



THORON DETECTION IN RADON SOURCES

EDWARD TIMKO, STEPHEN STYLES, AND LIAM WRUBLESKI

ABSTRACT. The detection of radon levels is important for reducing home and workplace exposure to ionizing radiation. For the detector under consideration, this detection is done by counting the number of alpha decays in any given time period, and a single measurement cannot distinguish between the different isotopes of radon that may be present. In this paper, we present a method to approximate the relative amounts of radon-222 and radon-220 in a particular sampling sequence by performing a linear regression to the theoretical expected count rate from each of these isotopes.

1. PROBLEM STATEMENT

Residential radon (Rn) progeny exposure is “the leading cause of lung cancer in non-smokers, and the second leading cause of lung cancer in smokers” [1]. Uranium (U) and thorium (Th) in the soil eventually decay into radon, which can then seep into basements and low-lying areas of the house. The two main radon isotopes are Rn-222, which is part of the U-238 decay chain, and Rn-220, also called thoron, which is part of the Th-232 decay chain. There is currently much interest in the Rn-220 contribution to radon progeny exposure, which has so far been largely ignored. Though Rn-220 has a relatively short half life and usually decays before it reaches the living areas in a house, its radioactive progeny can still pose a problem.

Radon is chemically inert, and is most often detected by its decays. Environmental Instruments Canada (EIC) produces a Radon Sniffer [2], which is used by radon mitigators and building scientists to find radon entry points. The sniffer works by pumping air through a filter that removes all radon progeny, after which it is passed into a detector that counts alpha particle emissions from the decaying particles. The detector only counts the total number of alpha decays in a given period, so it cannot distinguish between Rn-222, Rn-220, or their progeny without further processing. Currently, these sniffers assume the only radon species present is Rn-222.

The problem we were presented with was to determine a sampling scheme and algorithm that can reliably determine the approximate amounts of Rn-220 and Rn-222 in any particular radon-containing sample of air, given the existing capabilities of the sniffer.

2. METHOD

2.1. Primer on Radioactive Decay. The nuclei of some atoms can randomly and spontaneously decay. Such atoms, and the substances they comprise, are said to be *unstable*, or *radioactive*. Nuclear decay is accompanied by the emission of a particle.

There is a variety of particles that can be emitted on decay, but here we are only concerned with two : alpha (α) particles and beta (β) particles. The type of particle emitted in the decay is called the *mode* of that decay. Alpha or beta decay change the *nuclear species* (i.e. the number of protons and neutrons) of the given nucleus. Another common mode of decay is gamma decay, but this does not alter the nuclear species, and for this and other reasons it does not play a part in the problem at hand. The nuclear species of an atom is also known as that atom's *nuclide*.

When an atom decays, the result may also be unstable. A sequence of such decays form what is known as a *decay chain*, where an unstable substance A decays into another unstable substance B , which itself decays into substance C , etc. These decay chains continue until a stable nuclide is reached. An important quantity describing any radioactive substance is its half-life $t_{1/2}$, which is the amount of time over which a given particle of the substance has a 50% probability of having decayed. Equivalently, it is the time by which 50% of the substance is expected to have decayed.

The last point to make is that the probability of decay for each atom in any interval of time depends only on the length of that interval (i.e. the decay process is memoryless). It does not depend on how long that atom has existed overall, nor does it depend on the atoms around it. As such, mixing different radioactive substances does not change their individual behaviour. From the memoryless property of the process, it follows that the time it takes for a given atom to decay follows an exponential distribution, while the number of decays in a given period from a large amount of a pure radioactive substance approximately follows a Poisson distribution.

2.2. Modelling Expected Values. Suppose at time 0 you have a collection of $N(0)$ radioactive atoms, all of a single nuclide. On average, the number $N(t)$ of atoms you expect to find remaining at time t is given by

$$N(t) = N(0)e^{-\lambda t},$$

where λ , called the *decay constant* of the nuclide, is related to the half-life $t_{1/2}$ by $\lambda = \ln(2)/t_{1/2}$. The decay rate of this collection of atoms at time t is $\lambda N(t)$.

Before we proceed, a word about units. Radioactive quantities are frequently described not by mass or number of particles but in terms of *activity*, which in the equation above is the quantity $\lambda N(t)$. Activity carries units of decays per unit time. The most common units of activity in the applications being considered are becquerels (Bq) and picocuries (pCi). By definition, 1 Bq is 1 decay per second, and 1 pCi is equivalent to 0.037 Bq. In the context of radon mitigation, quantities are usually described in activity per volume, and the units are becquerels per cubic metre (Bq/m³) and picocuries per litre (pCi/L). One finds that 1 pCi/L is equal to 37 Bq/m³.

As mentioned before, the progeny of a radionuclide can also be radioactive, resulting in so-called decay chains. The word “chain” may be somewhat misleading, as

some radionuclides can decay in more than one way. For example, Bi-212 decays by alpha emission into Tl-208 with a probability of 33.7%, and decays by beta emission into Po-212 with a probability of 66.3%. The decay series which are relevant to this project are those of U-238 and Th-232, the relevant portions of which are illustrated in Figure 1.

As can be seen from Figure 1, many of the alternative decay branches occur only with small probability. Because these routes are improbable in most cases and so will contribute little to the decays seen by the detector, we neglect them in the model. Additionally, as Po-212 has a half-life of only $0.3 \mu\text{s}$ and the beta decay from Tl-208 is not seen, we neglect the different decay paths from Bi-212 and assume that Bi-212 decays directly to Pb-208 by alpha emission. Normally, we end a chain at a stable nuclide where it will no longer decay. However for Rn-222, we stop the decay at Pb-210 as it's half-life is 22 years, which can be considered stable for our model. More precise values for these decay chains can be obtained from NuDat[3], but does not significantly change the model. As a consequence, the decay chains we consider in our model are linearly ordered.

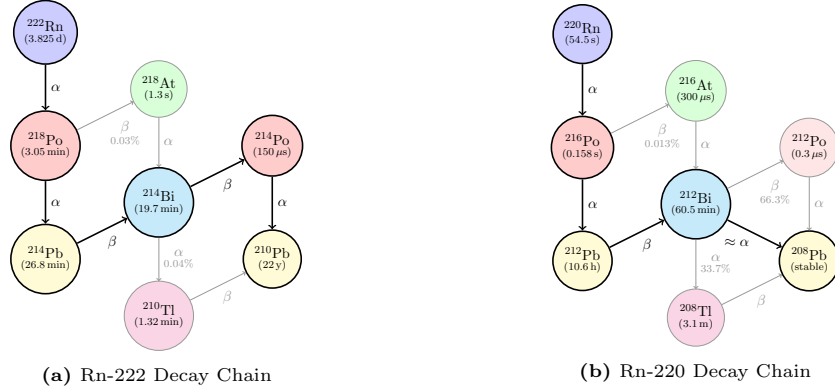


Figure 1. Decay series for Rn-222 and Rn-220, respectively. These are based on [4, Chapter 15]. The effective linear chains are indicated in bold.

Consider a (linearly ordered) decay chain $X_0 \rightarrow X_1 \rightarrow \dots \rightarrow X_\ell$ in which the decay constant of X_j is λ_j for $j = 0, 1, \dots, \ell - 1$, and X_ℓ is stable. We assume throughout that $\lambda_0, \lambda_1, \dots, \lambda_{\ell-1}$ are distinct and positive. Starting at time 0 with a collection $N_0(0)$ atoms of X_0 , $N_1(0)$ atoms of X_1 , etc., let $N_0(t), N_1(t), \dots, N_{\ell-1}(t)$ denote the expected number of atoms of the respective type in the given decay chain remaining at time t . When $0 < j < \ell$, the number $N_j(t)$ can change over time either by the decay of an atom of type X_j into one of type X_{j+1} (decreasing $N_j(t)$ in the process) or by an atom of X_{j-1} decaying into an atom of X_j (thereby increasing $N_j(t)$). The time evolution of the ensemble is thus described by the following system of ordinary differential equations:

$$\frac{dN_0}{dt} = -\lambda_0 N_0, \quad \frac{dN_j}{dt} = -\lambda_j N_j + \lambda_{j-1} N_{j-1} \quad (0 < j < \ell).$$

Compare with Equation (5) in [4, Chapter 15].

Recall that the detector does not measure the number of atoms remaining $N_j(t)$ but instead decay counts, the expected number of which in an interval $(s_1, s_2]$ is given by a sum of expressions of the form $\int_{s_1}^{s_2} \lambda_j N_j(s) ds$. Nevertheless, it will be in our interest to solve for N_j first. The solutions of the system are provided by the *Bateman Equation* [5] when $N_1(0) = \dots = N_{\ell-1}(0) = 0$:

$$N_j(t) = \frac{N_0(0)}{\lambda_j} \sum_{r=0}^j \left(\prod_{q=0, q \neq r}^j \frac{\lambda_q}{\lambda_q - \lambda_r} \right) \lambda_r e^{-\lambda_r t}.$$

Thus the number of decays by atoms of the j -th type in the time interval $(s_1, s_2]$ is

$$(2.1) \quad \int_{s_1}^{s_2} \lambda_j N_j(s) ds = N_0(0) \sum_{r=0}^j \left(\prod_{q=0, q \neq r}^j \frac{\lambda_q}{\lambda_q - \lambda_r} \right) (e^{-\lambda_r s_1} - e^{-\lambda_r s_2})$$

Since our detector only detects alpha decays, the number of counts we expect to see in the time interval $(s_1, s_2]$ is given by

$$\sum_{j \text{ } \alpha\text{-decays}} \int_{s_1}^{s_2} \lambda_j N_j(s) ds.$$

As each of these expressions shares the same independent variable $N_0(0)$, we may determine the total counts expected from this decay chain in any fixed interval as some function of only $N_0(0)$, provided we are working with a single decay chain. For the problem under consideration, we must work with two. The decay chains for Rn-222 and Rn-220 do not intersect, and thus the total counts from a collection of both of these isotopes of radon is the sum of the counts from the individual decay chains. In particular, the expected total number of counts in a time interval $(s_1, s_2]$ from an initial mixture of $N_{222}(0)$ atoms of Rn-222 and $N_{220}(0)$ atoms of Rn-220 is given by an expression of the form

$$N_{222}(0) \int_{s_1}^{s_2} f_{222}(t) dt + N_{220}(0) \int_{s_1}^{s_2} f_{220}(t) dt,$$

where f_{222} and f_{220} are functions derived from (2.1). This indicates that linear regression will provide an effective means of extracting the coefficients to determine the quantities of Rn-222 and Rn-220 present in a given sample.

Using the expected number of decays and the observed counts, we use a linear regression model to estimate our initial amount of each isotope. Letting P be the total number of sampling periods, we fit the model:

$$y_i = N_{222}(0)x_{222i} + N_{220}(0)x_{220i} + \epsilon_i$$

where y_i is the number of alpha particle decays observed in the time interval $(s_{i-1}, s_i]$, $x_{222i} = \int_{s_{i-1}}^{s_i} f_{222}(t) dt$, $x_{220i} = \int_{s_{i-1}}^{s_i} f_{220}(t) dt$, s_i denotes the end of sample period i , for $1 \leq i \leq P$, and s_0 denotes the beginning of sample period 1. Note that

the errors $\epsilon_1, \dots, \epsilon_P$ are random but not independent of one another.

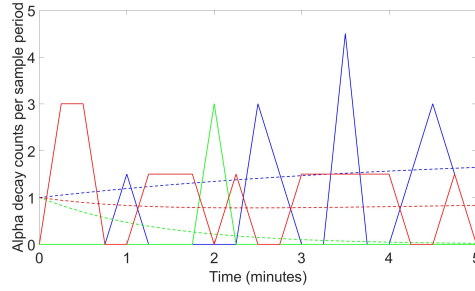
2.3. Simulating Experimental Values. In order to design the optimal sampling scheme, we must be able to evaluate the effectiveness of the linear regression model for any particular scheme. Determining this analytically is not straightforward, as the counts in any particular period are random and depend on the state underlying the counts in the previous period, as well as the length of each period. Even for a single particle, the time at which it decays into any other particle in the chain is the sum of random variables from several exponential distributions with different parameters. Therefore, we elected to evaluate the effectiveness of a scheme over test data, which we generated. We also narrowed the scope of sampling schemes under consideration to schemes where counts are reported at fixed times, over a fixed duration of time, where the contents of the cell are held constant over that duration. This sampling scheme does also require that we consider a time offset in order to account for the time it takes to fill the chamber, or else we would underestimate the Rn-220. Phrased in terms of the previous section, $s_i = T_s \cdot i + T_f$ for $0 \leq i \leq P$, where T_s is time of the sampling period and T_f is the time taken to fill the chamber.

This differs from the typical method used by radon mitigators in the field, which is continuous sampling of some number of environments, interspersed with flushing periods. However, based on our early results, we do not believe that this is a viable scheme to determine the relative amounts of Rn-220 and Rn-222, as the half-lives are too different and variations during pumping are not well understood. This change in procedure is a reasonable compromise, as in most circumstances knowing the Rn-220 concentration isn't necessary, and in those situations where it is the sampling scheme we propose is acceptable.

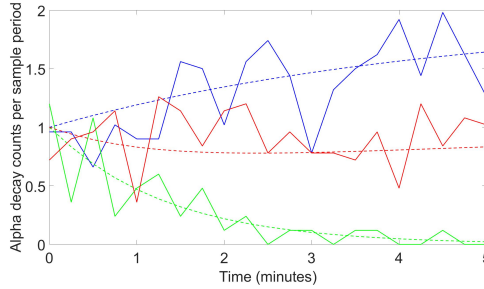
Generating test data for a sampling scheme of this type is relatively straightforward, but as the random processes involved are complicated, the best method we were able to come up with is not particularly efficient, having time complexity of $O(N \cdot \ell)$, where N is the number of particles in the chain at time $t = 0$ and ℓ is the length of the chain. We assume that at time $t = 0$, for a given decay chain, that only the substance at the head of the chain is present. Then, for each particle of that substance, we sample the *particle lifetimes*, which is the time that particle will take to decay from each substance in the decay chain to the next substance. Then the time at which that particle decays into each substance is the cumulative sum of the particle lifetimes up to that point. Using these decay times for the particles, we can then count how many alpha decays occur in each sampling period, which gives us the observed count for that sampling period.

2.4. Model Visualization. To demonstrate the relationship between the expected counts and the generated count data, we graph the expected count rate and a corresponding sample run for a series of concentration levels in Figure 2. Homeowners are advised to take action if the radon activity is greater than 4 pCi/L in air [6],

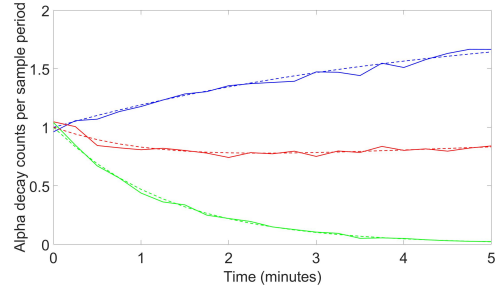
sub-slab measurements can be on the order of 100 pCi/L, and the largest value that Environmental Instruments Canada has reported to us was around 8000 pCi/L. By showing these values, we can see the amount of variation implicit to this problem, and note that higher activities make it much easier to determine the amounts of Rn-222 and Rn-220.



(a) Activity level of 4 pCi/L



(b) Activity level of 100 pCi/L



(c) Activity level of 8000 pCi/L

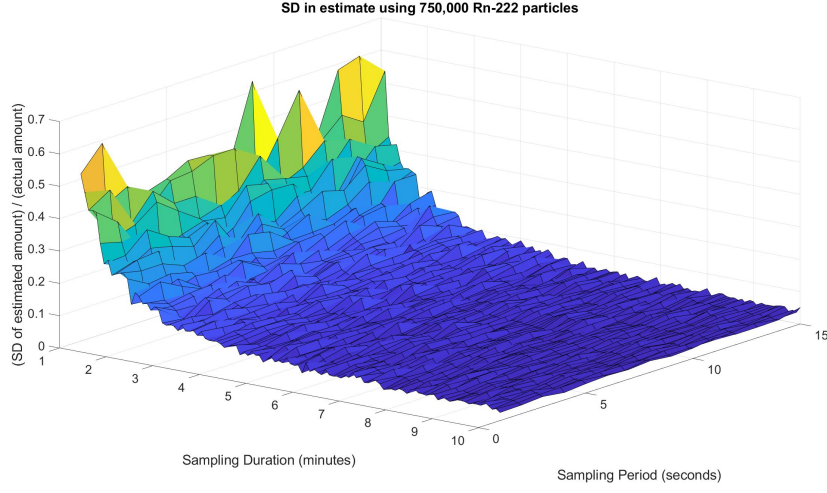
Figure 2. Sample runs and expected values for various mixtures of Radon isotopes (blue: pure Rn-222, red: Rn-222/Rn-220 mixture, green: pure Rn-220; dashed: expected counts, continuous: observed counts)

3. PARAMETER ESTIMATION

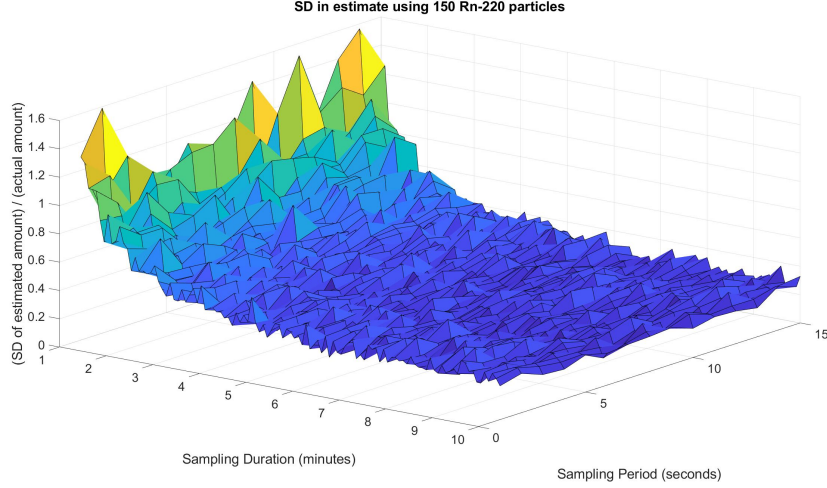
3.1. Grid Search. To find the sampling time and sampling period which minimize the variance in the estimates, we run a grid search that calculates multiple regressions on a fixed set of concentrations. Running a new sample set of decays for each regression is done to avoid over-fitting the model to a training set of data. From here, we calculate the mean and standard deviation of the estimated coefficients for various combinations of the sampling time and period to find appropriate values, attempting to minimize both the error relative to the true concentration, and the standard deviation in the estimates. With any sampling time below 90 seconds, we were not able to see a separation between the Rn-222 and Rn-220. When we sample for longer, our estimates start to level off in accuracy and increased time is no longer needed. From figure 3 we see that the standard deviations level off at approximately 5 minutes duration and are mostly independent of sampling period.

For each potential sampling time and period, the actual model with which estimations can be made takes the form of the $2 \times P$ matrix described by the left pseudo-inverse of X , $(X^T X)^{-1} X^T$, where the $P \times 2$ matrix $X = \begin{bmatrix} x_{222} & x_{220} \end{bmatrix}$ is formed using the values for x_{222} and x_{220} determined using the grid search, and when given a sample vector y , the estimated amounts \hat{N}_{222} and \hat{N}_{220} are determined by

$$\begin{bmatrix} \hat{N}_{222} \\ \hat{N}_{220} \end{bmatrix} = (X^T X)^{-1} X^T y$$



(a) Standard deviation of 25 different linear regression estimates for radon



(b) Standard deviation of 25 different linear regression estimates for thoron

Figure 3. Standard deviation in estimated values over varied sampling parameters

3.2. Results. After tuning our sampling parameters, we requested and received data on different radon and thoron sources from EIC. Using this data, we fit our linear regression model and solve for our initial concentrations of Rn-222 and Rn-220. Using those estimates rounded up to the next whole number, we then solve for what the expected number of alpha decays would be over our 5 minute interval and generate test data for the same amount of each nuclide. The following graphs highlight the accuracy of our model.

From figures 4 (a)-(d), we see that our model can predict the radon levels for a typical basement as well as the extreme cases where we are measuring high concentrations of Rn-222. These measurements were done all in the same day so the progeny of the previous test would be included in each subsequent test, therefore causing more error into our model. Although our predictions appear to be fairly accurate, they rely on the assumption that our radon sniffing chamber is empty. This progeny remaining in the sniffer will cause our model to overestimate the true values and a correction to the alpha decay count could increase the effectiveness of our model.

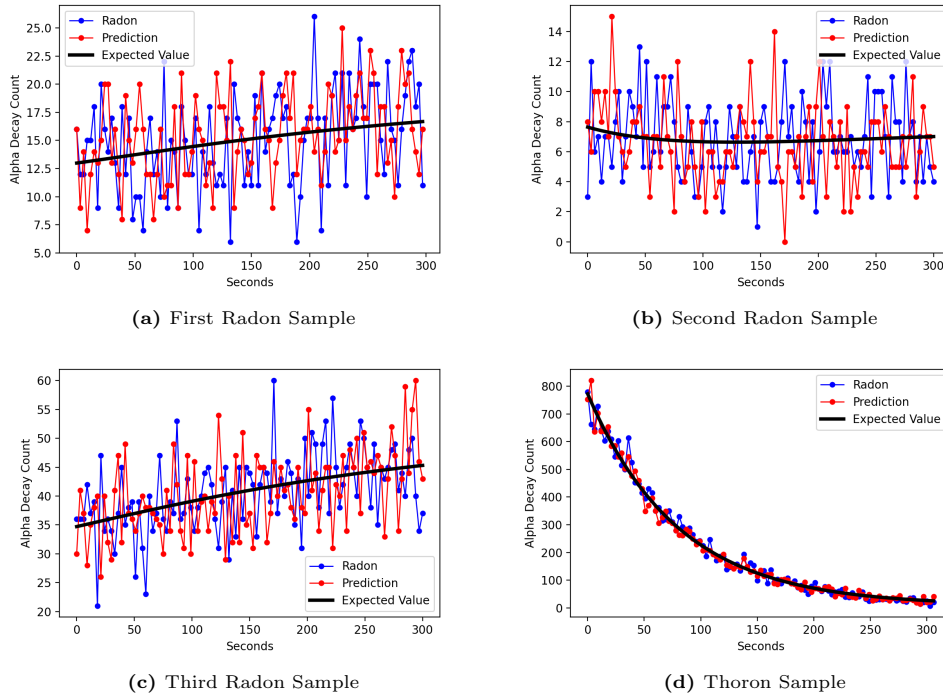


Figure 4. Observed and Predicted Values of Radon Samples

3.3. How to use the results. Overall, based on the efficacy of our model on these sample data, and in consultation with Environmental Instruments Canada, we have selected a sampling period of 3 seconds, over a total duration of 5 minutes. The particular sampling schedule is as follows:

- (1) Flush the cell for 90 seconds (this is probably an overestimation of the time required).
- (2) Pump air from the spot to be sampled into the cell for 90 seconds (this is also probably an overestimation of the time required).
- (3) Stop the pump, to let the sample be measured. Counting for this sequence should begin here.
- (4) Let the sample sit for 5 minutes, reporting counts every 3 seconds.
- (5) Take the reported counts as a 100 element column vector.
- (6) Multiply this vector by the 2×100 matrix $(X^T X)^{-1} X^T$, where X is the 100×2 matrix $\begin{bmatrix} x_{222} & x_{220} \end{bmatrix}$, using x_{222} and x_{220} determined from the solutions to the Bateman equations over the intervals (Python code to generate each of x_{222} , x_{220} , X , and $(X^T X)^{-1} X^T$ is available on the project's GitHub repository)[7]. Note also that:
 - the values x_{222} and x_{220} also need to account for the delay of pumping (in this particular setup, this can be done simply by computing the solutions for the first 130 intervals, and ignoring the first 30); and
 - this matrix is constant for a given sampling schedule, and does not need to be recomputed each time.
- (7) This gives the 2-element column vector $[\hat{N}_{220} \quad \hat{N}_{222}]^T$, which estimates the number of particles of each isotope that were in the chamber at the beginning of pumping (under the various assumptions of the underlying physics within the chamber).
- (8) Assuming that λ is given in s^{-1} , the absolute activity of each can then be determined in Bq as $\lambda \hat{N}$, and the volumetric activity can then be determined using the volume of the scintillation cell.

4. FURTHER RESEARCH

Further analysis could be done on the model to consider various constraints. A non-negative least squares regression would constrain the estimated amounts of each isotope to be positive. In our experimentation with this model, it tended to remove one of the covariates and predict either pure Rn-222 or pure Rn-220. Other regression methods such as Poisson regression or negative binomial regression could be considered to account for the discrete output, however when we compared our results to a normal Q-Q plot, the model fairly accurately followed a normal distribution, so these other methods may be unnecessary. Adding a correction term to the alpha decay count to account for the progeny left in the chamber could also increase the accuracy of the model.

To speed up the optimization, instead of considering a new dataset for each regression we could consider a fixed number of data sets for training. This would allow us to reduce the computation time, allowing us to check a finer grid of parameters.

Grid search is not a particularly effective method of finding the optimal configuration, so another optimization algorithm may be useful. The reason this was not

done here is that the optimization variables include the sampling period and duration, and as these determine not only the distributions of the random variables under consideration, but also the constants used in any optimization model, this makes creating an analytic model which can be given to a solver very difficult. Stochastic methods using randomly generated data are better suited to this task, but still encounter issues.

Developing a model which determines the amount of each nuclide over time when the chamber is being filled from a source of particular concentration may allow this model to use additional data from the fill period or the purge period, and may also allow it to be extended to continuous sampling.

The relative error in the amount of Rn-220 is quite high. To achieve activity commensurate with that of Rn-222, one requires much lower concentrations of Rn-220, so the relative variability in counts per period is much higher. There may be heuristic methods of determining whether there is Rn-220 that could supplement this model, which may reduce the error caused by this, but we do not consider them here.

Due to the aforementioned inefficiency, the method we used for generating random data is only viable for relatively small numbers of particles ($N \ll 10,000,000$). For numbers of particles greater than this, another method should be developed. We sidestepped this issue by running trials in parallel, but this solution will not work for significantly larger numbers of particles.

5. CONCLUSION

Using the 90 second procedure of continuous sniffing, we were able to measure how much Rn-222 is in our measurements. However, using this same sample scheme, we were unable to consistently differentiate between data obtained by measuring pure Rn-222 and that obtained from a mixture of Rn-222 and Rn-220. As such, we evaluated different sampling times and periods in which to most accurately estimate the quantity of each radon isotope. With this new sampling scheme of 3 second samples counted over a time of 5 minutes, we developed an algorithm by which the amount of Rn-220 in a sub-slab measurement of radon isotopes can be estimated through a linear regression model. In tandem with existing methods, this can be used to assist radon mitigators in making mitigation decisions by ensuring they have as much information as possible about the situation, without drastically increasing the time for each test.

ACKNOWLEDGEMENTS

We wish to thank Kai Kaletsch from Environmental Instruments Canada for assisting us in his role as an industry mentor, and for providing an interesting challenge to work on. We also wish to thank the Pacific Institute for Mathematical Sciences (PIMS) for running the *Math^{Industry}* workshop, and the organizers within PIMS who put in a great deal of effort to make it happen.

REFERENCES

- [1] Lung Cancer Canada, "Radon", <https://www.lungcancercanada.ca/Lung-Cancer/Radon.aspx#:~:text=Health%20Canada%20recently%20increased%20its,deaths%20globally%20are%20radon%20induced>.
- [2] Environmental Instruments Canada, "CT007-R Radon Sniffer", <https://radonsniffer.com/>
- [3] Brookhaven National Laboratory, "NuDat 2", <https://www.nndc.bnl.gov/nudat2/>
- [4] R. B. Leighton, *Principles of Modern Physics*, McGraw-Hill, New York, 1959.
- [5] H. Bateman, *The solution of a system of differential equations occurring in the theory of radioactive transformations*. Proc. Cambridge Philos. Soc, Vol. 15, No. pt V (1910) pp. 423–427.
- [6] US Environmental Protection Agency, "Health Risk of Radon", <https://www.epa.gov/radon/health-risk-radon>
- [7] <https://github.com/StephenStyles/Radon>