OAK RIDGE National Laboratory

# Mapping Estimates Of Residential Radon Exposure At Small Spatial Scales

Heidi A. Hanson, PhD
Senior Scientist, Advanced Computing for Health Sciences | Computational Sciences and Engineering Division

Oak Ridge, TN
October 31, 2023

ORNL is managed by UT-Battelle LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

# Goal: Develop models that predict indoor radon exposure at small spatial scales

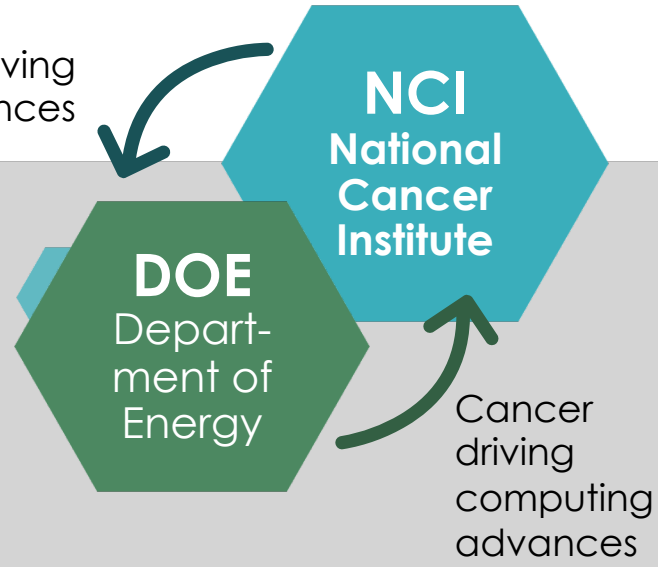
- Radon is a widespread environmental hazard requiring monitoring and management.[1]
- The carcinogenic gas is the second-leading cause of lung cancer in the United States (US) and is linked to lung cancer incidence.[2-4]
- Radon gas is estimated to contribute to 21,000 lung cancer deaths per year, which accounts for 16% of all lung cancer deaths.[5,6]
- Radon-attributable lung cancer deaths are more common than the approximately 8,700 annual deaths from melanoma, another common adult cancer.[5,6]
- Ou and Hanson found that measures of Radon Hazard Potential in the Soil were better than publicly available measures.[7]


1. Chahine T, Schultz BD, Zartarian VG, Xue J, Subramanian SV, Levy JI. Modeling joint exposures and health outcomes for cumulative risk assessment: the case of radon and smoking. Int J Environ Res Public Health. 2011;8(9):3688-711. Epub 2011/09/13. doi: 10.3390/ijerph8093688. PubMed PMID: 22016710.
2. Krewski D, Lubin JH, Zielinski JM, Alavanja M, Catalan VS, Field RW, Klotz JB, Letourneau EG, Lynch CF, Lyon JI, Sandler DP, Schoenberg JB, Steck DJ, Stolwijk JA, Weinberg C, Wilcox HB. Residential radon and risk of lung cancer: a combined analysis of 7 North American case-control studies. Epidemiology (Cambridge, Mass). 2005;16(2):137-45. Epub 2005/02/11. PubMed PMID: 15703527.
3. Krewski D, Lubin JH, Zielinski JM, Alavanja M, Catalan VS, Field RW, Klotz JB, Letourneau EG, Lynch CF, Lyon JL, Sandler DP, Schoenberg JB, Steck DJ, Stolwijk JA, Weinberg C, Wilcox HB. A combined analysis of North American case-control studies of residential radon and lung cancer. Journal of toxicology and environmental health Part A. 2006;69(7):533-97. Epub 2006/04/13. doi: 10.1080/15287390500260945. PubMed PMID: 16608828.
4. Lantz PM, Mendez D, Philbert MA. Radon, Smoking, and Lung Cancer: The Need to Refocus Radon Control Policy. American Journal of Public Health. 2013;103(3):443-7. doi: 10.2105/AJPH.2012.300926. PubMed PMID: PMC3673501.
5. Environmental Protection Agency. Health Risk of Radon 2021 [cited 2020 September 6]. Available from: https://www.epa.gov/radon/health-risk-radon.
6. Surveillance E, and End Results Program. Cancer Stat Facts: Lung and Bronchus Cancer: National Cancer Institute; 2021 [cited 2021 January 3]. Available from: https://seer.cancer.gov/statfacts/html/lungb.html.
7. Ou JY, Ramsay JM, Lee G, VanDerslice JA, Taddie M, Kirchhoff AC, Divver E, Akerley W, Kepka D, Hanson HA. Patterns of indoor radon concentrations, radon-hazard potential, and radon testing on a small geographic scale in Utah. J Environ Radioact. 2023 Jan;256:107046. doi: 10.1016/j.jenvrad.2022.107046. Epub 2022 Oct 31. PMID: 36327525.

MOSSAIC

Modeling Outcomes Using Surveillance Data & Scalable Artificial Intelligence For Cancer

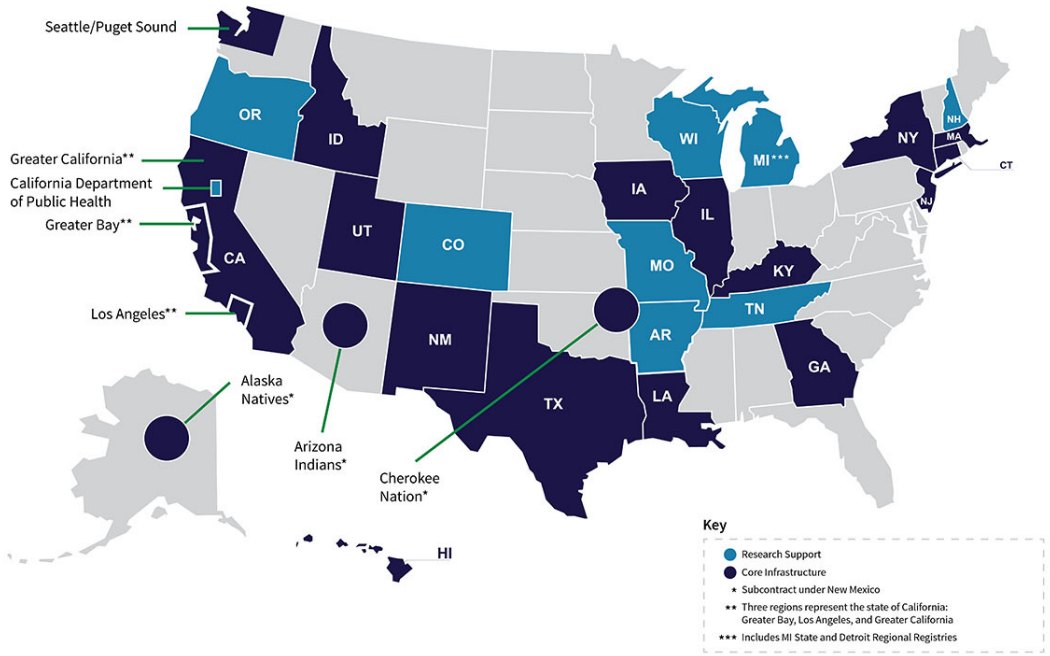
# Investors and technical leads




- Principal Investigators
  - Dr. Lynne Penberthy, Associate Director of the Surveillance Research Program, NCI
  - Dr. Heidi Hanson, Group Leader, Biostatistics and Biomedical Informatics, ORNL
- Technical Leads
  - Dr. Betsy Hsu, Chief, Surveillance Informatics Branch, NCI
  - Dr. John Gounley, Group Leader, Scalable Biomedical Modeling, ORNL

# Real world evidence: AI for near real-time health surveillance covering 48% of the US population

Surveillance Epidemiology End-Results (SEER) Registries > 850,000 Diagnoses Annually




**https://seer.cancer.gov/registries/**

**Auto-Extraction from Pathology Reports:**

Accuracy: Auto-coding of 23-27% of path reports (N ~ 230,000) with > 98% accuracy across all data elements

- Phenotype classifications:
  - Site = 70 categories
  - Sub-site = 324 categories
  - Histology = 626 categories
  - Laterality = 7 categories
  - Behavior = 4 categories

**Automatic identification of "hallucination" of model that can be autotuned**

**Production implementation Hierarchical Self Attention Model (HiSAN) with Deep Abstention:**

- Total 16 registries (additional 4 since March 2022; ~31% of US population)
- Default as part of any new DMS installation, regardless of SEER affiliation
- 5+ new registries anticipated in 2023/2024
- Testing phase with the Veteran's Health Administration

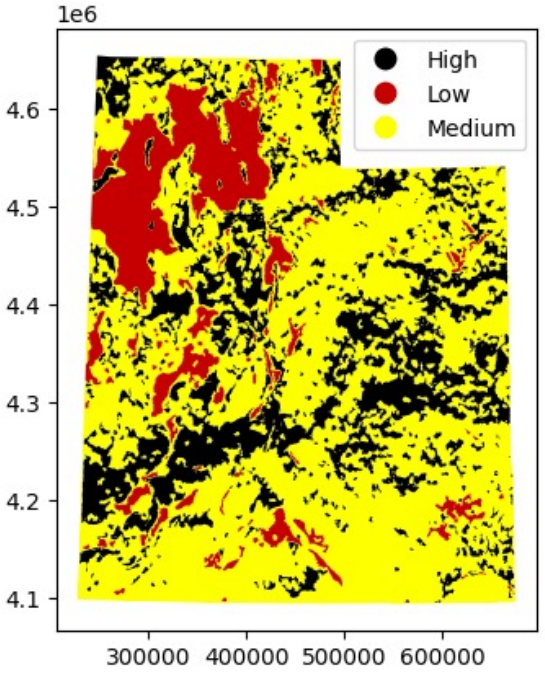
# SEER Residential History Data

Zaria Tatalovich, Heidi Hanson, Lynne Penberthy, Kevin Henry, David Stinchcomb

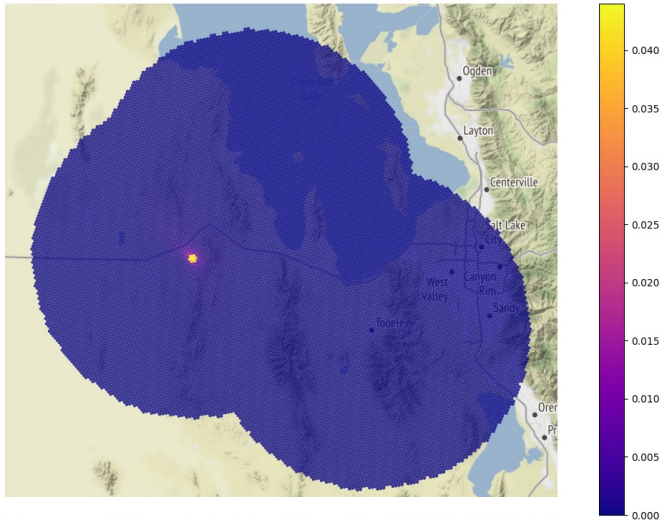
## LexusNexus Residential History data

- 11 SEER registries have been linked (3.2 million individuals diagnosed from 2005 - 2022)
- 15 SEER registries should be linked by the end of 2023
- Residential history constructed for Louisiana and expected to be expanded to all linked registries this year
- High quality data from 1995 – 2020
- 83% are geocoded to the point location
- IRB approved studies may gain access to the information

## Indoor Radon




## RSEI Microdata




estimated average daily concentration of PCBs (µg/m$^3$) along the Wasatch Front in 2005

## Air Pollution Data

OAK RIDGE National Laboratory

# Modeling indoor radon exposure at smaller spatial scales

- Case 1: Pennsylvania Zip Code Data

Dakotah Murdock, Jeremy Logan, Heechan Lee

# 637,706* Pre-mitigation residential indoor radon tests

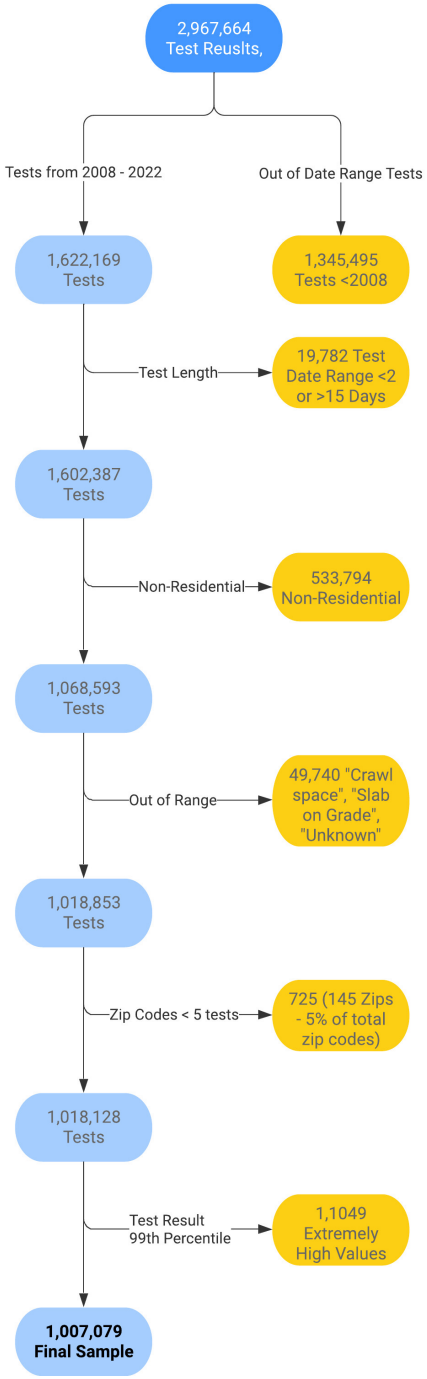
Data:  Radon Test Results from the Pennsylvania Department of Environmental Protection (DEP)

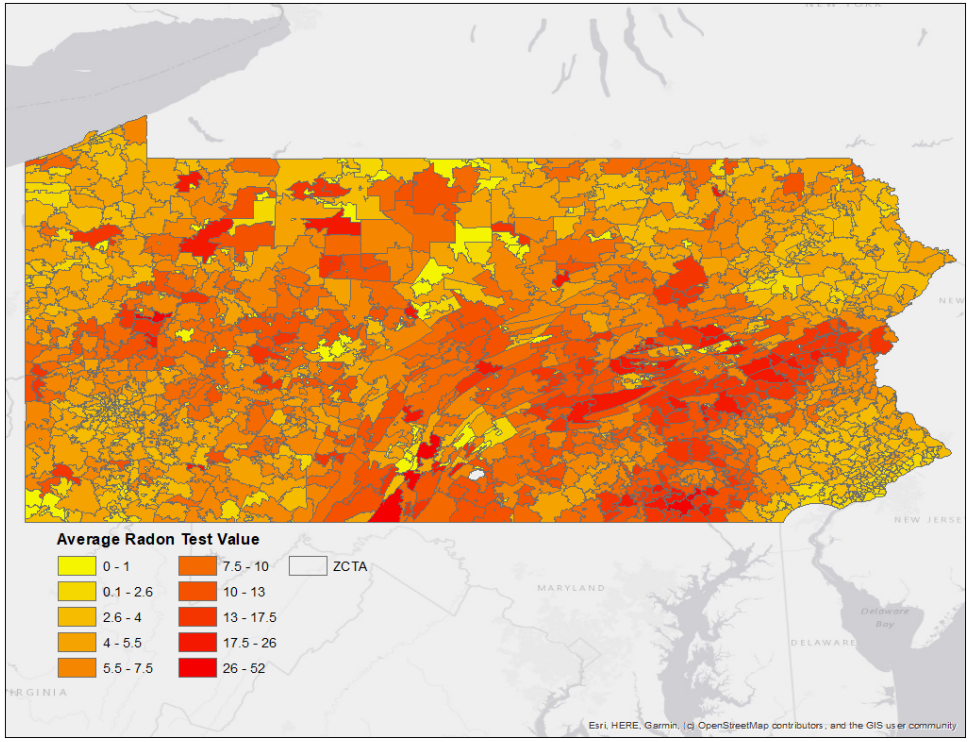
Indoor Radon Measure Value – Radon measures in pCi/L

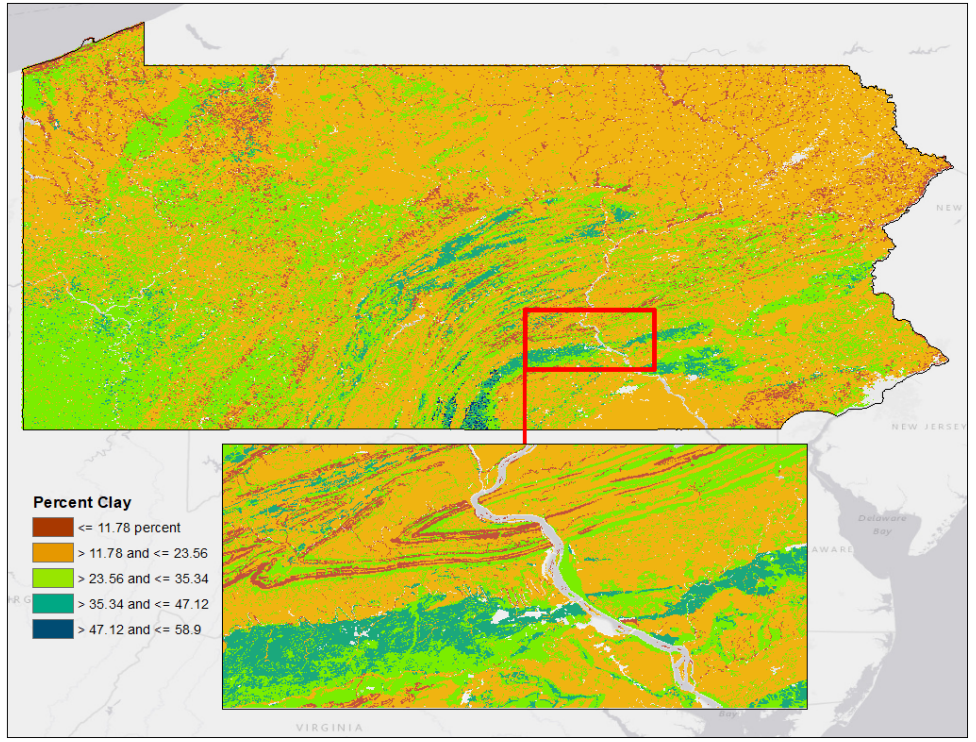
Spatial Resolution: Zip code

*Tests from 2018 – 2022 are not included in these results

https://data.pa.gov/Energy-and-the-Environment/Radon-Test-Results-September-1986-Current-Annual-C/vkjb-sx3k






Heterogenous ensemble methods for estimating indoor radon exposure over space and time.

- Geographically and Temporally Weighted Regression methods
  - Local regression method that gives higher weight to the best performing method in a region/space
- **Random Forest**
- Neural Network
- Stochastic Gradient Boosting Machine
- Generalized Linear Model

Datasets

gNATSGO
US Census
Elevation
Daymet – Climatedata

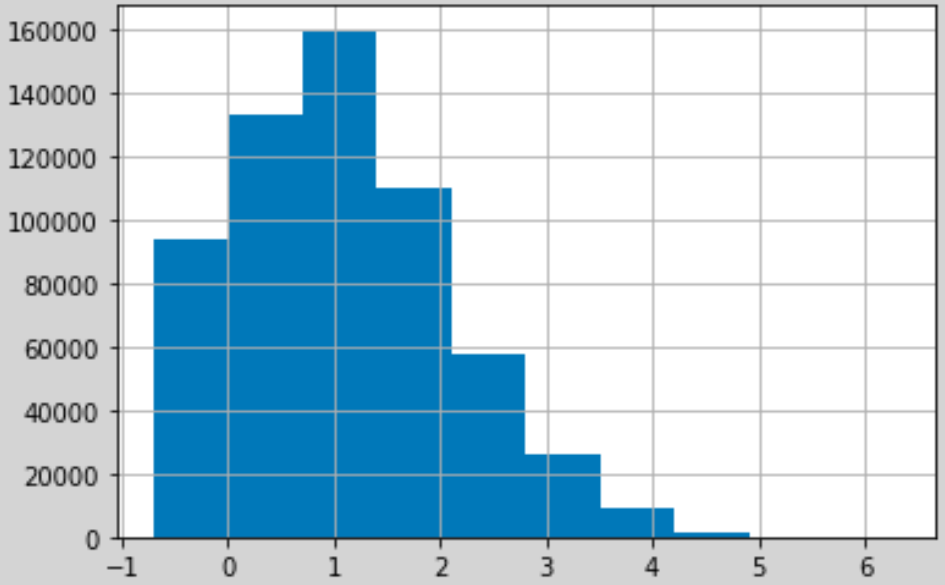
Variables included in the model

Soil drainage type
Saturated Hydraulic Conductivity
Hydrolic Soil Group
Depth to water table annual minimum
Available water storage
Soil Temperature Regime
Depth to Bedrock
Slope
Parent material dominant condition
Percent K
Ppm eU
Ppm eTh
Aquifer permeability class
Percent Sand in soil
Min elevation in watershed
Max elevation in watershed
Percent flat land
Percent flat land in watershed
Elevation (DEM)
Daymet data
%Basements in area
Year Structure Built
Rooms in House
House Heating Fuel
Lab
Year test was analyzed
Test location/floor

OAK RIDGE National Laboratory

# Methods

- Random Forest Regression
- 70% of the data used for training; 30% for testing
- Hyperparameter grid search (number of trees, depth of tree, number of features, minimum impurity, min number of samples per split)
  - Final Model used: Trees = 600, Depth = 10, Min Impurity Decrease = 0.00005, Minimum Samples per Split = 20
- Spatial Crossfold Validation at the county level




Logged Indoor Radon Test Values

MINELE: Minimum Elevation in Watershed

Elevation: Global Multi-Resolution Terrain Elevation Data 30m Resolution

B25040_00X: House Heating Fuel

RELIEF: Difference in the elevation between the basin outlet and the highest point in the drainage divide.

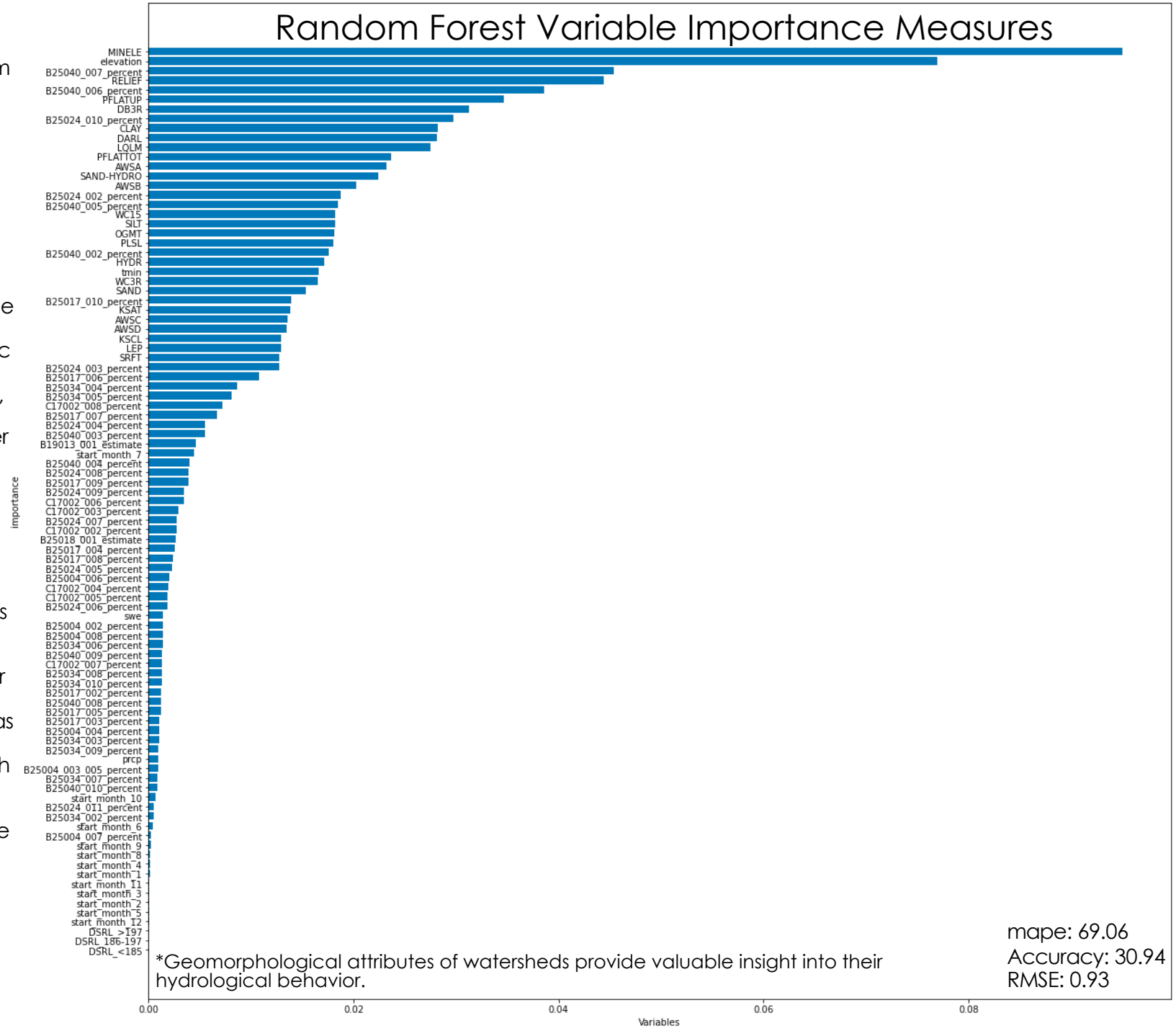
PFLATUP: Percent flat land (slope less than 1 percent) in watershed upland

DB3R: Bulk density, one-third bar, is the ovendry weight of the soil material less than 2 millimeters in size per unit volume of soil at water tension of 1/3 bar, expressed in grams per cubic centimeter. Bulk density data are used to compute linear extensibility, shrink-swell potential, available water capacity, total pore space, and other soil properties. The moist bulk density of a soil indicates the pore space available for water and roots. Depending on soil texture, a bulk density of more than 1.4 can restrict water storage and root penetration. Moist bulk density is influenced by texture, kind of clay, content of organic matter, and soil structure.

B25024_010E: % Units in Housing Structure Mobile Homes

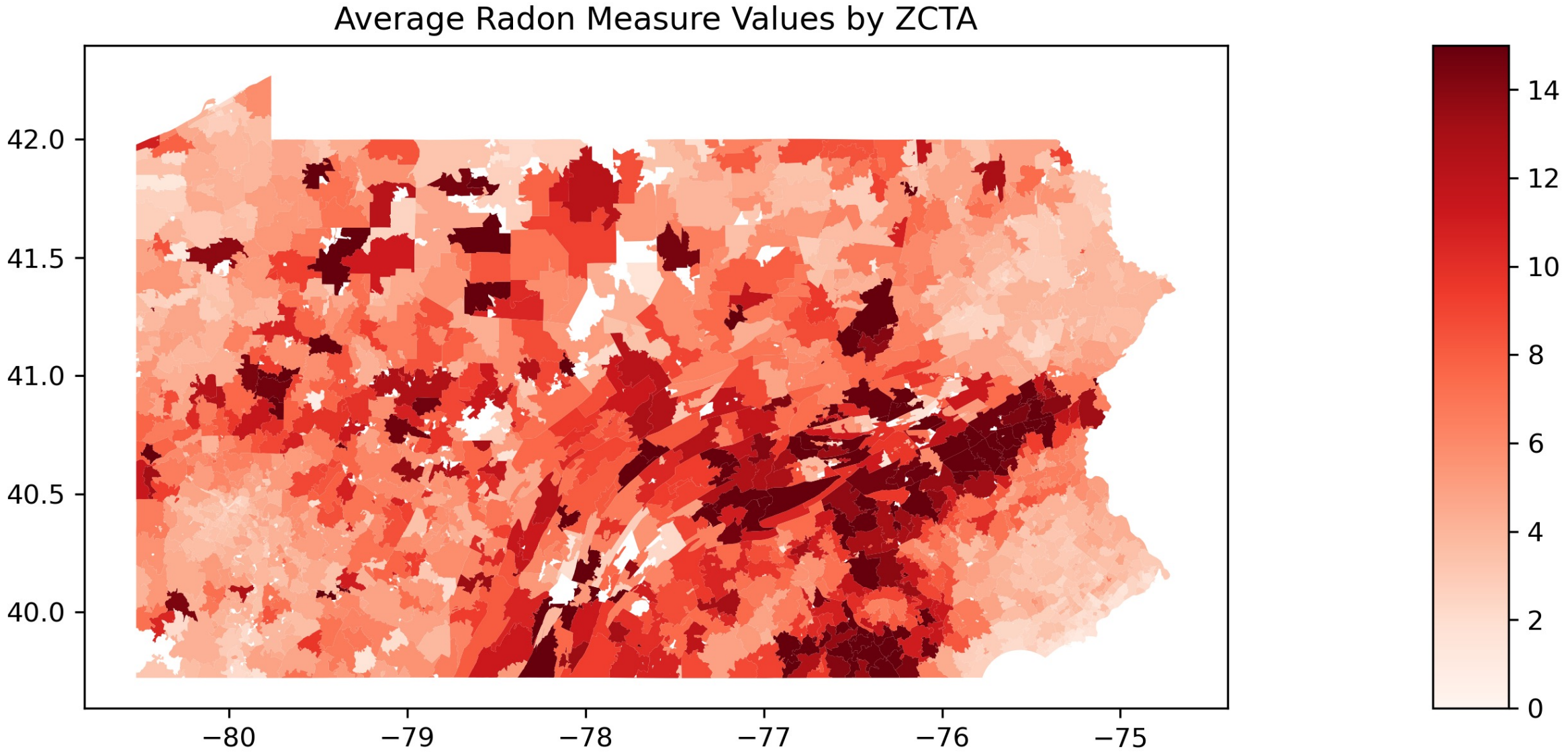
CLAY: Clay as a soil separate consists of mineral soil particles that are less than 0.002 mm in diameter. The estimated clay content of each soil layer is given as a percentage, by weight, of the soil material that is less than 2 mm in diameter

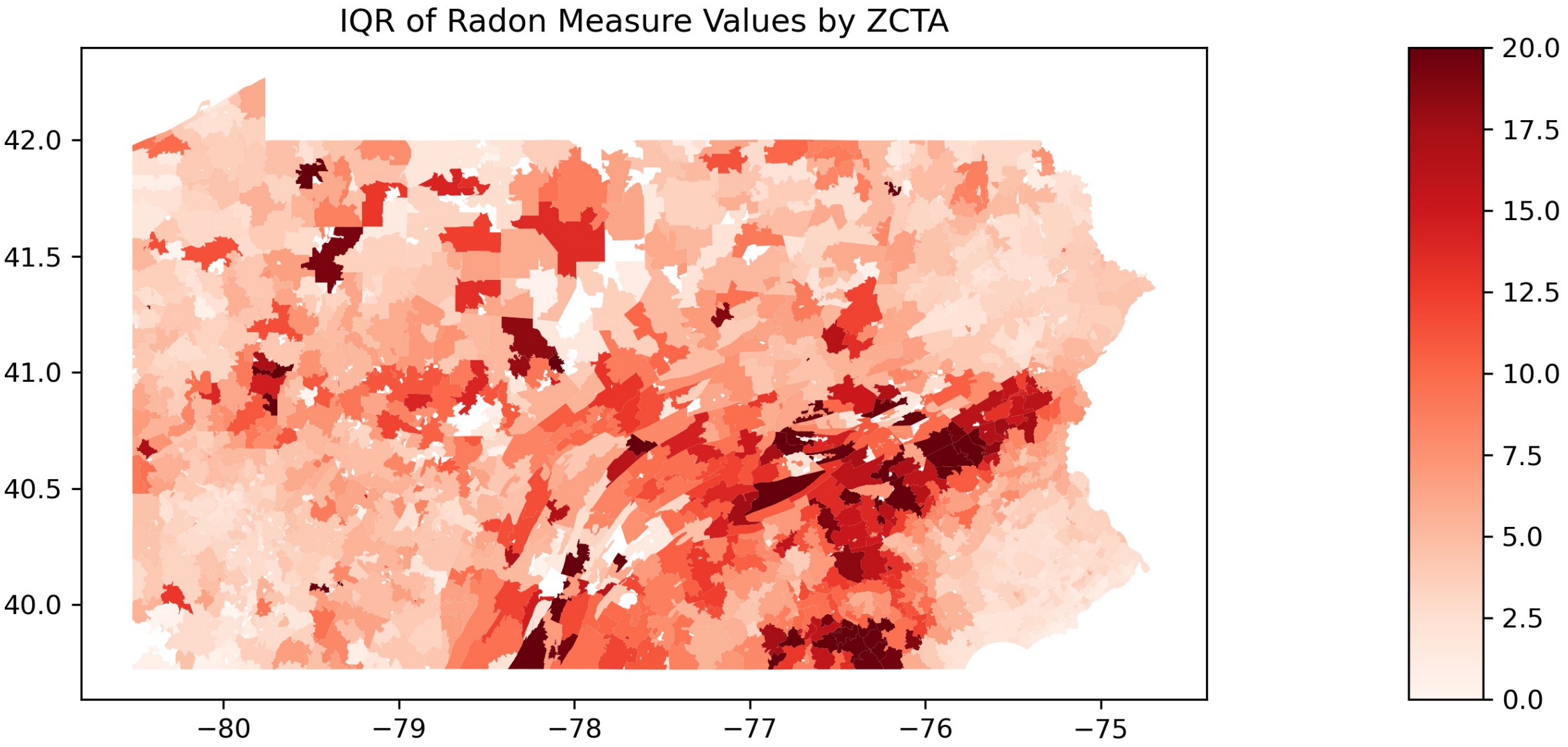
DARL: A "restrictive layer" is a nearly continuous layer that has one or more physical, chemical, or thermal properties that significantly impede the movement of water and air through the soil or that restrict roots or otherwise provide an unfavorable root environment. Examples are bedrock, cemented layers, dense layers, and frozen layers. This theme presents the depth to any type of restrictive layer that is described for each map unit. If more than one type of restrictive layer is described for an individual soil type, the depth to the shallowest one is presented. If no restrictive layer is described in a map unit, it is represented by the "greater than 200" depth class.






OAK RIDGE National Laboratory


Average Radon Measure Values by ZCTA
42.0
41.5
41.0
40.5
40.0
−80
−79
−78
−77
−76
−75
14
12
10
8
6
4
2
0




We need point level data!

OAK RIDGE National Laboratory

# Utah

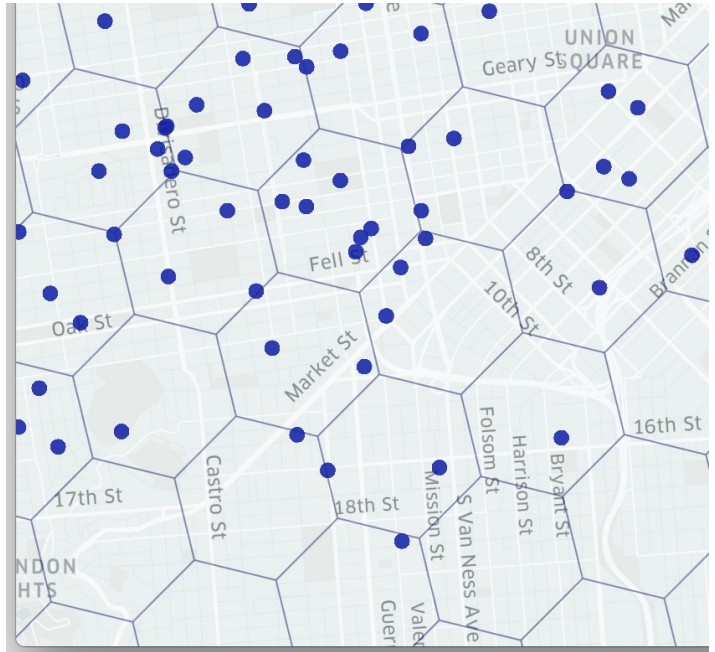
- 20,368 pre-mitigation radon tests conducted in Utah from 2001 – 2017
- Mailing addresses of radon tests were geocoded at the point level

HEALTH
UNIVERSITY OF UTAH

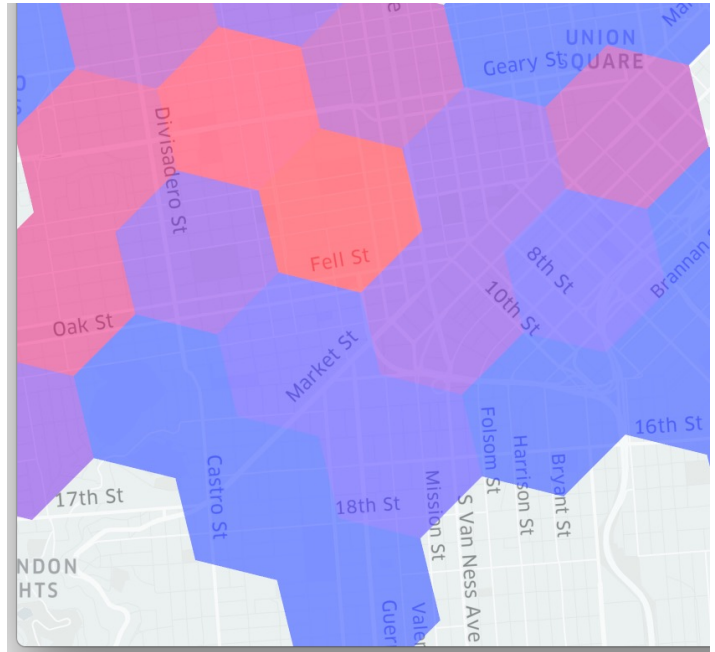
Joemy Ramsy, Marissa Taddie, Jame VanDerslice, Eleanor Divver, Dakotah Maguire, Jeremy Logan, Heechan Lee

# Stacking Environmental Datasets at small spatial scales with Uber H3 Hex


Coordinate information




Overlay with hexagon "fishnet"




Aggregate to new polygons

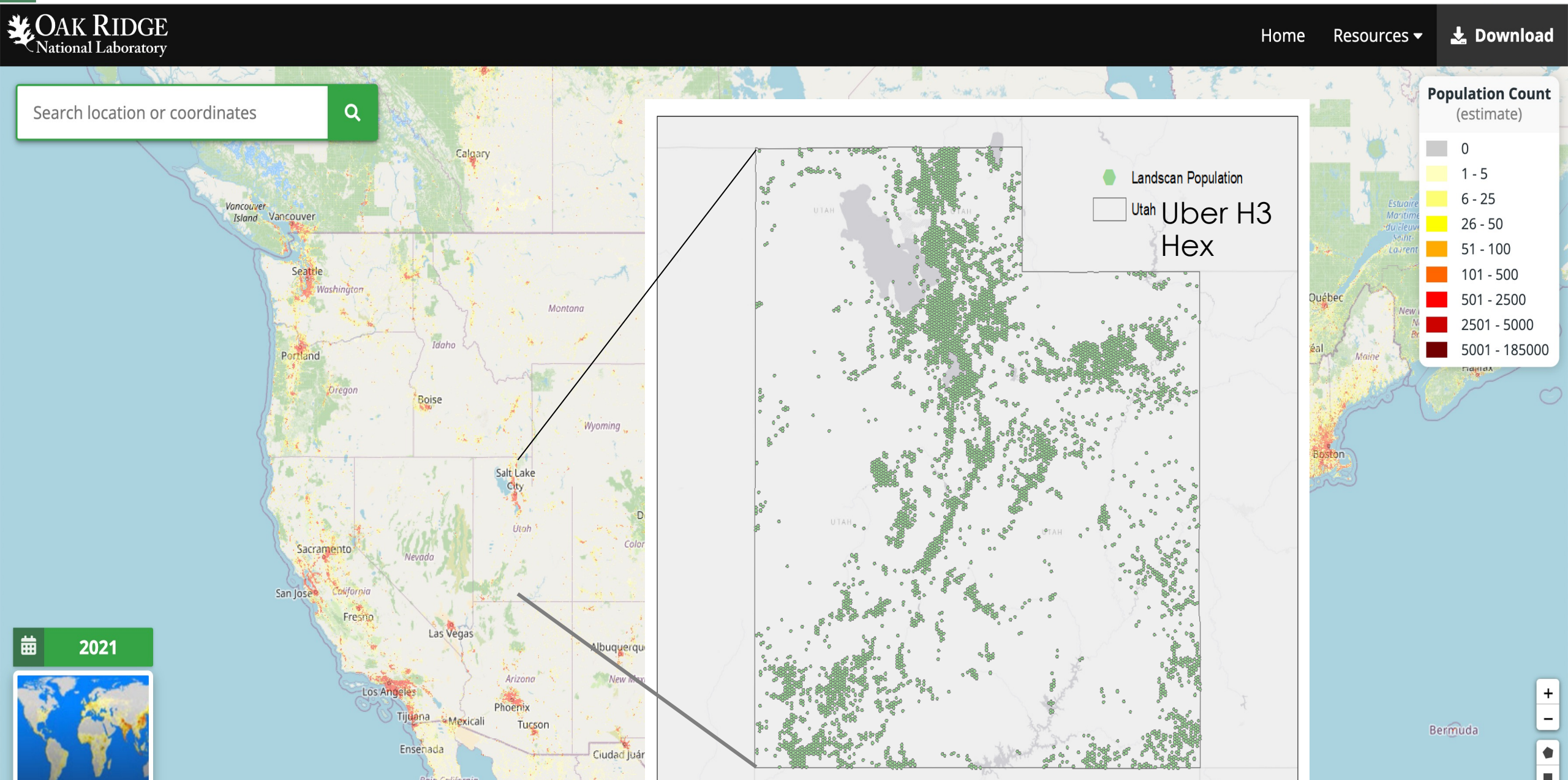
A spatial join can be used to map all datasets to the same spatially linked polygon. All environmental datasets can be stacked to create multi-exposure measures for a region.



OAK RIDGE National Laboratory

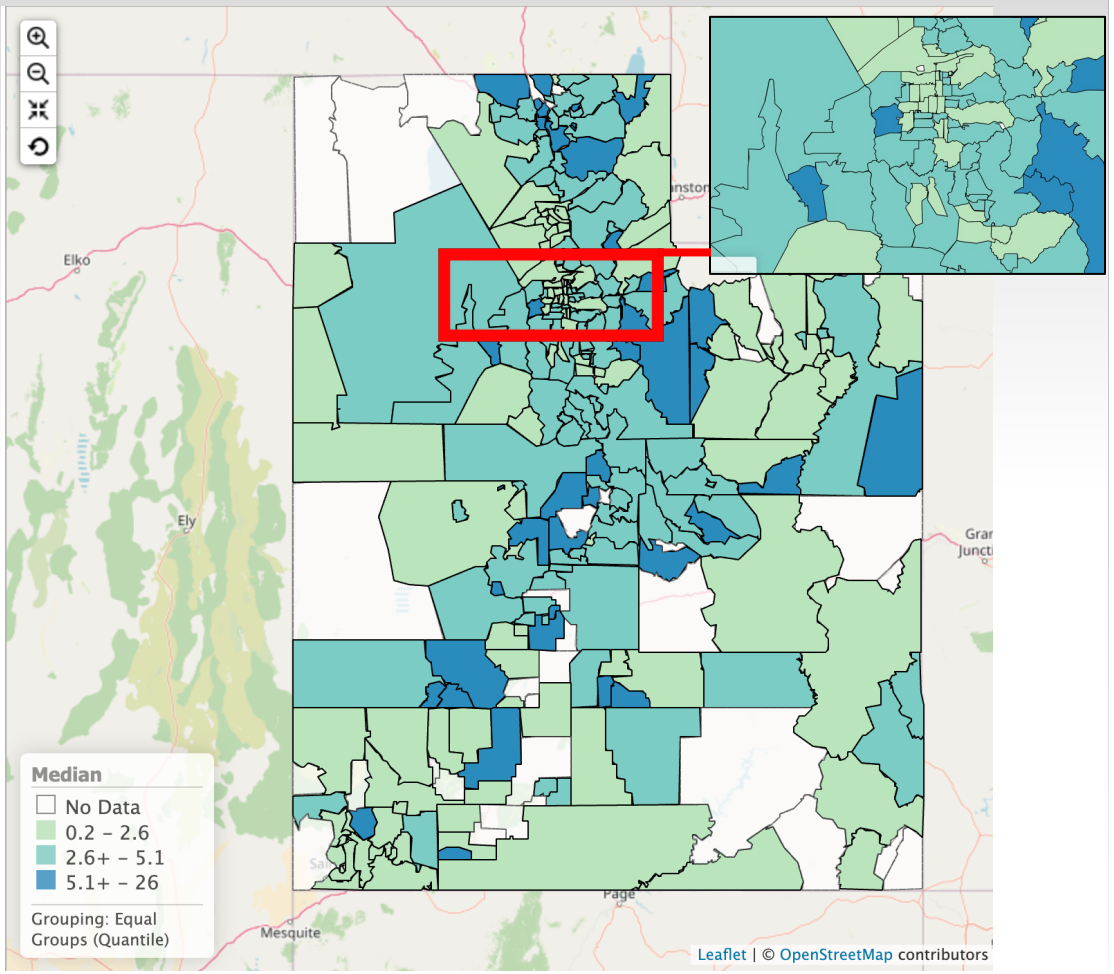
https://www.uber.com/en-FR/blog/h3/

# LandScan Population Data for Identifying Bias



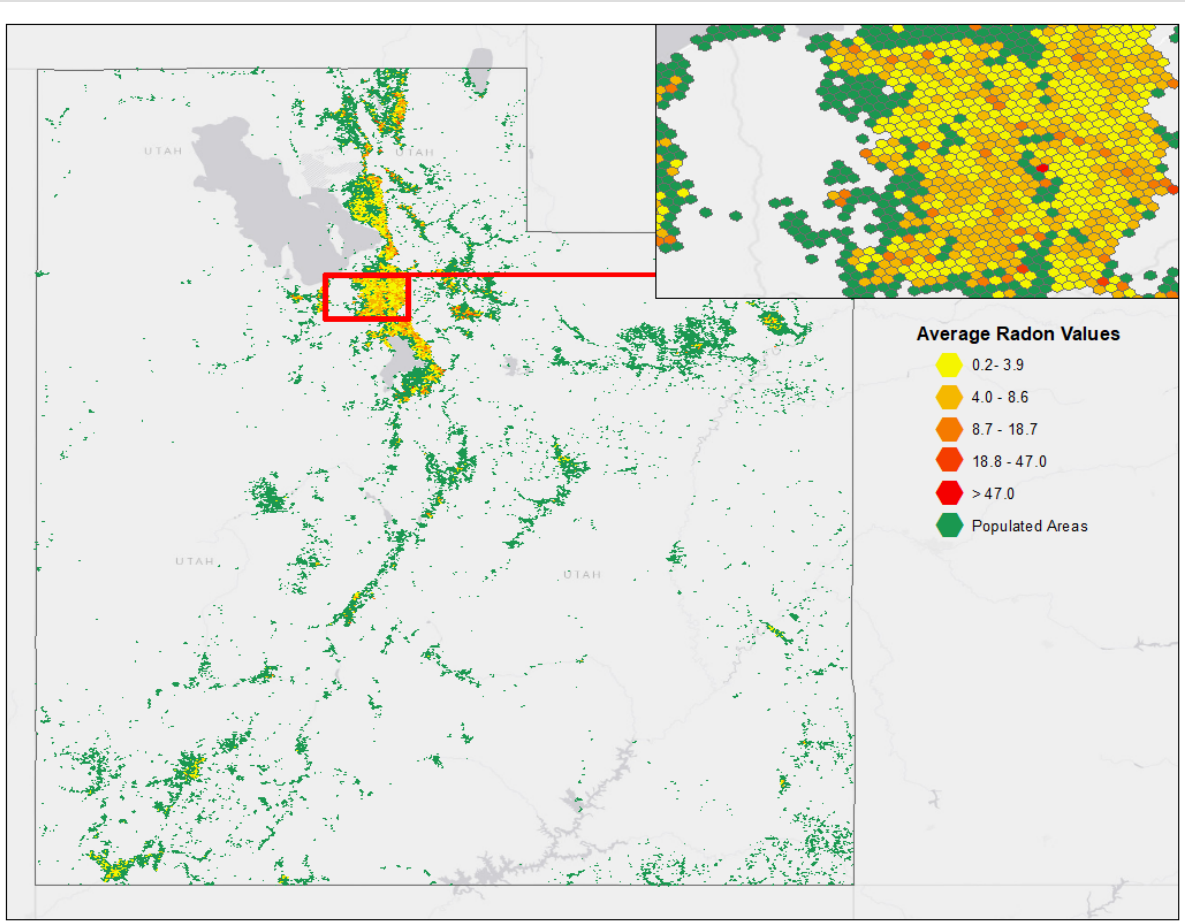
# Measuring indoor radon exposure at the zip code vs. small spatial scales

## Median Radon Levels – Zip Code




## Mean Radon Levels – Uber Hex Res 8

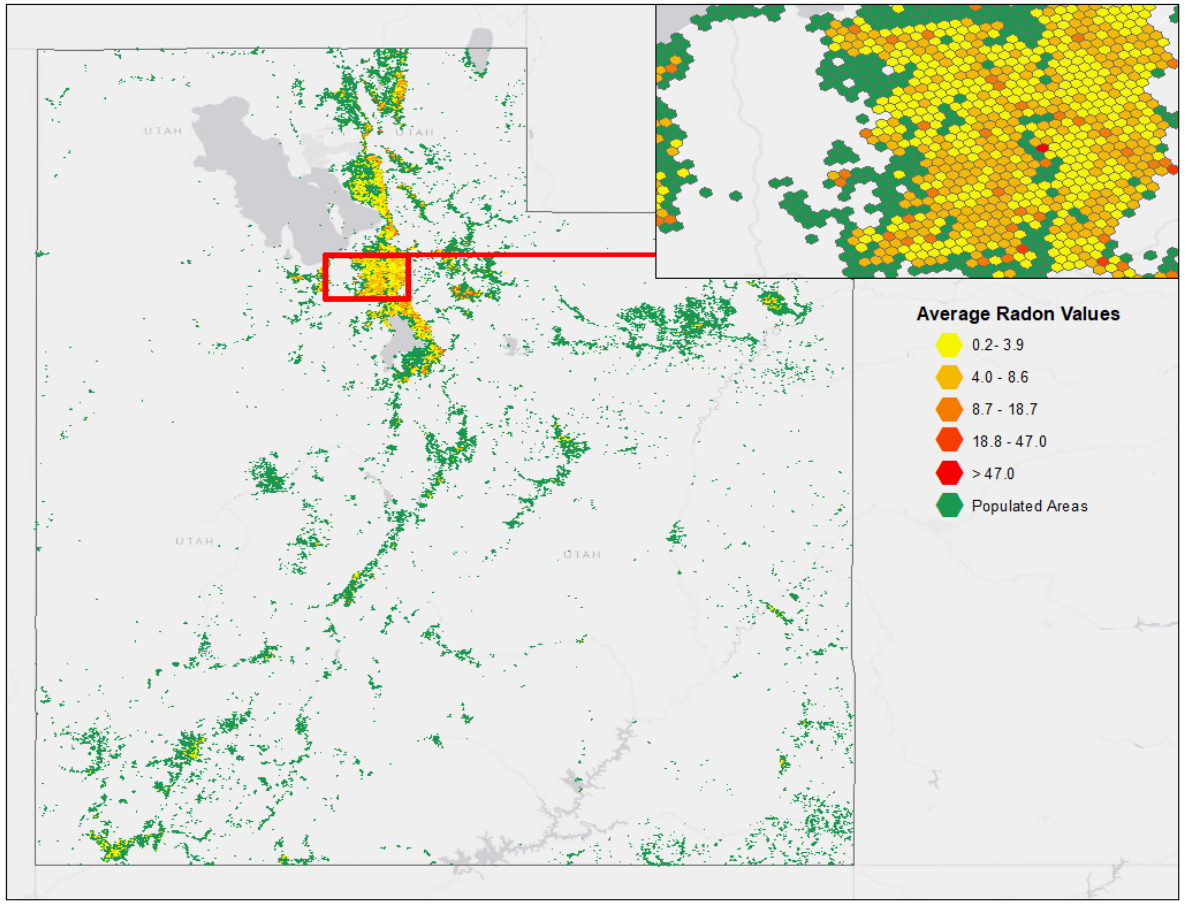
## Utah DEQ



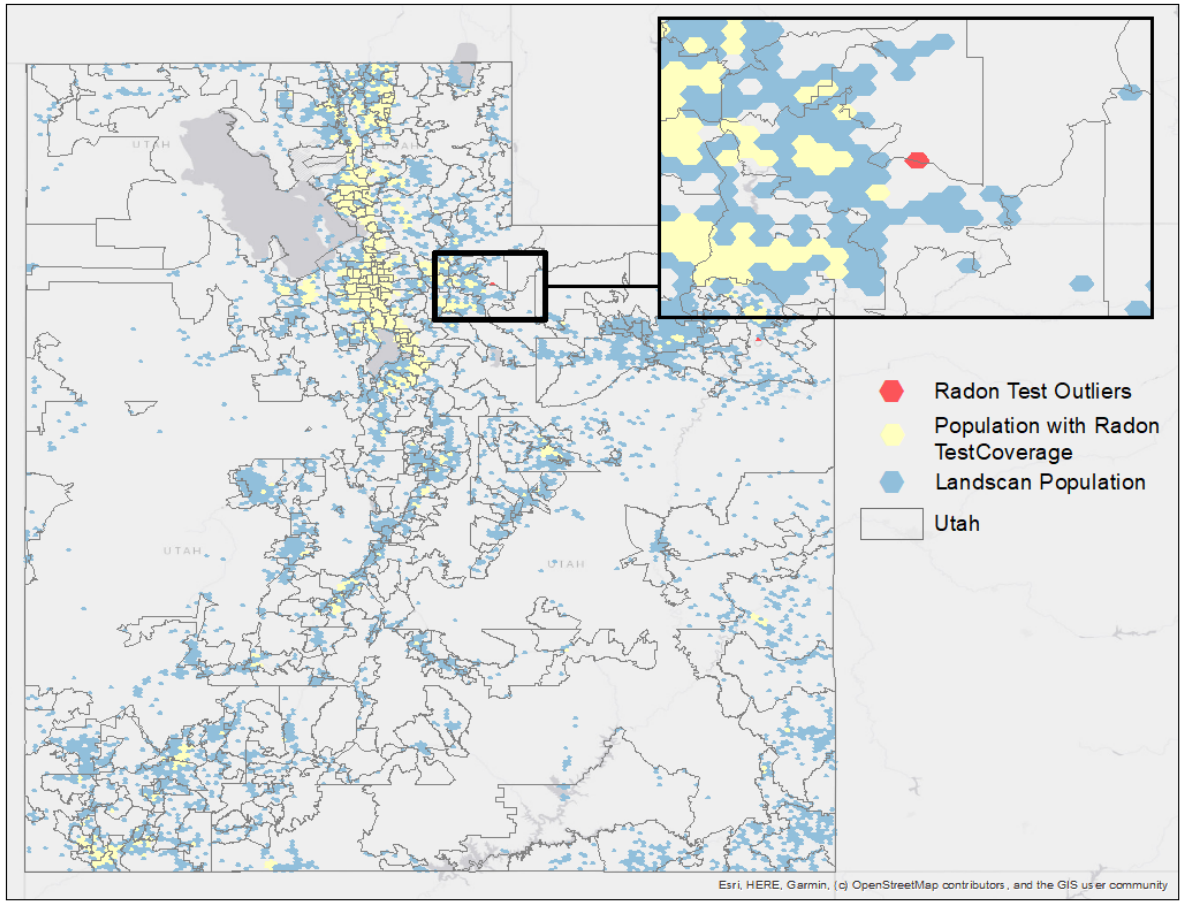

https://epht.health.utah.gov/epht-view/query/result/radon/RadonTestResult/Count.html

# Population Masked Average Indoor Radon Test Results

## Population Masked Average Indoor Radon at Uber H3 Level 8




## Locations Missing Test Results



# Next Steps

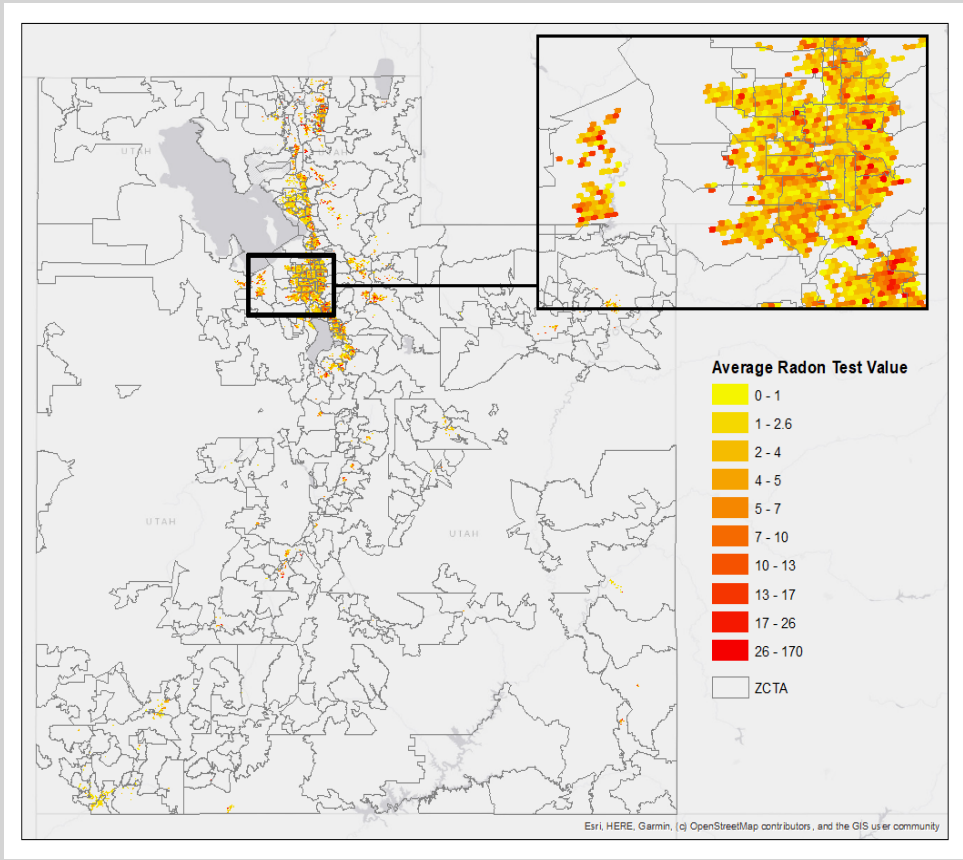
Heterogenous Ensemble Models

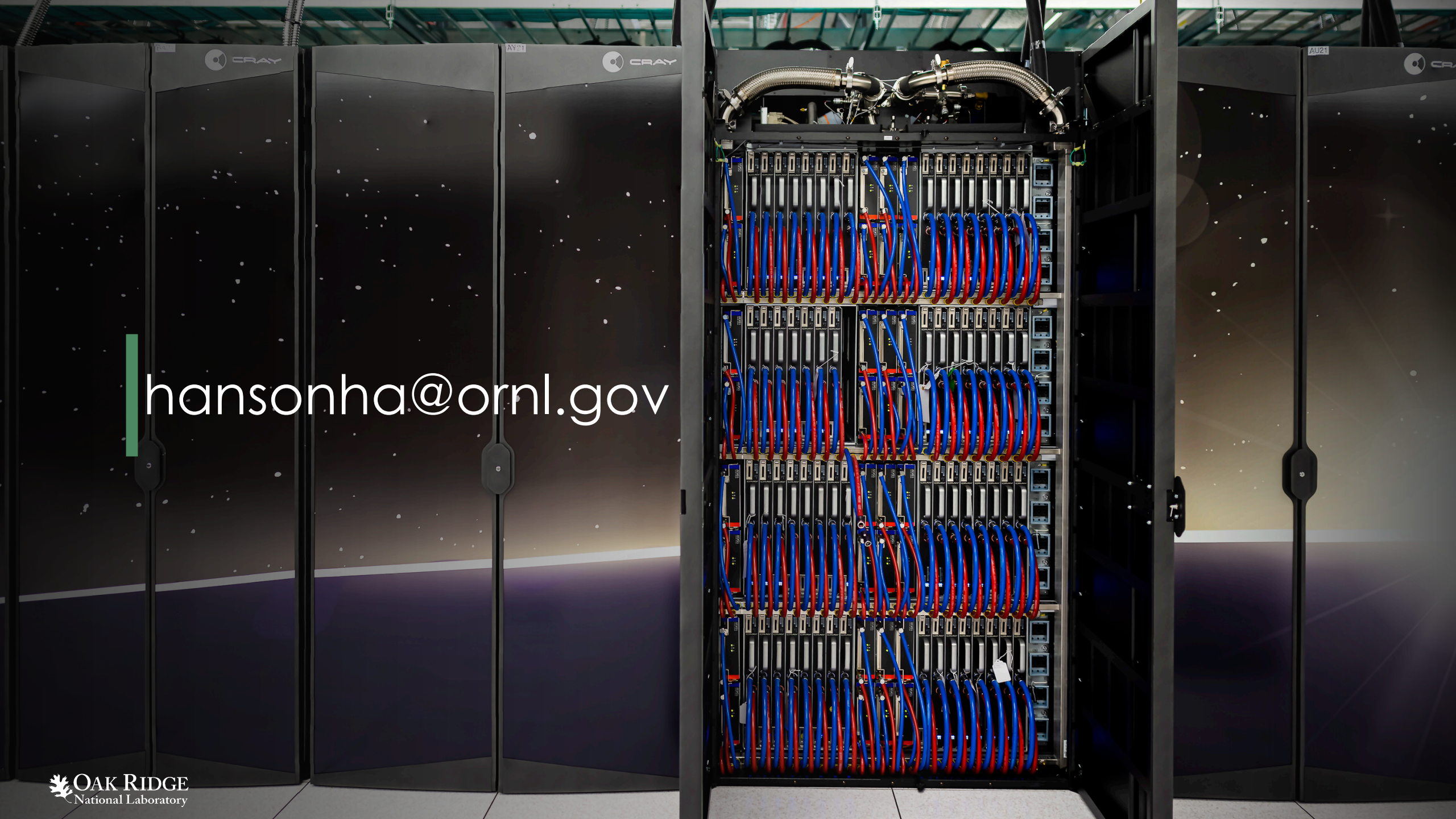
- Geographically and Temporally Weighted Regression methods
  - Local regression method that gives higher weight to the best performing method in a region/space
- Random Forest
- Neural Network
- Stochastic Gradient Boosting Machine
- Generalized Linear Model

Additional Data

- HISDAC-US: Historical Settlement Data
  - Housing characteristics for over 200 years
- **More point level indoor home radon tests results**
  - Point level cancer data: Kentucky, Iowa, Louisiana, Oregon, New Jersey, New Mexico

Utah: Variation in radon test values by zip code





hansonha@ornl.gov