps identify areas prone to high radon concentrations. Such maps allow for geographic targeting of radon awareness, testing, and remediation campaigns, and they can also encourage new policies (World Health Organization 2009). The radon risk map of Ireland divided the country into grid squares and mapped the proportion of homes whose indoor concentration exceeded the national threshold of 200 Bq m^{-3} (Long & Fenton 2011). Those grid squares where >10% of homes were estimated to exceed the national threshold were designated as high radon areas (HRAs). After completion of the map, an updated building code required that all new buildings be fitted with a standby radon sump that could be installed at a later date. Buildings within the HRAs were required to install a radon barrier in addition to the standby sump (Long & Fenton 2011). The choice of threshold concentration for use in such mapping is generally based on the recommended threshold used in the geographic jurisdiction for which the map is being prepared. However, the choice of threshold will affect the size of the spatial area classified as high risk and any resulting policy, and it may affect the accuracy of the classification. If the concentration threshold in Ireland was higher or lower than 200 Bq m^{-3} it would have changed the designation of HRAs and the requirement of additional radon protection measures in new buildings.

Ultimately, the objective of any radon risk map is to effectively delineate areas at risk of high indoor radon concentrations and, therefore, greater rates of radon-induced lung cancer. Temporal trends in the annual crude ratio of lung cancer mortality can be used as an exploratory tool for investigating spatial differences in radon distribution (Henderson et al. 2014). As such, we expect that an effective radon risk map would show distinct differences in lung cancer mortality trends between regions defined as higher and lower risk. However, the delineation of higher and lower risk areas depends on the chosen concentration threshold.

Our objective is to explore how different radon concentration thresholds are associated with the accuracy of risk classification, geographic areas classified as higher or lower risk, populations classified as higher or lower risk, and observed temporal trends in lung cancer mortality. Understanding these relationships has important implications for informing policy on appropriate concentration thresholds. Following Branion-Calles et al. (2015) we map the radon vulnerability of geologic units using eight thresholds ranging from 50 to 600 Bq m^{-3}. Radon vulnerability refers to the potential for a geographic area to exceed a specified concentration threshold. Maps of indoor radon vulnerability are then used to explore the association between radon concentration thresholds and lung cancer mortality trends stratified by sex and smoking prevalence.

3.3 Study Area

The study area was the province of British Columbia (BC), on the west coast of Canada. Many parts of BC are prone to high radon concentrations, including both small and large communities, primarily within the interior and northern regions (Branion-Calles et al. 2015; Henderson et al. 2014; Henderson et al. 2012). In the 2011 census BC had a population of approximately 4.4 million people, 3.79 million living in urban areas and 609,000 living in rural areas. The majority of the population lives within a small area in the southwestern region (Figure 3.1).

3.4 Data

3.4.1 Bedrock Dissemination Areas

The province was divided into 36,061 mapping units based on an intersection of census dissemination areas and simplified bedrock lithology. Each mapping unit was labelled as a “Bedrock Dissemination Area” (BDA) and was assumed to represent a homogenous spatial area with respect to the environmental and housing conditions that would affect potential susceptibility to high radon concentrations. In order to enable the classification of indoor radon risk each BDA were associated with variables derived from overlapping geospatial datasets including: indoor radon concentration data, geologic, soil, meteorological, hydrological and neighbourhood housing data.

Indoor Radon Concentrations and Vulnerability Class

Indoor radon concentration data are archived at the BC Center for Disease Control (BCCDC) and consist of five disparate surveys conducted between 1991 and 2014. Surveys were conducted by the BCCDC, the Northern Health Authority, the BC Lung Association, the Donna Schmidt Foundation and a private contractor. The BCCDC survey consisted of two surveys, the first of which was designed to oversample in areas with known high ambient radiation levels and the second oversampled in areas with moderate ambient radiation levels. The remaining four surveys collected measurements through volunteers. Attributes common to each survey were a six digit postal code, date of test period and the observed radon concentration. Each indoor radon concentration observation was assigned a geographic coordinate based on its associated six digit postal code and date through geocoding. A total of 4352 indoor radon concentrations were successfully assigned a spatial location.

Indoor radon concentration values were used to construct the response variable for the purposes of statistical classification. We used the same classification of indoor radon risk, termed indoor radon vulnerability, developed in previous work (Branion-Calles et al. 2015). Multiple binary response variables were defined where each BDA with observed concentrations was assigned an indoor radon vulnerability classification based on the following thresholds: 50, 100, 150, 200, 300, 400, 500, and, 600 Bq m^{-3}. These values were selected based on the premise that they cover the range of radon threshold concentrations used in countries throughout the world and represent multiple scenarios of provincial radon risk. Of the total 36,051 BDAs, there were 1,054 that contained at least one indoor radon measurement. These BDAs comprised the training dataset, leaving the remaining 34,972 BDAs to be classified using model results. A binary indicator of either high or low vulnerability was assigned to each BDA in the training dataset based on whether the observed 95th percentile radon measurement was greater or less than each concentration threshold. This resulted in eight different class distributions for the training dataset (Figure 3.2).

Independent Variables

The potentially predictive independent variables were selected based on their theoretical association with local radon concentrations, either individually or in combination. For example, soils that allow for a greater rate of radon transport towards the subsurface may increase the quantity of radon available to be transported into homes (Arnold 2006; Nazaroff 1992; Shweikani et al. 1995). Similarly, colder ambient temperatures may increase the difference between indoor air and outdoor air and therefore increase the rate at which soil gas is drawn indoors (Al-Ahmady & Hintenlang 1994). The transport of radon into homes can be further affected by specific housing characteristics, such as cracks in the foundation and the ventilation rate (Appleton 2007). Although we did not have such data about the individual homes, we do have neighbourhood data on average home age and state of repair from the 2011 National Household Survey (Statistics Canada 2013).

The specific independent variables we constructed for each BDA were: (1) simplified bedrock lithological class from the BC Digital Geology Open File (BCDGOF) (Cui et al. 2013); (2) geologic fault presence from the BCDGOF (Cui et al. 2013) ; (3) dominant soil parent material from the Soil Landscapes of Canada Version 2.2 (SLC) (Agriculture and Agri-Food Canada 2013); (4) dominant soil drainage class from the SLC (Agriculture and Agri-Food Canada 2013); (5) dominant rooting depth class from the SLC (Agriculture and Agri-Food Canada 2013); (6) dominant soil coarse fragment content from the SLC (Agriculture and Agri-Food Canada 2013); (7) dominant kind of surface material from the SLC (Agriculture and Agri-Food Canada 2013); (8) average winter temperature (climate normals) from the Climate Western North America database (CWNA)(Wang et al. 2012); (9) average winter precipitation (climate normals) from the

CWNA(Wang et al. 2012); (10) distance to nearest major river from the Freshwater Atlas of BC (GeoBC 2014); (11) dominant age of home from the 2011 National Household Survey of Canada (NHSC) (Statistics Canada 2013); (12) proportion of homes in need of major repairs from the 2011 NHSC (Statistics Canada 2013); and (13) distance to nearest uranium mineralization. The last variable was not included our previous work (Branion-Calles et al. 2015), but homes built on materials with high uranium content may be more prone to higher radon concentrations (Appleton 2007). Distance to nearest uranium mineralization was obtained by calculating the Euclidean distance from spatially referenced locations of known mineral occurrences with a significant quantity of uranium. Mineral occurrence data in British Columbia are available from the British Columbia Ministry of Energy and Mines (BC Ministry of Energy and Mines 2015). Each mineral occurrence in the database had a spatial location as well as a description of the present elements or substances that had economic potential. Detailed rationale and methods for the other 12 variables is given elsewhere (Branion-Calles et al. 2015).

Population Estimates

Estimates of the resident population for each BDA were made using data from the Dissemination Area (DA) level of the 2011 national census. Theses spatial areas generally include between 400-700 persons. Because BDAs represent the intersections between DAs and the bedrock geography, BDAs are smaller than their parent DAs. The population of a BDA was therefore estimated based on the proportion of its total area relative to the area of its parent DA. For example, if a 2 km^2 DA had 500 residents and it was split into two 1 km^2 BDAs, each would be assigned an estimated population of 250.

3.4.2 Mortality Records

Mortality records provided by the provincial Vital Statistics agency are archived at the BCCDC. These data include information about age, sex, underlying cause of death, and postal code of residence for each decedent. The underlying cause of death is coded according to the International Classification of Diseases 10th Revisions (ICD-10). We extracted deaths due to all natural causes (excluding ICD-10 codes starting with T through Y) and lung cancer (ICD-10 code C34) for adults aged 20 and over from 1998 through 2012. Each death was anonymously mapped by geocoding its residential 6-digit postal code.

3.4.3 Smoking Prevalence

There are 89 local health areas (LHAs) in BC, and these are the smallest spatial area at which health services are administered. Data on smoking prevalence were available at the LHA level from the BC Ministry of Health, which contracted Statistics Canada to oversample in BC during the 2008-2009 Canadian Community Health Survey (Statistics Canada 2009; Statistics Canada 2011a). Data from some of the smaller LHAs were combined to ensure statistical validity, resulting in 83 rather than 89 estimates. Each LHA was assigned a binary classification of higher or lower smoking based on whether its smoking prevalence was above or below the median of all 83 estimates.

3.5 Methods

3.5.1 Indoor Radon Vulnerability Modelling and Mapping

Following the methods outlined in Branion-Calles et al (2015) we used a balanced random forest algorithm to classify radon vulnerability based on whether model estimates were above or below the eight threshold values (50, 100, 150, 200, 300, 400, 500, and 600 Bq m^{-3}). Indoor radon concentrations result from a complex combination of environmental and housing characteristics and therefore necessitate a modelling technique that can capture this complexity (Nazaroff 1992). Random forests are used to model complex environmental processes because they are a non-parametric ensemble classifier with a high predictive ability and the flexibility to accommodate mixed variable types, non-linear relationships, and high order interaction effects (Cutler et al. 2007; Prasad et al. 2006). The random forest algorithm works by combining the results of a user specified amount of maximally grown classification trees. Each classification tree is created by randomly selecting a bootstrapped sample of the training data and continually splitting the sample into two subsets based on the value of an independent variable, until all subsets can no longer be split. For each split the algorithm first selects a random subset of all available independent variables and, second, searches all possible binary splits based on the whole range of values within the selected subset of independent variables. The split that is chosen maximizes the class homogeneity within each resulting subset. When results are aggregated over all trees, the variability between trees in the forest reduces over-fitting and susceptibility to outliers in the model (Cutler et al. 2007; Prasad et al. 2006). The balanced approach modifies the random forest algorithm by ensuring that there is equal representation in each bootstrapped sample from which each tree is grown in order to more effectively classify the minority class in an imbalanced dataset (Chen et al. 2004).

Estimates of predictive accuracy can be made without an independent validation dataset by using “out-of-bag” (OOB) data. This refers to the observations that were left out of any given bootstrapped sample (Breiman 2001). In order to obtain an unbiased estimate of predictive performance, the OOB observations for each classification tree are dropped down and assigned a predicted classification. The final prediction for each observation is given based on its majority classification over all trees for which it was OOB. For all observations the OOB prediction can be compared with its observed class to derive an unbiased estimate for the predictive performance of the model through an assessment of the so-called confusion matrix.

Class accuracy, class precision, and a kappa statistic can be all be generated from the OOB confusion matrix to evaluate model performance. Class accuracy refers to the proportion of a given observed class that was correctly classified. Class precision refers to the proportion of the given predicted class that were correctly classified. A kappa statistic quantifies the improvement of the classifier compared with a random classifier, which can be a robust measure to evaluate overall classifier performance for imbalanced datasets (Fatourechi et al. 2008).

An individual balanced random forests model was trained on the subset of 1054 BDAs that had observed vulnerability classes based on the eight selected radon thresholds (Breiman 2001; Chen et al. 2004). To ensure stable results each model combined twenty balanced random forest algorithm runs consisting of 10,000 individual classification trees. The model performance was compared by evaluating class accuracy, class precision, and kappa scores. A vulnerability classification was assigned to the unmeasured BDAs from each model, resulting in eight different maps. Approximately 23% of BDAs in the province had independent variable values not contained within the training dataset, which made them ineligible for prediction. For each map the regional differences in vulnerability were assessed by comparing (1) the geographic areas classified as high and (2) the number of people living within those areas, where regions were defined by census division boundaries (Statistics Canada 2011b).

3.5.2 Comparing Lung Cancer Mortality Trends

The effects of radon thresholds on lung cancer mortality trends was assessed by comparing the annual ratio of lung cancer mortality to all natural mortality in high and low vulnerability regions. Each death was spatially assigned to a radon vulnerability class for each of the eight predictive maps. Additionally, each death was assigned to higher or lower smoking prevalence based on the LHA in which it occurred. By attributing each death with both radon and smoking classifications we were able to compare trends across high and low radon vulnerability conditioned on smoking prevalence. For each radon reference threshold, the annual sum of lung cancer deaths was divided by annual sum of all natural deaths in the high and low vulnerability areas. The values for 1998 through 2013 were plotted, and a trend line was fitted using a LOESS smoother. The same was done to explore potential differences between males and females, as was previously observed in BC (Henderson et al. 2014).

3.6 Results

3.6.1 Indoor Radon Vulnerability

The difference in overall classification accuracy between models improved as the radon threshold decreased (Table 3.1). The Kappa score for each model improved with each reduction in concentration threshold. The greatest gains in performance as measured by Kappa were found in the reductions from 600 to 500 Bq m^{-3}, 300 to 200 Bq m^{-3}, and 150 to 100 Bq m^{-3}, with gains of 0.11, 0.08, and 0.11, respectively. Reductions from 500 to 300 Bq m^{-3}, 200 to 150 Bq m^{-3}, and 100 to 50 Bq m^{-3} resulted in minimal improvement to the Kappa score.

Models for lower radon thresholds were better able to accurately the predict the high vulnerability classification, which lead to the observed gains in the Kappa score. Estimates of the accuracy for high vulnerability classification increased from 0.69 to 0.86. Class precision also increased with each successive reduction in radon threshold, from 0.22 to 0.84. Conversely, estimates for the accuracy of low vulnerability classification decreased from 0.83 to 0.77 and estimates of its class precision decreased from 0.97 to 0.8 (Table 3.1). The gains in class accuracy and class precision for the high vulnerability class with the use of a lower concentration threshold were much greater than the decreases in class accuracy and class precision in the low vulnerability class.

The overall provincial prevalence of high vulnerability areas increased with lower concentration thresholds (Figure 3.3), but some regions were more affected than others. Census divisions in the central and northeast had the largest increases across decreasing radon thresholds, but the highly populated southern coastal areas were generally not affected. The relative ranks of areas at risk were minimally affected by changes in concentration thresholds. For example, census divisions within the Kootenay economic region were at highest risk across all thresholds. The Northeast economic region was most affected, showing a rapid increase in high vulnerability with decreasing threshold values (Figure 3.4).

The total number of residents living in high vulnerability areas increased as the concentration threshold decreased, but the rate of increase varied regionally. Census divisions within the Thompson Okanagan, Cariboo, and Kootenay economic regions consistently had higher numbers of residents living in high vulnerability areas compared with the rest of the province, regardless of the threshold. Although there were large increases in the geographic area classified as high vulnerability in the Northeast and Nechako economic regions, the number of people living in those areas remained low due to their sparse populations (Figure 3.4).

3.6.2 Lung Cancer Mortality Trends

The trends for the entire province showed that areas with high radon vulnerability consistently had higher proportions of lung cancer mortality across all radon thresholds. Further, there was little change in the distance between the high and low vulnerability lines with decreasing radon thresholds. When plots were stratified by higher and lower smoking prevalence there was no clear separation between the lung cancer trends in areas with higher smoking. However, in areas with lower smoking prevalence, the high radon vulnerability areas had consistently higher proportions of lung cancer mortality, though the separation between lines decreased as the radon threshold decreased. The trends in lung cancer mortality for low radon vulnerability areas with lower smoking were flat and stable at ~7.5% for all thresholds while the trends in higher smoking areas were curved and increasing (Figure 3.5).

When lung cancer mortality trends in high and low vulnerability areas were stratified by sex, high vulnerability areas were consistently associated with higher proportions of lung cancer mortality across radon thresholds for both males and females. For each threshold, the trend lines for males in both high and low vulnerability areas showed a slight decrease through time while they were increasing through time for females. The ratios for females in high vulnerability appeared unstable for thresholds greater than 300 Bq m^{-3} (Figure 3.6).

3.7 Discussion

Different regions, countries, and organizations recommend different radon concentration thresholds that essentially classify the associated risk of lung cancer as being acceptable or unacceptable. In reality, however, radon is a non-threshold carcinogen and any level of exposure carries some risk (Darby et al. 2005). Established guideline concentration values reflect a balance between the health evidence, what is practically achievable, and other political and public health priorities. To date there has been little systematic evaluation of how decisions about threshold values affect variables such as the accuracy with which risk can be classified, the extent of geographic areas classified as high risk, the size of the populations classified as high risk, and the observed relationships between risk areas and lung cancer mortality trends. Here we have addressed this gap by exploring the impacts of thresholds ranging from 50 – 600 Bq m^{-3}

in one Canadian province with previously demonstrated spatial variability in radon risk (Branion-Calles et al. 2015; Henderson et al. 2014; Rauch & Henderson 2013).

We found that the accuracy of risk classification was improved as the threshold decreased, likely due to increasing balance of the training data. Though the balanced random forest algorithm is more effective at classifying imbalanced datasets than an unmodified random forest, it is still designed to minimize the overall error. This appeared to be more effective when high and low vulnerability were delineated using a lower threshold, resulting in a more balanced dataset (Chen et al. 2004). Due to the potential for misclassification of individual BDAs, each threshold map should be interpreted at regional scale rather than at the individual mapping unit.

Unsurprisingly, the geographic extent of areas classified as high risk became larger as the thresholds decreased. However, much of BC is sparsely populated, so it was more important to consider changes in the populations classified as high and lower vulnerability as the thresholds changed. in the number of people living in high vulnerability areas increased from approximately 326,000 to 824,800 when the threshold was reduced from the current guideline value of 200 Bq m^{-3} to to the minimum value of 50 Bq m^{-3}. Given that high vulnerability areas were associated with higher prevalence of lung cancer mortality, the increase in exposed population indicates that adoption of a higher threshold value has the potential to mask some risk.

Regardless of threshold employed, the majority of the provincial population lived in areas classified as low vulnerability. At the current guideline value of 200 Bq m^{-3} only 7% of BC residents were estimated to live in areas of high vulnerability. However, there is direct evidence that indoor radon concentrations contribute to lung cancer mortality in the general population at concentrations less than the Health Canada guideline (Darby et al. 2005). A study in the UK estimated that approximately 96% of radon-related lung cancer deaths resulted from exposures to indoor concentrations less than 200 Bq m^{-3}, due to the large number of people exposed to these lower risk concentrations (Gray et al. 2009). Given that the majority of the BC population is exposed to concentrations lower than 200 Bq m^{-3} it is likely the majority radon-related lung cancer deaths result from exposures less than 200 Bq m^{-3}.

The crude lung cancer mortality ratio within areas classified as high vulnerability was higher than within areas classified as low vulnerability for every concentration threshold through time. Smoking rates are the primary predictor of population-level lung cancer risk, which results in geographic variations in lung cancer mortality trends being dominantly associated with geographic variations in smoking prevalence (Youlden et al. 2008; Alberg & Nonemaker 2012; Jemal et al. 2010). In BC areas with higher smoking prevalence, we observed no differences in lung cancer mortality trends between high and low vulnerability areas. In areas with lower smoking prevalence, however, differences between radon vulnerability areas were clear across all threshold values. Radon and smoking have a synergistic relationship at the individual level (Saccomanno et al. 1988; Office of Radiation and Indoor Air 2003), but radon vulnerability appeared to have little effect on population mortality trends in high smoking areas. Furthermore, the difference in trends between males and females suggests that sex may have a modifying effect on lung cancer mortality ratios in the province (Henderson et al. 2014). The difference in trends between high and low vulnerability in areas with a lower smoking prevalence suggests that the methods first developed in Branion-Calles et al (2015) were able to delineate areas of higher and lower radon risk.

In areas with lower smoking prevalence areas classified as high vulnerability had a distinctly higher rates of lung cancer mortality than areas classified as low vulnerability, even when a threshold as low as 50 Bq m^{-3} was used. Given that overall rates of cigarette smoking are declining in BC (Health Canada 2013a), we hypothesize that the trends in areas with higher smoking will approach those in areas with lower smoking over time. However, the smoking prevalence data were only available at the LHA level and, as a result, we could not account for geographic variability in smoking within LHAs. Overall, the elevated trends in lung cancer mortality associated with the high vulnerability areas indicate that radon exposure is an important risk factor in BC.

The Health Canada radon guideline is a threshold value that provides a frame of reference for making informed decisions about radon testing and remediation, but Canadian residential radon values are not regulated (Health Canada 2006; Health Canada 2013b). In the absence of binding federal policy, provincial governments have the authority to independently enact radon protection legislation through changes to provincial building codes (Dunn & Cooper 2014). Although BC has adopted radon mitigation measures for newly constructed buildings in its provincial code, there is no legal requirement for new buildings to test below a specific concentration threshold (Dunn & Cooper 2014). Based on the results of our study and the principle that no radon concentration is safe, there is evidence to support the recommendation of a concentration threshold as low as 50 Bq m^{-3} for BC. Though further research is needed to quantify the absolute number of lung cancer deaths related to indoor radon across the province, a lower threshold value may have the potential to reduce burden of disease attributable to radon, especially if it was legally enforced for new buildings. While such measures would not affect the existing building stock, they would be an important step towards protecting the BC population from radon exposure in future.

3.8 Conclusions

We examined how different radon concentration thresholds were associated with classification accuracy, estimated areas and populations at risk, and lung cancer mortality trends in BC. Lowering the threshold from its current guideline value of 200 Bq m^{-3} to 50 Bq m^{-3} resulted in better classification accuracy, a 2.5-fold increase in the relatively small population at risk, and persistent separation in lung cancer mortality trends between areas of high and low vulnerability. We suggest that it would be appropriate for BC to consider mandating a 50 Bq m^{-3} threshold value to maximize the reduction of radon-related lung cancer in the province.

Acknowledgements

We would like to thank the BC Lung Association, Northern Health Authority, Donna Schmidt Foundation and Peter Chataway for sharing their data. This work has been supported by the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada.

References

Agriculture and Agri-Food Canada, 2013. Soil Landscapes of Canada (SLC). *Government of Canada*. Available at: http://sis.agr.gc.ca/cansis/nsdb/slc/index.html [Accessed May 15, 2014].

Al-Ahmady, K.K. & Hintenlang, D.E., 1994. Assessment of temperature-driven pressure differences with regard to radon entry and indoor radon concentration. In *AARST*. Atlantic City: The American Association of Radon Scientists and Technologists.

Alberg, A.J. & Nonemaker, J., 2012. Who is at high risk for lung cancer? Population-level and individual-level perspectives. *Seminars in Respiratory and Critical Care Medicine*, 29(3), pp.223–232.

Appleton, J.D., 2007. Radon: sources, health risks, and hazard mapping. *Ambio*, 36(1), pp.85–89. Available at: http://www.jstor.org/stable/4315791.

Arnold, B.W., 2006. Radon Transport. In C. K. Ho & S. W. Webb, eds. *Gas Transport in Porous Media*. Springer Netherlands, pp. 333–338.

BC Ministry of Energy and Mines, 2015. MINFILE Mineral Inventory. Available at: http://www.empr.gov.bc.ca/mining/geoscience/minfile/Pages/default.aspx# [Accessed January 1, 2015].

Branion-Calles, M.C., Nelson, T.A. & Henderson, S.B., 2015. A geospatial approach to the prediction of indoor radon vulnerability in British Columbia, Canada. *Journal of Exposure Science and Environmental Epidemiology*, 00, pp.1–12.

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp.5–32. Available at: http://link.springer.com/article/10.1023/A:1010933404324 [Accessed September 25, 2013].

Chen, C., Liaw, A. & Breiman, L., 2004. *Using random forest to learn imbalanced data*, University of California, Berkeley. Available at: http://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf [Accessed August 7, 2014].

Chen, J., Moir, D. & Whyte, J., 2012. Canadian population risk of radon induced lung cancer: a re-assessment based on the recent cross-Canada radon survey. *Radiation protection dosimetry*, 152(1-3), pp.9–13.

Cui, Y. et al., 2013. British Columbia Digital Geology: BCGS Open File 2013-04. *BC Geological Survey*. Available at: http://www.empr.gov.bc.ca/MINING/GEOSCIENCE/PUBLICATIONSCATALOGUE/OPENFILES/2013/Pages/2013-4.aspx [Accessed April 15, 2014].

Cutler, D.R. et al., 2007. Random forests for classification in ecology. *Ecology*, 88(11), pp.2783–2792. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18051647.

Darby, S. et al., 2005. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ*, 330(7485), pp.223–226. Available at: http://www.bmj.com/cgi/doi/10.1136/bmj.38308.477650.63 [Accessed September 25, 2013].

Dunn, B. & Cooper, K., 2014. *Radon in Indoor Air : A Review of Policy and Law in Canada*, Toronto.

Fatourechi, M. et al., 2008. Comparison of Evaluation Metrics in Classification Applications with Imbalanced Datasets. In *Machine Learning and Applications, 2008. ICMLA '08. Seventh International Conference on*. San Diego: IEEE, pp. 777–782. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4725065 [Accessed August 16, 2014].

GeoBC, 2014. Freshwater Atlas. Available at: http://geobc.gov.bc.ca/base-mapping/atlas/fwa/ [Accessed April 15, 2014].

Gray, A. et al., 2009. Lung cancer deaths from indoor radon and the cost effectiveness and potential of policies to reduce them. *BMJ (Clinical research ed.),* 338(7688), p.a3110.

Health Canada, 2013a. Current Smoking Prevalence by Age, Canada, 1985-2012. Available at: http://www.hc-sc.gc.ca/hc-ps/tobac-tabac/research-recherche/stat/ctums-esutc_2012-eng.php [Accessed April 1, 2015].

Health Canada, 2006. *Report of the Radon Working Group on a New Radon Guideline for Canada*, Available at: http://carst.ca/Resources/Documents/2006 Report of the Radon Working Group in Canada.pdf.

Health Canada, 2013b. Responses to Peer Reviewers' Comments on the Proposed Revision to the Radon Guideline. Available at: http://www.hc-sc.gc.ca/ewh-semt/radiation/radon/peer-pair-comment-radon-eng.php [Accessed April 1, 2015].

Henderson, S.B. et al., 2014. Differences in lung cancer mortality trends from 1986-2012 by radon risk areas in British Columbia, Canada. *Health Physics*, 106(5), pp.608–613. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24670910 [Accessed October 16, 2014].

Henderson, S.B., Kosatsky, T. & Barn, P., 2012. How to Ensure That National Radon Survey Results Are Useful for Public Health Practice. *Can J Public Health*, 103(3), pp.231–234.

Jemal, A. et al., 2010. Global patterns of cancer incidence and mortality rates and trends. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 19(8), pp.1893–1907.

Long, S. & Fenton, D., 2011. An overview of Ireland's National Radon Policy. *Radiation protection dosimetry*, 145(2-3), pp.96–100.

Nazaroff, W.W., 1992. Radon transport from soil to air. *Reviews of Geophysics*, 30(2), pp.137–160. Available at: http://dx.doi.org/10.1029/92RG00055 [Accessed November 14, 2013].

Prasad, A.M., Iverson, L.R. & Liaw, A., 2006. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*, 9(2), pp.181–199.

Rauch, S.A. & Henderson, S.B., 2013. A comparison of two methods for ecologic classification of radon exposure in British Columbia: residential observations and the radon potential map of Canada. *Canadian journal of public health*, 104(3), pp.e240–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23823889.

Saccomanno, G. et al., 1988. Relationship of radioactive radon daughters and cigarette smoking in the genesis of lung cancer in uranium miners. *Cancer*, 62(7), pp.1402–1408.

Shweikani, R., Giaddui, T.G. & Durrani, S. a., 1995. The effect of soil parameters on the radon concentration values in the environment. *Radiation Measurements*, 25(1-4), pp.581–584. Available at: http://linkinghub.elsevier.com/retrieve/pii/135044879500188K [Accessed November 13, 2014].

Statistics Canada, 2009. *Canadian Community Health Survey (CCHS): annual component user guide for the 2008 microdata files*, Ottawa, Ontario.

Statistics Canada, 2011a. *Canadian Community Health Survey (CCHS): supplement to the user guide for the British Columbia sample buy-in*, Ottawa, Ontario.

Statistics Canada, 2011b. Census Dictionary 2011. *Statistics Canada Catalogue no. 98-301-X2011001*. Available at: http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/98-301-X2011001-eng.pdf.

Statistics Canada, 2013. National Household Survey (NHS) Profile, 2011. [2013]. *Statistics Canada*. Available at: http://hdl.handle.net/10573/42928.

Synnott, H. & Fenton, D., 2005. *An Evaluation of Radon Reference Levels and Radon Protocols in European Countries: A report of the ERRICCA 2 European project*, Available at: https://www.epa.ie/pubs/reports/radiation/RPII_ERRICA_Measure_Report_05.pdf.

US Environmental Protection Agency, 2003. *EPA Assessment of Risks from Radon in Homes*, Available at: http://www.epa.gov/radiation/docs/assessment/402-r-03-003.pdf.

Wang, T. et al., 2012. ClimateWNA—High-Resolution Spatial Climate Data for Western North America. *Journal of Applied Meteorology and Climatology*, 51(1), pp.16–29. Available at: http://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-11-043.1 [Accessed October 27, 2014].

World Health Organization, 2007. *International Radon Project Survey on Radon Guidelines, Programmes and Acvitivites*, Geneva. Available at: http://www.who.int/ionizing_radiation/env/radon/IRP_Survey_on_Radon.pdf.

World Health Organization, 2009. *WHO Handbook on Indoor Radon: A Public Health Perspective* H. Zeeb & F. Shannoun, eds., Geneva.

Youlden, D.R., Cramb, S.M. & Baade, P.D., 2008. The International Epidemiology of Lung Cancer: geographical distribution and secular trends. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, 3(8), pp.819–831.

Table 3.1 - The classification metrics for each balanced random forest algorithm. Accuracy is defined as the proportion of an observed class that was correctly classified. Precision is defined as the proportion of a predicted class that was correctly classified. Kappa can be interpreted as the percent improvement in overall accuracy of a classifier compared with the expected overall accuracy of a random classifier. Values in bold indicate the highest value between threshold models.

	Threshold in Bq m^{-3}	**Lower-than-threshold Accuracy**	**Lower-than-threshold Precision**	**Higher-than-threshold Accuracy**	**Higher-than-threshold Precision**	**Kappa**	**Kappa Gain**
a)	600	0.81	**0.97**	0.69	0.22	0.25	0
b)	500	**0.83**	0.96	0.72	0.32	0.36	**0.11**
c)	400	**0.83**	0.96	0.74	0.37	0.39	0.03
d)	300	0.8	0.94	0.73	0.42	0.41	0.02
e)	200	0.8	0.91	0.76	0.55	0.49	0.08
f)	150	0.77	0.88	0.76	0.6	0.5	0.01
g)	100	0.79	0.86	0.83	0.75	0.61	**0.11**
h)	50	0.77	0.8	**0.86**	**0.84**	**0.63**	0.02

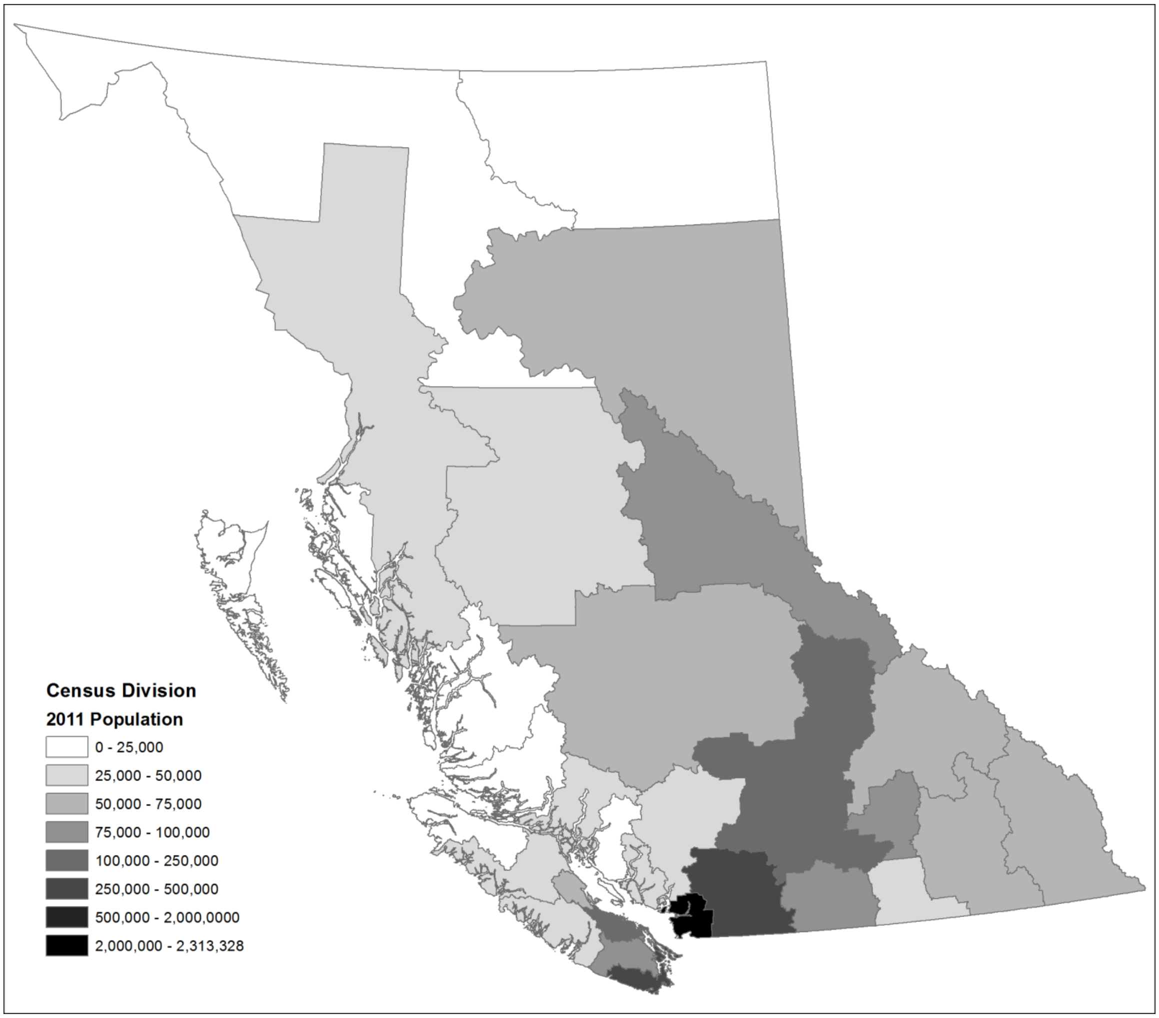

Figure 3.1 - The study area of British Columbia, Canada. The spatial distribution of the provincial population by census division boundaries is shown.

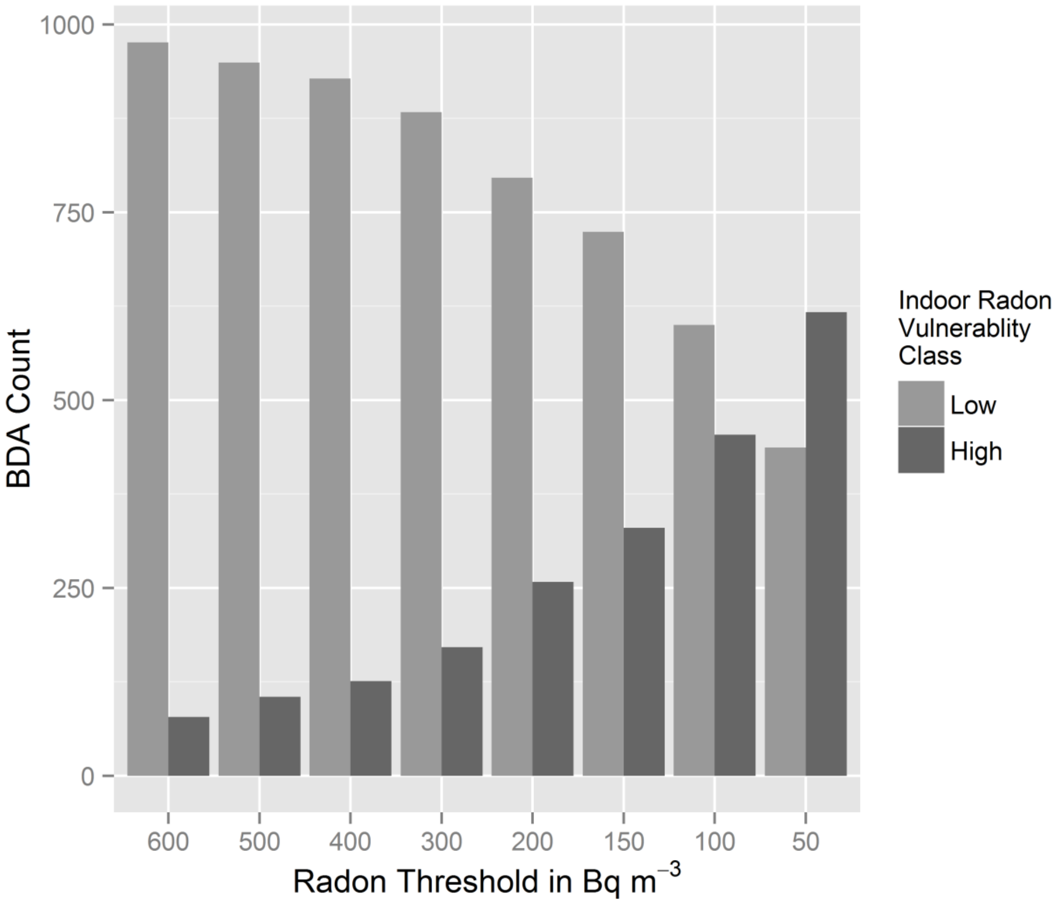

Figure 3.2 - The class distribution of bedrock dissemination areas (BDAs) in the training dataset using each threshold value.

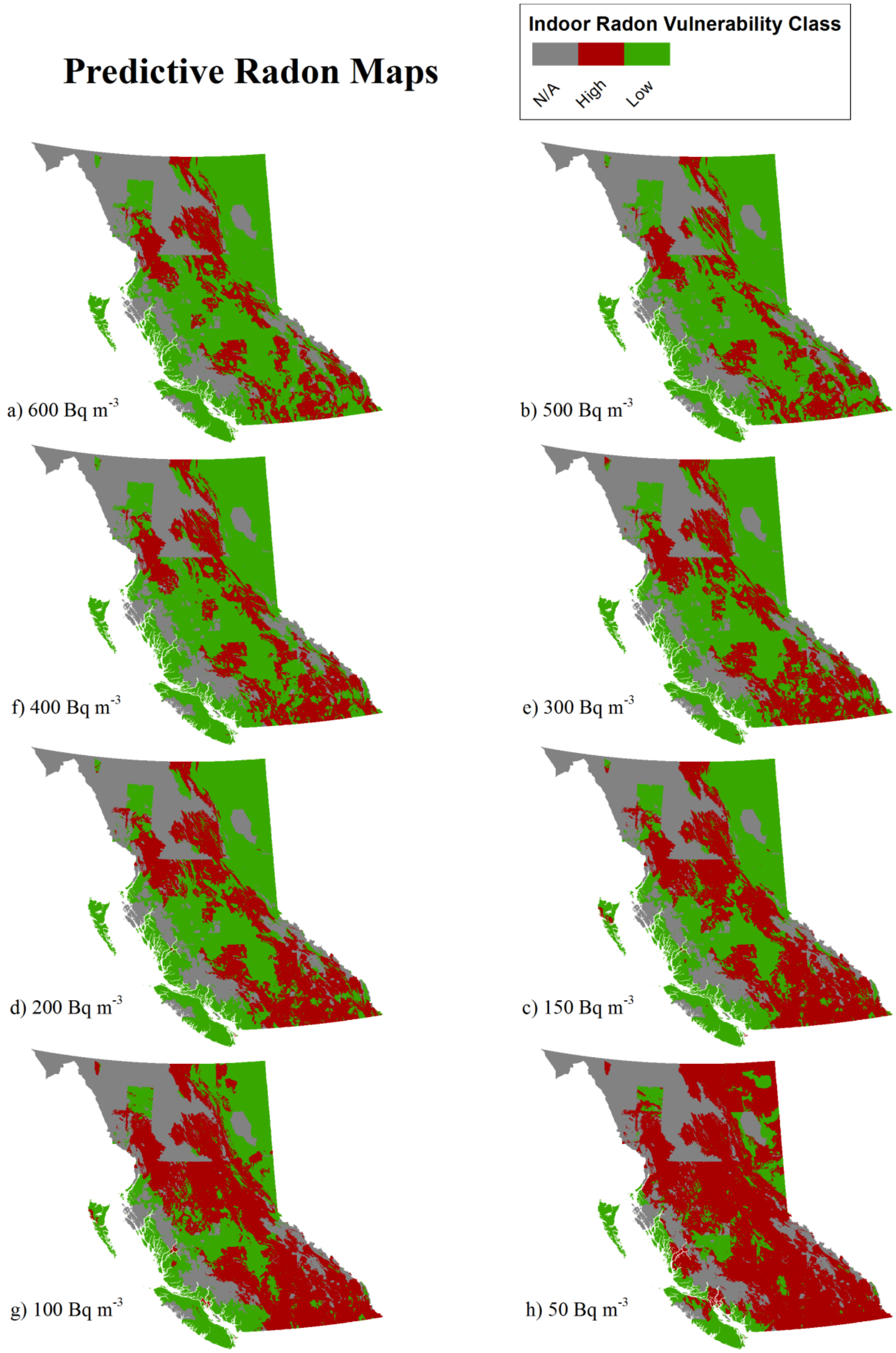

Figure 3.3 - Estimated vulnerability maps for each of the eight radon threshold. Red areas indicate high vulnerability, green areas indicate low vulnerability, and grey areas indicate regions without adequate data for modelling.

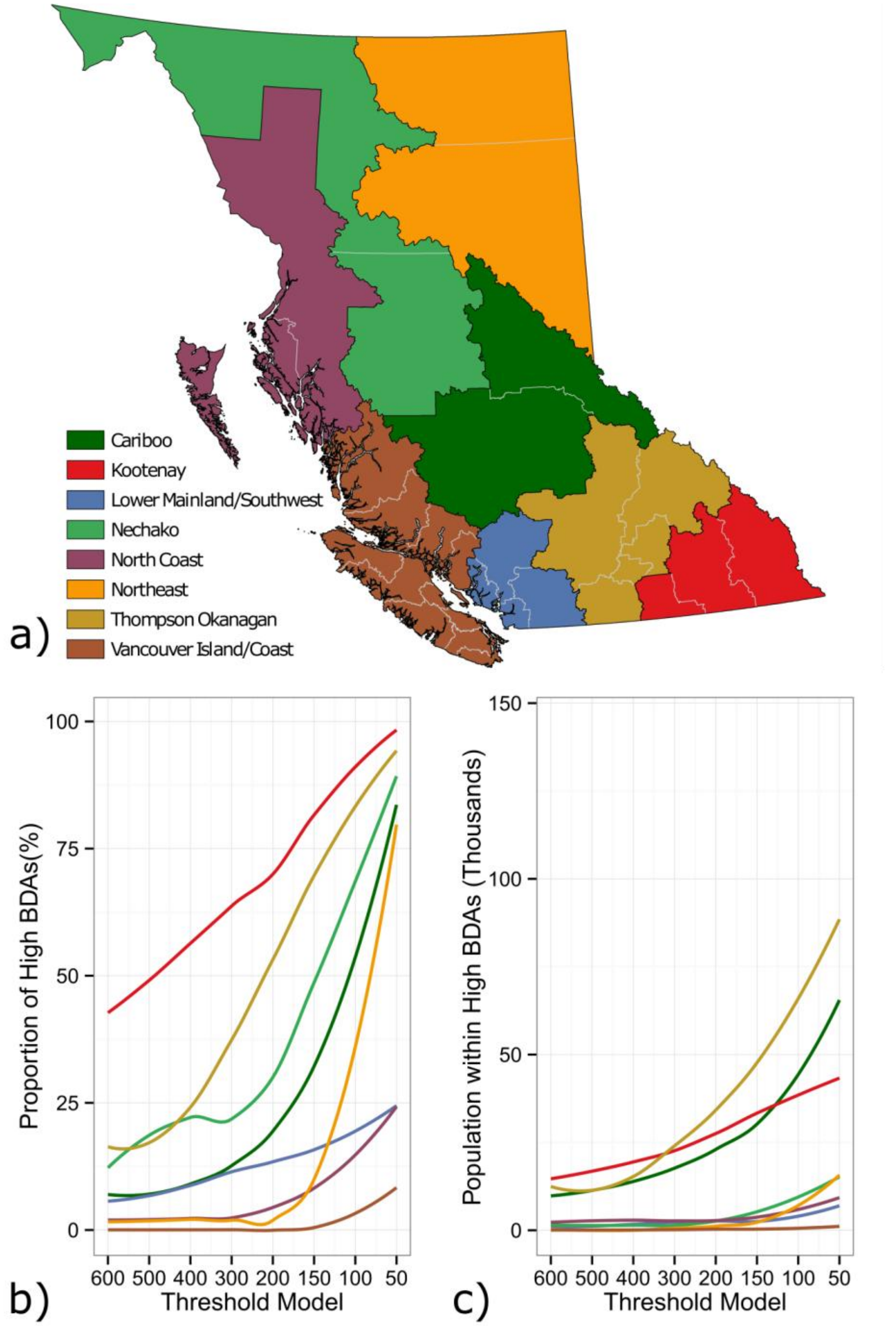

Figure 3.4 - Changes in regional vulnerability classification based on changes in threshold values plotted by the proportion of high BDAs by census division (b) and the estimated population living within high BDAs (c). The colours all correspond to the legend in (a). Census divisions are demarcated by grey lines in (a), and they aggregate up to the coloured economic regions (a). Trends in (b) and (c) were fitted using a locally-weighted LOESS smoother.

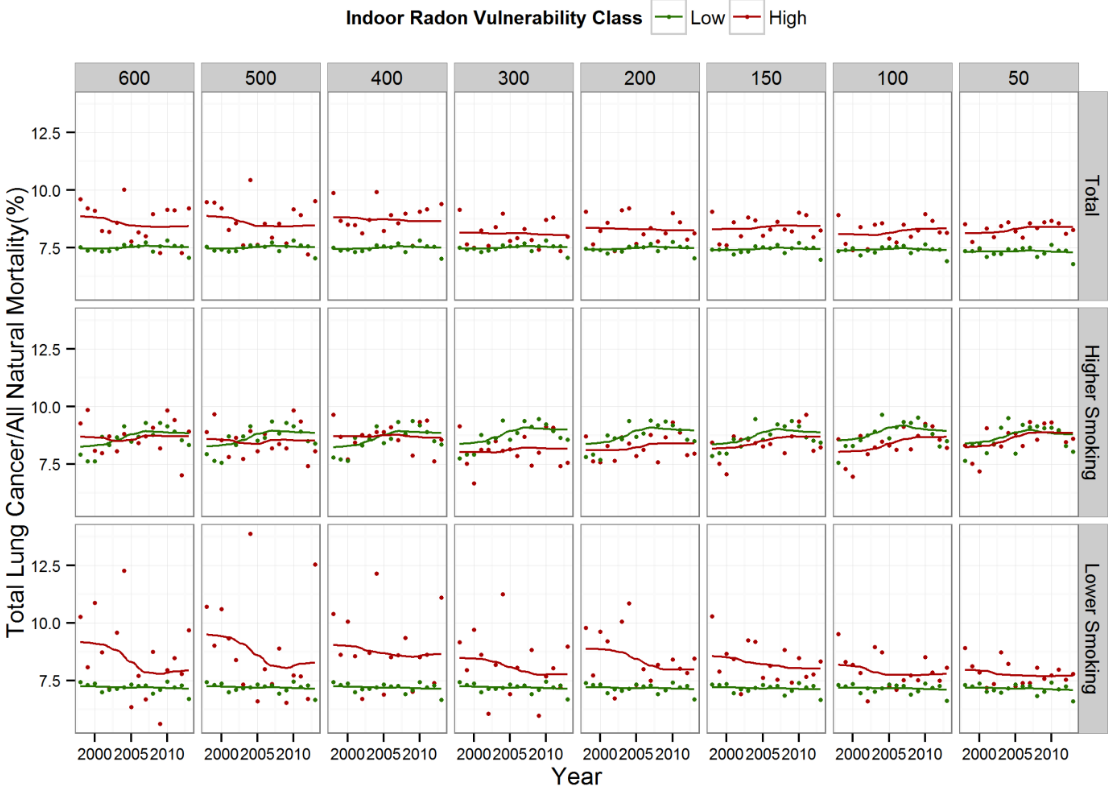

Figure 3.5 - The annual ratio of lung cancer mortality to all natural mortality (the crude lung cancer mortality ratio) within high and low vulnerability areas plotted from 1998-2013 for each predictive map based on eight threshold values. The columns show the threshold values in Bq m^{-3}, which were used to delineate low and high vulnerability. The rows show the total trends, and the trends when stratified by higher smoking LHAs and lower smoking LHAs. The lung cancer mortality trends were fitted with a locally-weighted LOESS smoother.

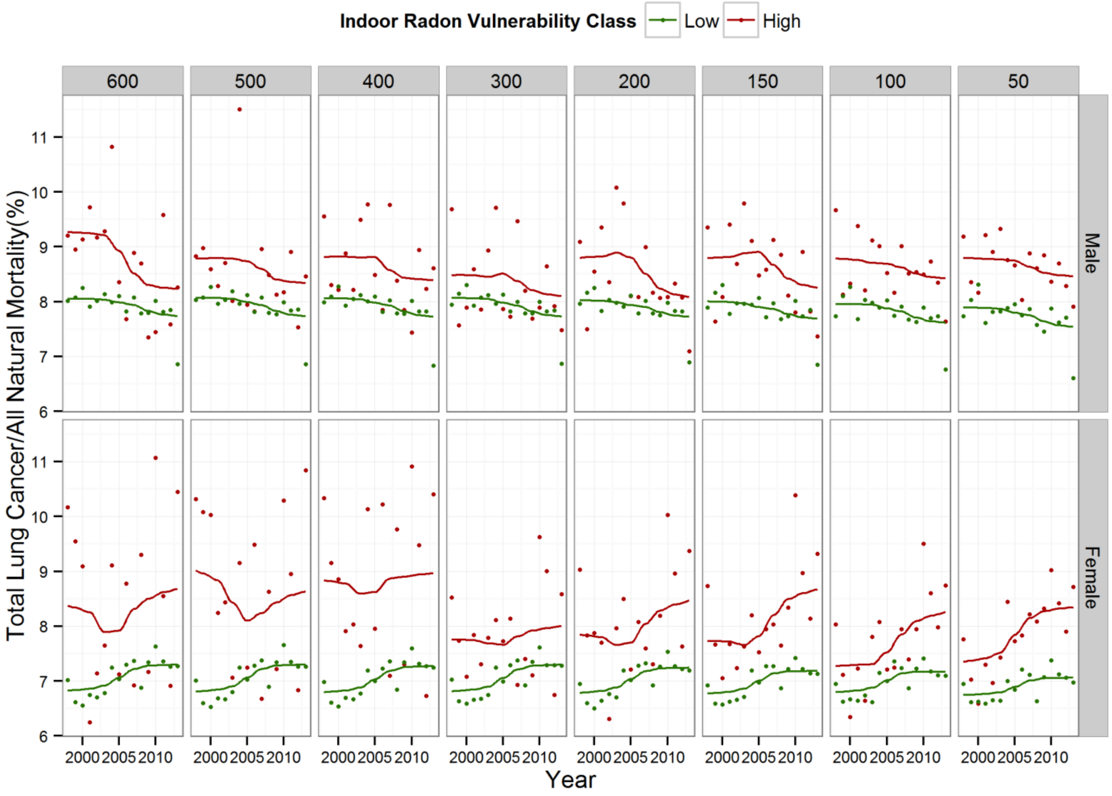

Figure 3.6 - The annual ratio of lung cancer mortality to all natural mortality (the crude lung cancer mortality ratio) within high and low vulnerability areas plotted from 1998-2013 for each predictive map based on eight threshold values. The columns show the threshold values in Bq m^{-3}, which were used to delineate low and high vulnerability. The rows show the trends stratified by sex.

4.0 CONCLUSIONS

4.1 Discussion and Conclusions

Indoor radon is recognized as an important environmental hazard and a major public health concern worldwide. It is second only to smoking in contributing to lung cancer incidence globally (World Health Organization 2009). In Canada, radon is estimated to be a factor in over 3,000 lung cancer deaths annually (Chen et al. 2012). British Columbia has many radon prone communities with both large and small populations (Henderson et al. 2012), and indoor radon is a major contributor to lung cancer mortality in the province (Henderson et al. 2014). Population level mitigation of the health effects of exposure to indoor radon require an understanding of its spatial distribution, that is, which regions are at highest risk for homes with higher concentration and subsequently greater rates of lung cancer incidence and mortality. Mapping the regional variations in susceptibility to high indoor radon concentration is a vital tool for population-level reduction of exposure to indoor radon, as it can help target radon reduction resources to areas in greatest need, and guide the formation of policies (Miles & Appleton 2005; Long & Fenton 2011). Where data are not uniformly distributed in a study area, spatially continuous maps of indoor radon risk are produced by either aggregating observed indoor radon concentrations by large spatial units (Dubois 2005; Chen 2009; Henderson et al. 2014) or by inferring indoor radon risk from maps produced using proxy data, such as geochemical, rock or soil permeability, or soil-gas radon concentration data (Chen 2009; Appleton & Ball 2002; Ielsch et al. 2010; Kemski et al. 2008). Using a few large spatial units to describe radon risk in a jurisdiction can hide the true variability in indoor radon concentrations and oversimplify the resulting maps, while maps of radon risk based on proxy data can be inconsistent with observed indoor radon data (Rauch & Henderson 2013).

The development of our predictive indoor radon mapping method was based on the need to be able to produce accurate maps of indoor radon risk based directly on indoor radon concentrations at spatial units much smaller in size than in previous work. Moreover, since there is no safe threshold concentration of indoor radon, the delineation of high risk areas is subject to arbitrary decisions of concentrations assumed to be dangerous, the choice of which can serve to under or overestimate radon risk. Radon thresholds that distinguish dangerous from non-dangerous concentrations are provided by different countries, organizations and regions, but each is subject to uncertainty due to the fact there is no safe concentration of radon, and risk still exists at exposures below a given guideline. Our mapped scenarios of radon *vulnerability* using a range of different threshold concentrations from 50 to 600 Bq m^{-3}, provide vital information in understanding the geographic extent of the indoor radon problem in BC.

In Chapter 2 we develop a data driven method in order to predict ordinal classes of radon vulnerability in areas without any indoor radon observations, map the results and identify areas most at risk. We define low, medium and high vulnerability based on whether the observed 95th percentile concentration within the spatial unit of prediction were below, between, or above the two radon concentration thresholds recommended by Health Canada for individual homeowners of 200 and 600 Bq m^{-3}. We then link classes of radon vulnerability to select environmental and neighbourhood housing predictors to quantify relationships between predictor variables and classes of radon vulnerability. We use a balanced variant of the statistical classifier, random forests, to model radon vulnerability in the province. We elucidate the relationship between predictor variables and the probability of low, medium or high vulnerability through partial dependence plots. Our results reveal that a higher probability of moderate or high indoor radon vulnerability is associated with areas having colder and drier winters, in closer proximity to major rivers, and whose dominant soil parent material mode of deposition are either fluvioglacial or colluvial. Census divisions at greatest risk (largest proportion of moderate and high indoor radon vulnerability) include the Kootenays and Columbia-Shuswap.

In Chapter 3, using the balanced random forest algorithm and data outlined in Branion-Calles et al (2015), we explore the impact of using different radon concentration thresholds to define low and high vulnerability on several factors. These factors include: the accuracy of the balanced random forest classification algorithm, the change in geographic extent of high and low vulnerability in the resulting maps, the change in populations classified as higher and lower vulnerability, and the temporal trends in lung cancer mortality associated with higher and lower vulnerability. Eight threshold values between 50 and 600 Bq m^{-3} model indoor radon vulnerability. We compare high vulnerability areas, their estimated populations and the differences in lung cancer mortality trends stratified by sex and smoking prevalence. Our results indicate that the accuracy of the classification improves with lower thresholds as a result of a more balanced dataset, and that, unsurprisingly, the geographic extent of high vulnerability areas, as well as the number of residents within them, increases with a lower threshold. Census divisions in the Northeast economic regions show the greatest increase in number of high BDAs while census divisions within the Thompson Okanagan, Cariboo and Kootenay economic regions consistently contain greater number of residents within high vulnerability zones. The majority of the provincial population is estimated to live in low vulnerability areas.

By demonstrating that the crude lung cancer mortality ratio is consistently higher in areas of high radon vulnerability for each threshold concentration tested, we provide evidence for the efficacy of the mapping methods outlined in Chapter 2 (Branion-Calles et al. 2015). Moreover, by comparing the results using a range of different threshold concentrations, we establish that the potential to exceed a concentration as low as 50 Bq m^{-3} is associated with elevated temporal trends in lung cancer mortality in areas of lower smoking prevalence. We therefore recommend the adoption of a lower threshold concentration in BC than the currently recommended threshold by Health Canada.

4.2 Research Contributions

The papers presented in this thesis contribute to the literature surrounding radon in BC identifying it as an important public health threat (Henderson et al. 2014; Rauch & Henderson 2013; Henderson et al. 2012). This thesis demonstrates the utility of GIS approaches for surveillance of indoor radon in the province. By developing a spatial model of indoor radon risk we were able to not only predict areas with a greater susceptibility to higher indoor radon concentrations, but also identify predictor variables associated with higher risk areas, allowing for an assessment of possible environmental contributors to regional radon risk. Furthermore, the use of GIS approaches allowed for the generation of multiple scenarios of risk, and the overlay of population and mortality datasets to assess health outcomes. The ability to combine multiple related datasets in order to assess different characterizations of indoor radon's spatial distribution, their co-location to human populations, and associated health outcomes, is an invaluable tool to inform policy and promote interventions to reduce population level exposure to radon.

A key contribution of this thesis is a framework for producing maps of indoor radon risk at fine spatial resolutions based on existing geospatial datasets. This novel method contributes to the radon mapping literature and the final product can be used for risk communication, geographic targeting of radon awareness and monitoring campaigns, and could inform the development of radon policy in the province. The method presented in this thesis allows for the estimation of indoor radon risk based directly on observed indoor radon data for much smaller mapping units than would be possible using only the observed data. The method could theoretically be replicated in jurisdictions throughout the world given that similar existing geospatial datasets are available.

Furthermore, to my knowledge, little work has been done to assess the effect of the use of different threshold values on geographic characterizations of radon risk, its subsequent impact on population within high risk areas as well as the lung cancer mortality trends associated with high risk areas. While a variety of indoor radon risk maps are available throughout the world, they generally present only one scenario of geographic risk based on one set of thresholds (Dubois 2005; Sainz-Fernandez et al. 2014; Henderson et al. 2014). Radon presents some degree of risk at all concentrations, and those areas characterized as lower risk via an arbitrary threshold concentration could serve to create a false sense of security for residents. By assessing a range of possible characterizations of risk we gain a more comprehensive understanding of the spatial distribution of radon risk within BC. By assessing the lung cancer mortality trends associated with each scenario of risk, we can provide greater confidence in the predictive results when informing radon policy and targeting of radon reduction resources.

4.3 Research Limitations

It is well known that individual concentrations of residential radon are highly variable even between adjacent homes. For example, regression models that draw upon large indoor radon datasets with systematically collected and detailed explanatory variables, will often explain relatively little variance (Hunter et al. 2009; Hauri et al. 2012; Andersen et al. 2007). As a result of the complexity of factors contributing to indoor radon concentrations, radon risk maps cannot be used to infer or predict individual concentrations.

The method presented in this thesis for predictive radon mapping was unable to provide a spatially continuous estimate of indoor radon risk in the province, as there a number of large, sparsely populated areas where no predictions could be made. In order to predict a radon vulnerability class for a given mapping unit, the random forest algorithm requires that the attribution for that mapping unit contain values also present in the training data. Any mapping unit which has levels of an attribute not observed in the training data cannot be predicted for. Therefore, the production of a spatially continuous map of indoor radon risk using this method also requires that the spatial distribution of the training data cover the range of potential levels in attribution. As a result, the suggestion that the predictive mapping method would be improved with a smaller study area and more detailed attribution must also consider that observed indoor radon concentrations must overlap these areas and cover the range of predictor variable levels. Therefore with more detailed categorical attribution, the more extensive the spatial distribution of the sampled data has to be.

4.4 Research Opportunities

Though our balanced random forest approach was successful, there are other sampling methods and modeling techniques that can deal with imbalanced datasets and it would be interesting to see how they compare. The balanced random forest technique we employed was able to intrinsically down-sample the majority class by stratifying each bootstrapped sample used to grow an individual classification tree. Although we employed a down-sampling method in this research, there are a number of up-sampling methods such as SMOTE (Chawla et al. 2002), or ADASYN (He et al. 2008) that are available to potentially improve classification of imbalanced datasets. Furthermore, boosting, a popular algorithm for statistical classification that combines multiple weak learners into a stronger learner (Hastie et al. 2009), is an alternative modeling technique. Future research opportunities could lie in combining different sampling and modeling techniques using the indoor radon vulnerability database and comparing their results in order to select the optimal radon vulnerability map in terms of classification accuracy and precision.

The results of our method could be improved and made to be spatially continuous by employing a “multi-tier” approach as described by Chen (2009), to estimate a radon risk class for those mapping units for which a radon vulnerability class could not be assigned. The multi-tier approach is based on the creation of a scoring system that rates the radon potential of a geographic area based on the data available in that specific region (Appleton & Ball 2002; Chen 2009). Where predictions cannot be made using our predictive radon risk mapping methods, estimates of risk could be made using available related ancillary datasets individually or in combination with one another, such as soil geochemical data (Lett et al. 2008), airborne radiometric data (Natural Resources Canada 2015), or estimates of radon potential from the Radon Potential Map of Canada (Radon Environmental Management Corp. 2011).

There are also opportunities to employ more sophisticated methods for determining the difference in lung cancer mortality trends between classes of radon vulnerability than the time-series approach taken in this thesis. The comparison of the odds of dying of lung cancer between different estimates of ecological radon exposure could be potentially explored through a case-only study design (Kosatsky et al. 2012). This design could not only be applied to each threshold map to determine the optimal threshold for vulnerability mapping in the province, it could also be used to compare the odds of dying of lung cancer for other ecological classifications of radon risk such as those defined by the Radon Potential Map of Canada (Radon Environmental Management Corp. 2011) or Radon Risk Areas of BC (Rauch & Henderson 2013; Henderson et al. 2014).

References

Andersen, C.E. et al., 2007. Prediction of 222Rn in Danish dwellings using geology and house construction information from central databases. *Radiation protection dosimetry*, 123(1), pp.83–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16868014 [Accessed October 23, 2013].

Appleton, J.D. & Ball, T.., 2002. Geological radon potential mapping. In P. T. Bobrowsky, ed. *Geoenvironmental Mapping: Methods, Theory and Practice*. Exton, PA: A.A. Balkema Publishers, pp. 577–613.

Branion-Calles, M.C., Nelson, T.A. & Henderson, S.B., 2015. A geospatial approach to the prediction of indoor radon vulnerability in British Columbia, Canada. *Journal of Exposure Science and Environmental Epidemiology*, 00, pp.1–12.

Chawla, N. V. et al., 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, pp.321–357.

Chen, J., 2009. A preliminary design of a radon potential map for Canada: a multi-tier approach. *Environmental Earth Sciences*, 59(4), pp.775–782. Available at: http://link.springer.com/10.1007/s12665-009-0073-x [Accessed September 25, 2013].

Chen, J., Moir, D. & Whyte, J., 2012. Canadian population risk of radon induced lung cancer: a re-assessment based on the recent cross-Canada radon survey. *Radiation protection dosimetry*, 152(1-3), pp.9–13.

Dubois, G., 2005. *An Overview of Radon Surveys in Europe*,

Hastie, T., Tibshirani, R. & Jerome, F., 2009. *The elements of statistical learning: data mining, inference and prediction* Second., New York: Springer. Available at: http://www.springerlink.com/index/D7X7KX6772HQ2135.pdf [Accessed August 15, 2014].

Hauri, D.D. et al., 2012. A prediction model for assessing residential radon concentration in Switzerland. *Journal of environmental radioactivity*, 112(0), pp.83–89. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22683900 [Accessed September 25, 2013].

He, H. et al., 2008. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *Proceedings of the International Joint Conference on Neural Networks*. pp. 1322–1328.

Henderson, S.B. et al., 2014. Differences in lung cancer mortality trends from 1986-2012 by radon risk areas in British Columbia, Canada. *Health Physics*, 106(5), pp.608–613. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24670910 [Accessed October 16, 2014].

Henderson, S.B., Kosatsky, T. & Barn, P., 2012. How to Ensure That National Radon Survey Results Are Useful for Public Health Practice. *Can J Public Health*, 103(3), pp.231–234.

Hunter, N. et al., 2009. Uncertainties in radon related to house-specific factors and proximity to geological boundaries in England. *Radiation protection dosimetry*, 136(1), pp.17–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19689964.

Ielsch, G. et al., 2010. Mapping of the geogenic radon potential in France to improve radon risk management: methodology and first application to region Bourgogne. *Journal of environmental radioactivity*, 101(10), pp.813–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20471142 [Accessed September 25, 2013].

Kemski, J. et al., 2008. From radon hazard to risk prediction-based on geological maps, soil gas and indoor measurements in Germany. *Environmental Geology*, 56(7), pp.1269–1279. Available at: http://link.springer.com/10.1007/s00254-008-1226-z [Accessed November 5, 2013].

Kosatsky, T., Henderson, S.B. & Pollock, S.L., 2012. Shifts in mortality during a hot weather event in Vancouver, British columbia: Rapid assessment with case-only analysis. *American Journal of Public Health*, 102(12), pp.2367–2371.

Lett, R.E. et al., 2008. *GeoFile 2008-1: A Drainage Geochemical Atlas for British Columbia*, Available at: http://www.empr.gov.bc.ca/Mining/Geoscience/PublicationsCatalogue/GeoFiles/Pages/2008-1.aspx.

Long, S. & Fenton, D., 2011. An overview of Ireland's National Radon Policy. *Radiation protection dosimetry*, 145(2-3), pp.96–100.

Maantay, J.A. & Mclafferty, S., 2011. Environmental Health and Geospatial Analysis: An Overview. In J. A. Maantay & S. McLafferty, eds. *Geospatial Analysis of Environmental Health*. Dordrecht: Springer Netherlands, pp. 3–37. Available at: http://link.springer.com/10.1007/978-94-007-0329-2 [Accessed November 27, 2013].

Miles, J.C.H. & Appleton, J.D., 2005. Mapping variation in radon potential both between and within geological units. *Journal of Radiological Protection*, 25(3), pp.257–276. Available at: http://iopscience.iop.org/0952-4746/25/3/003/ [Accessed September 24, 2013].

Natural Resources Canada, 2015. Geoscience Data Repository for Geophysical Data. Available at: http://gdr.agg.nrcan.gc.ca/gdrdap/dap/search-eng.php.

Radon Environmental Management Corp., 2011. Radon Potential Map of Canada. Available at: http://www.radoncorp.com/pdf/presentationMappingPublic.pdf [Accessed November 14, 2013].

Rauch, S.A. & Henderson, S.B., 2013. A comparison of two methods for ecologic classification of radon exposure in British Columbia: residential observations and the radon potential map of Canada. *Canadian journal of public health*, 104(3), pp.e240–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23823889.

Sainz-Fernandez, C. et al., 2014. The Spanish Indoor Radon Mapping Strategy. *Radiation Protection Dosimetry*, 162(1-2), pp.58–62.

World Health Organization, 2009. *WHO Handbook on Indoor Radon: A Public Health Perspective* H. Zeeb & F. Shannoun, eds., Geneva.