# Norwegian bedrock and drift geology as a predictor of indoor radon concentrations

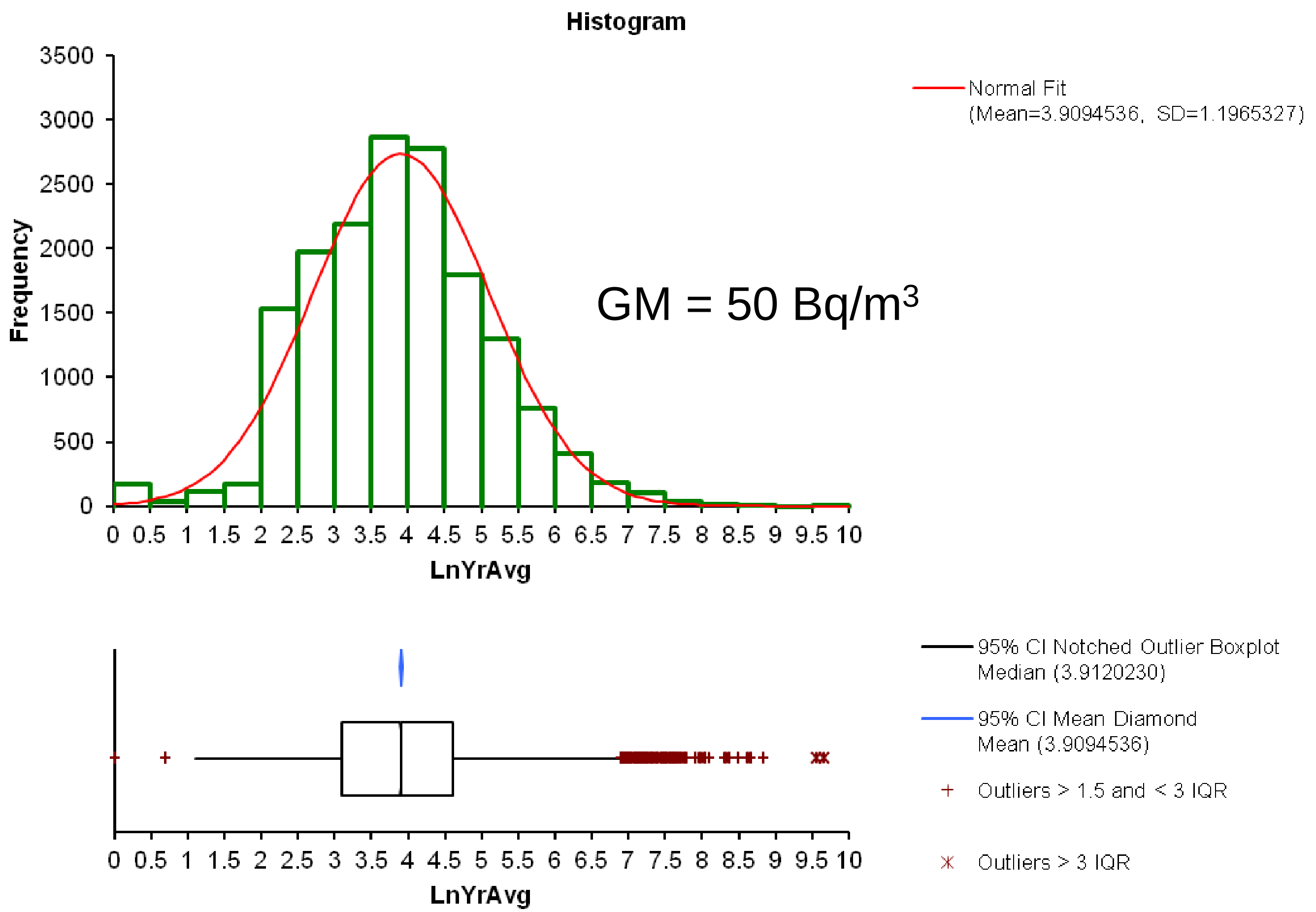
Mark A. Smethurst[1]
Robin J. Watson[2]
Ingvild Finne[3]
Guri Venvik Ganerød[2]
Anne Liv Rudjord[3]

[1] Avalonia Geophysics and University of Exeter, Cornwall Campus, Penryn, UK
[2] Geological Survey of Norway, Trondheim, Norway
[3] Norwegian Radiation Protection Authority, Østerås, Norway

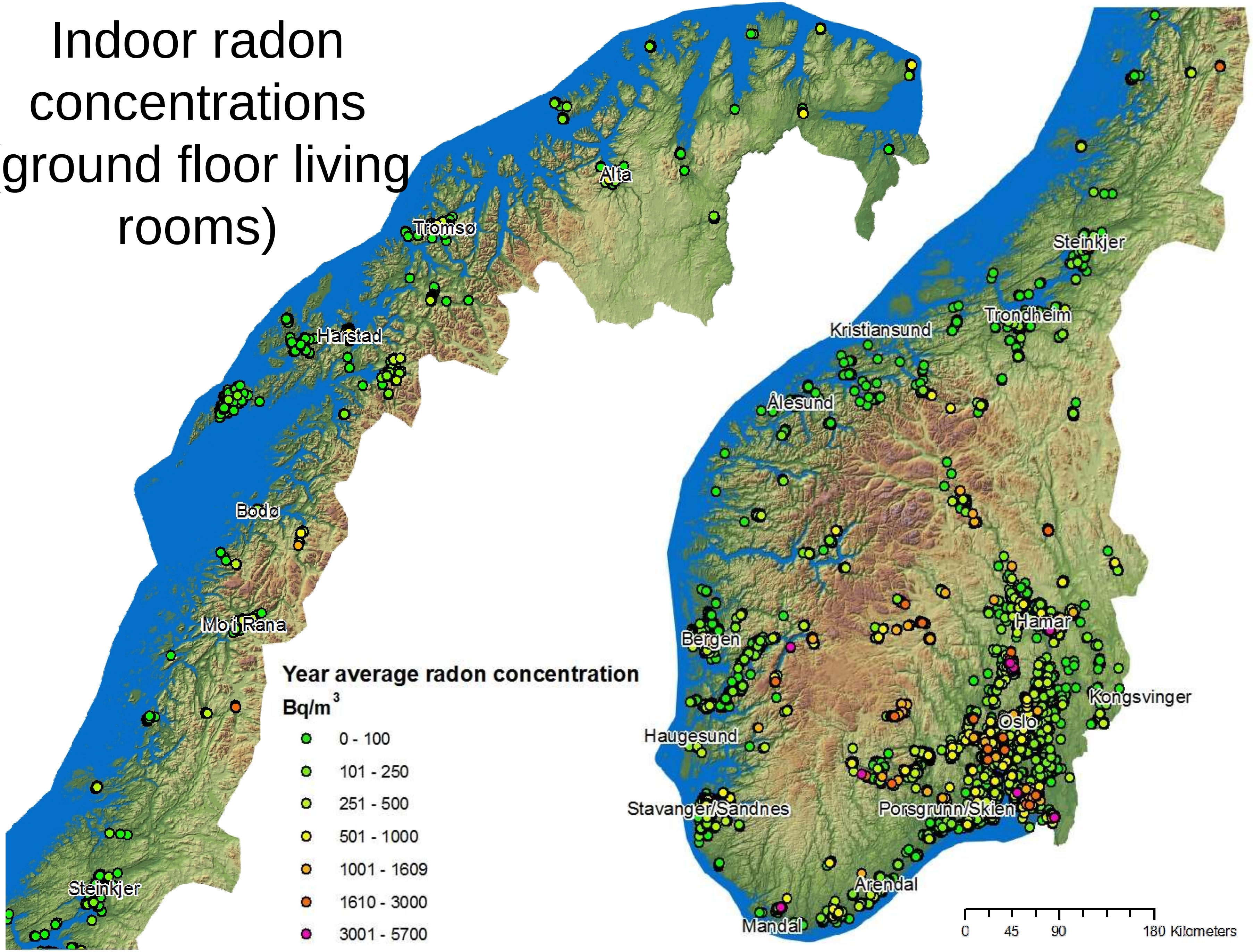
# Ongoing investigation

- Triggered by
  - requirement for maps
    - Simple radon susceptibility maps
    - Objectively derived confidence limits on estimates of radon susceptibility
  - Substantial increase in the availability of geo-referenced indoor radon measurements
- Founded on
  - Indoor radon concentrations held by Norwegian Radiation Protection Authority
  - Bedrock and drift geology held by Geological Survey of Norway

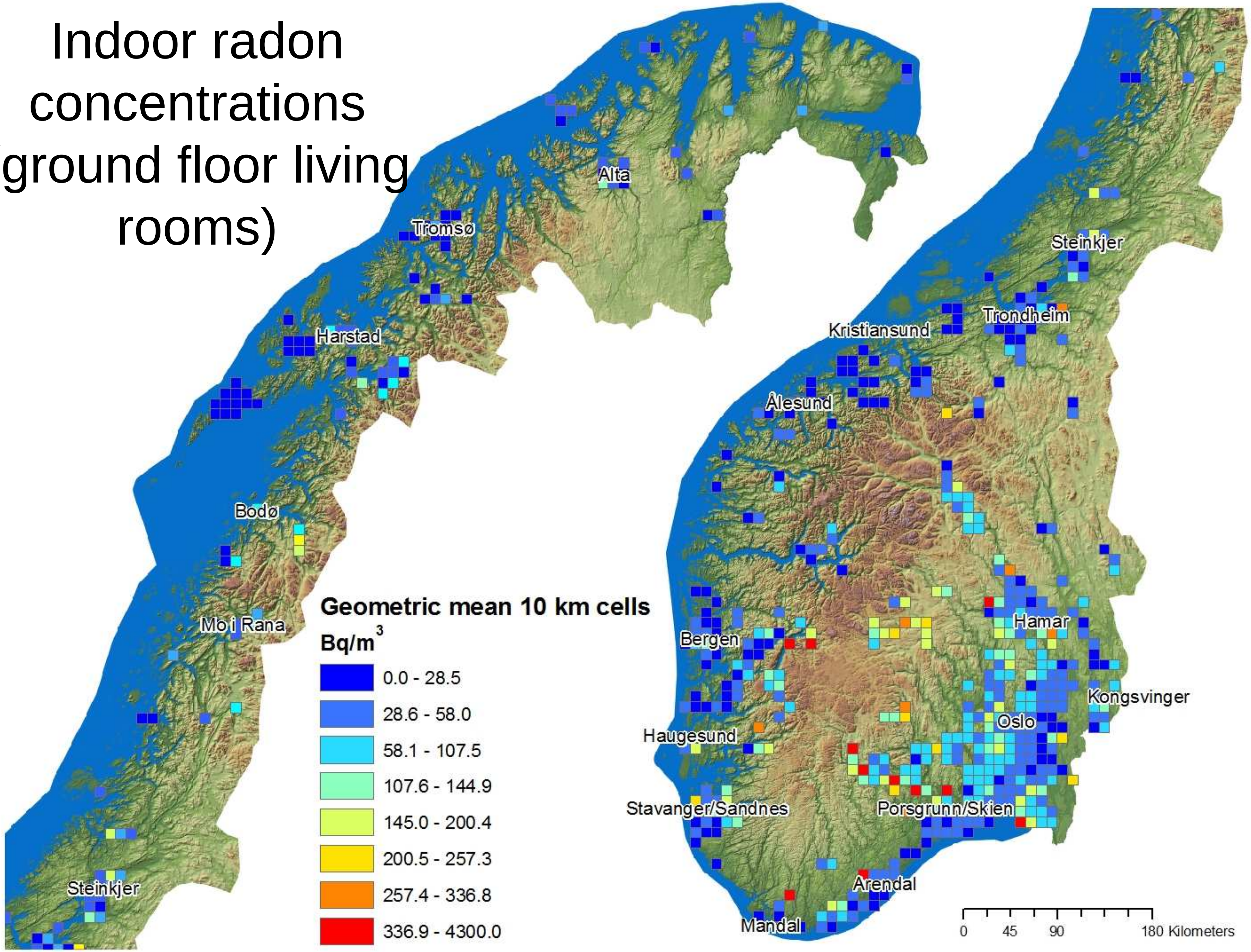
# Year average indoor radon concentrations



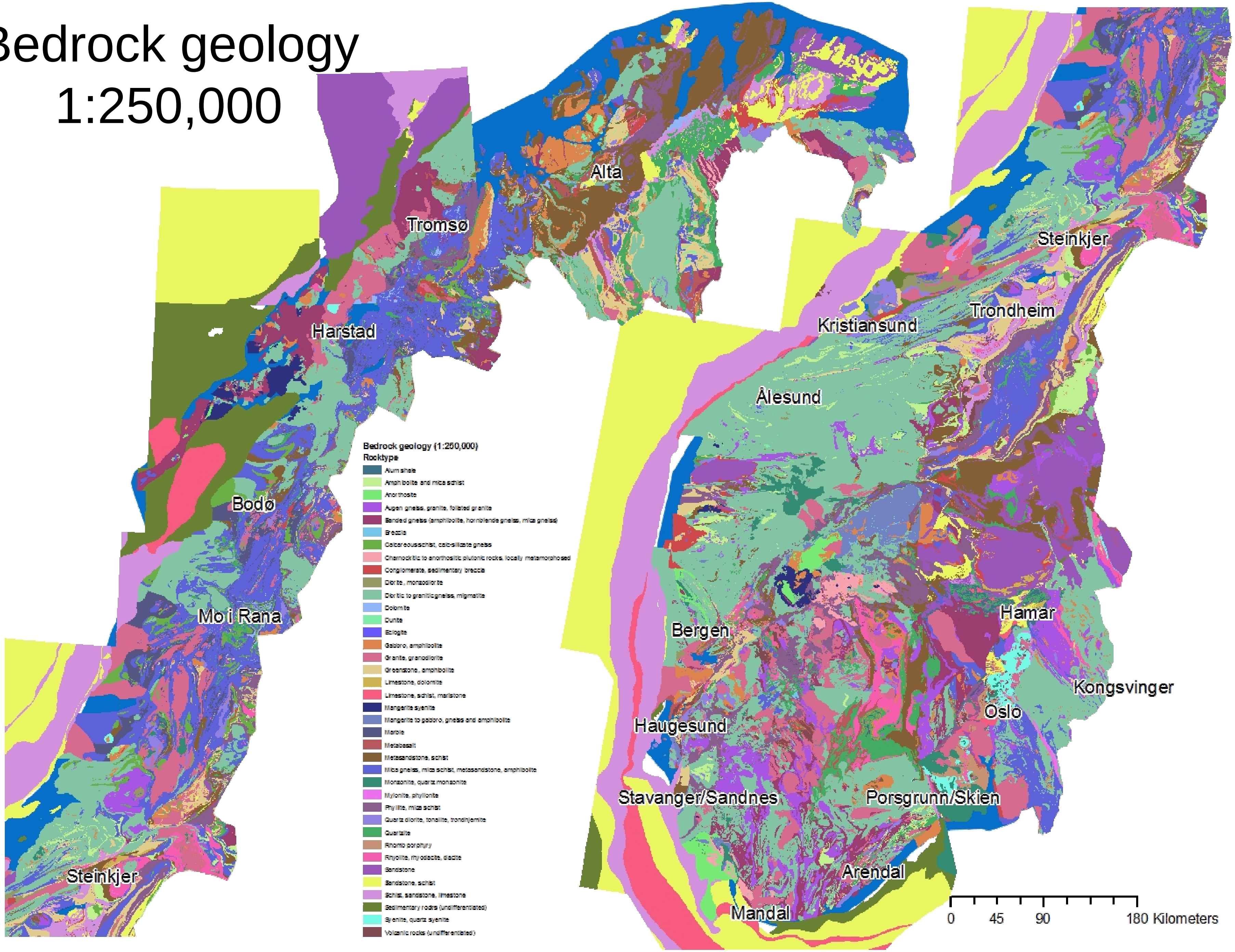
# Indoor radon concentrations (ground floor living rooms)



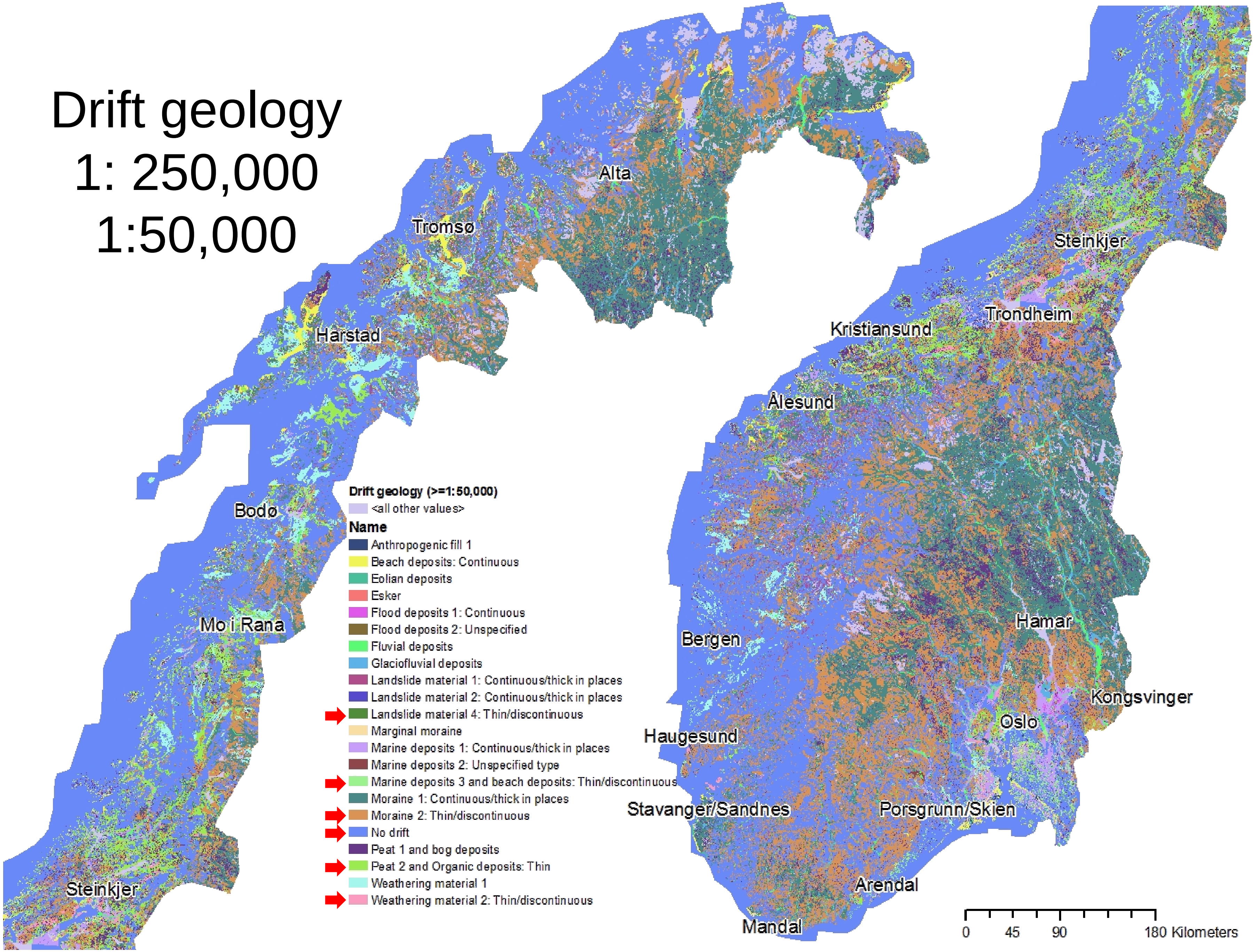
# Indoor radon concentrations (ground floor living rooms)



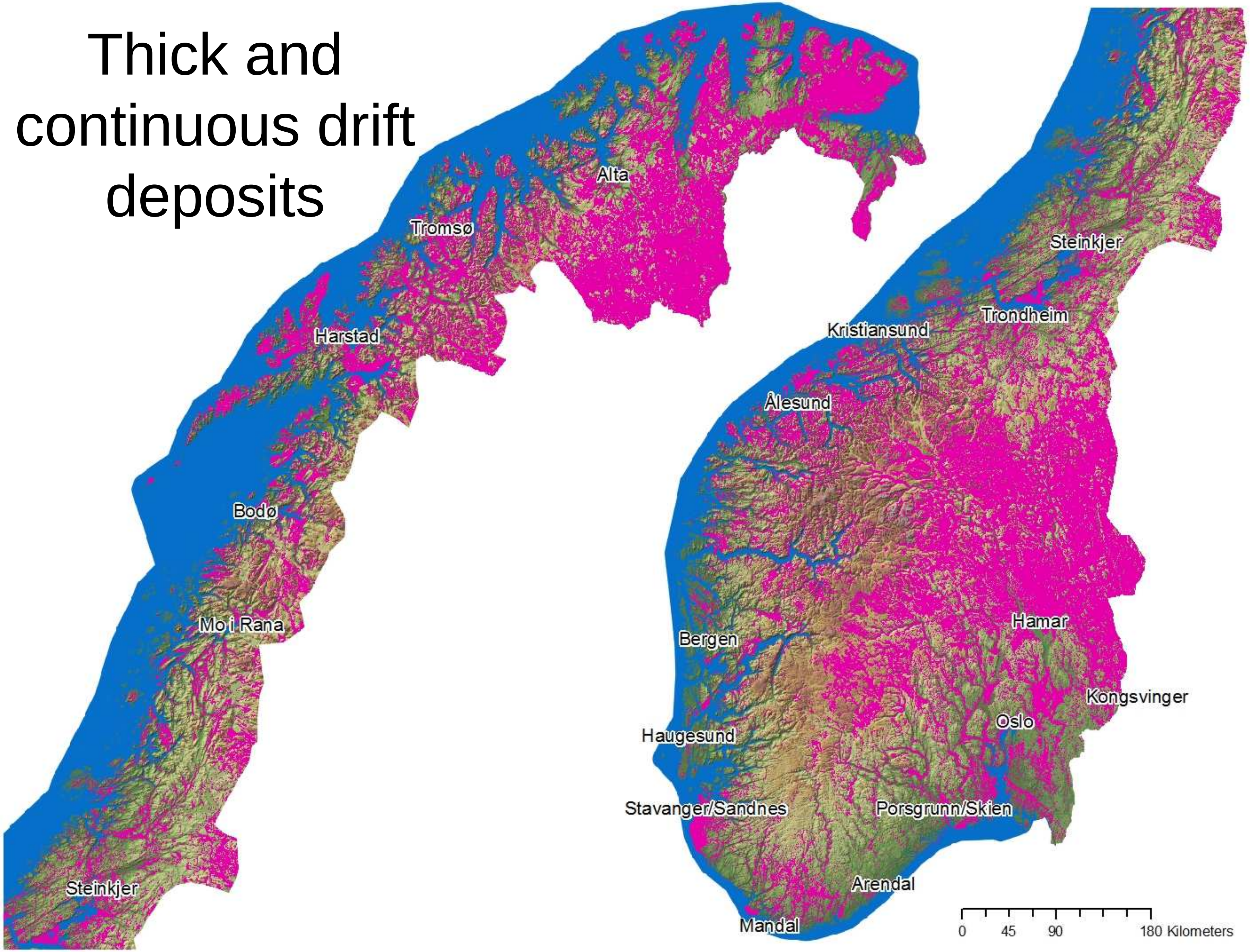
# Bedrock geology 1:250,000



# Drift geology
# 1: 250,000
# 1:50,000



# Thick and continuous drift deposits



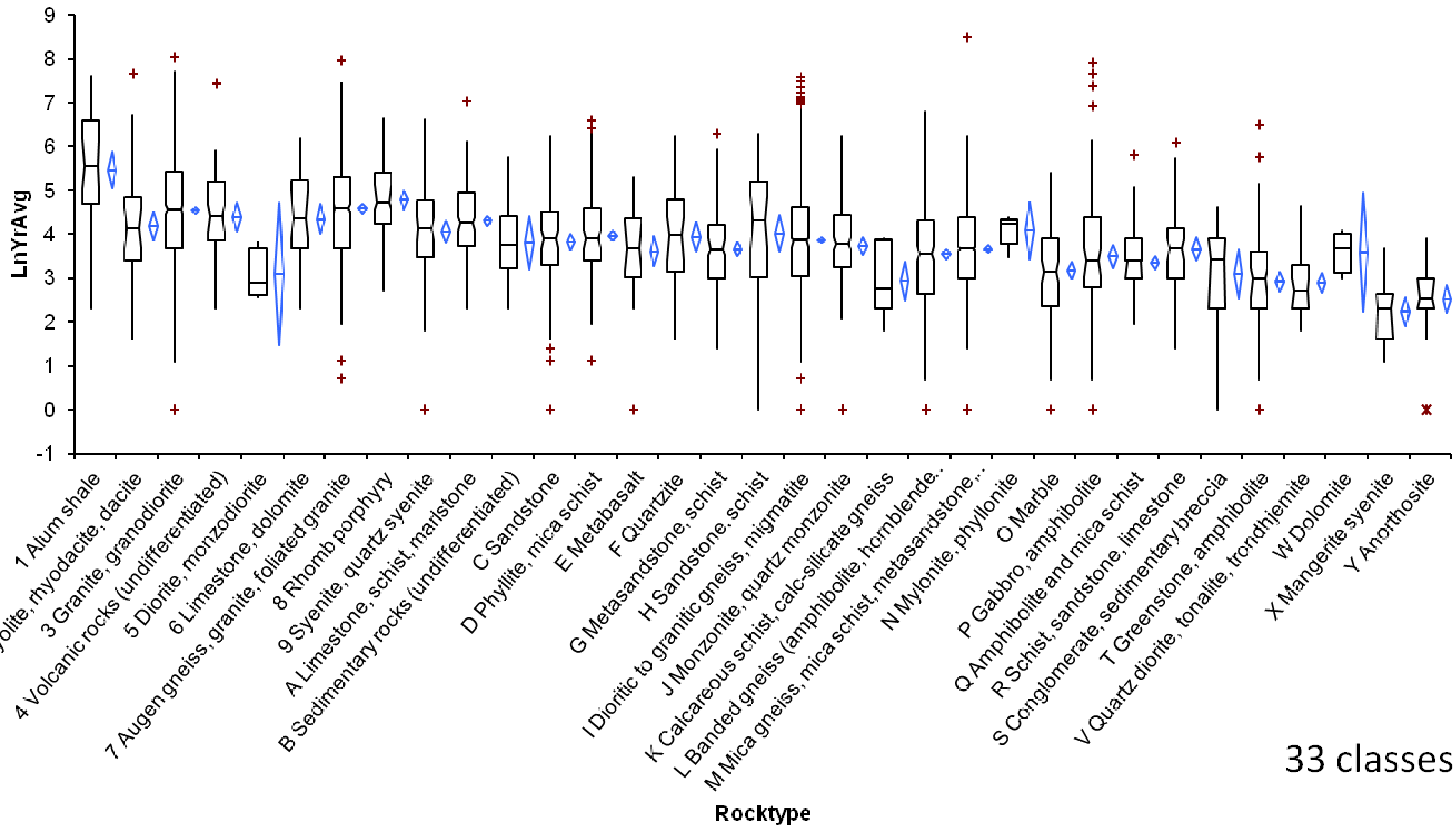
# Analytical approach

- Characterise bedrock, drift, and combinations of the two in terms of available indoor radon measurements
- Examine the strengths of the relationships between geological classes and indoor radon
- Use the relationships to produce radon susceptibility estimates where indoor data are few/absent

# Methods

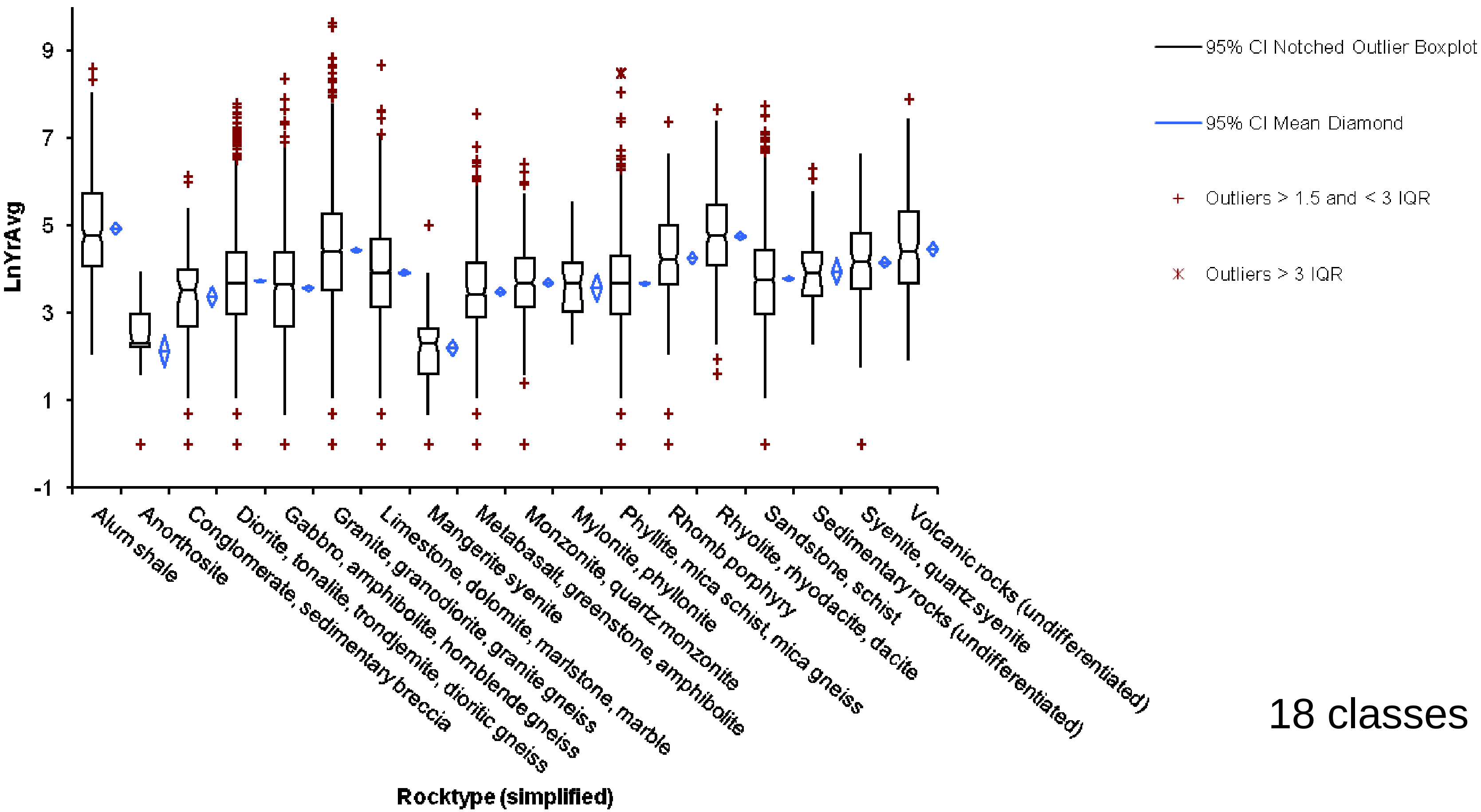
- Use 1-way and 2-way ANOVA to examine the influence of mapped geology on the variance in radon measurements
- Construct a naive Bayes classifier to identify radon prone areas from geological input

# Radon concentrations on bedrock types (no drift)



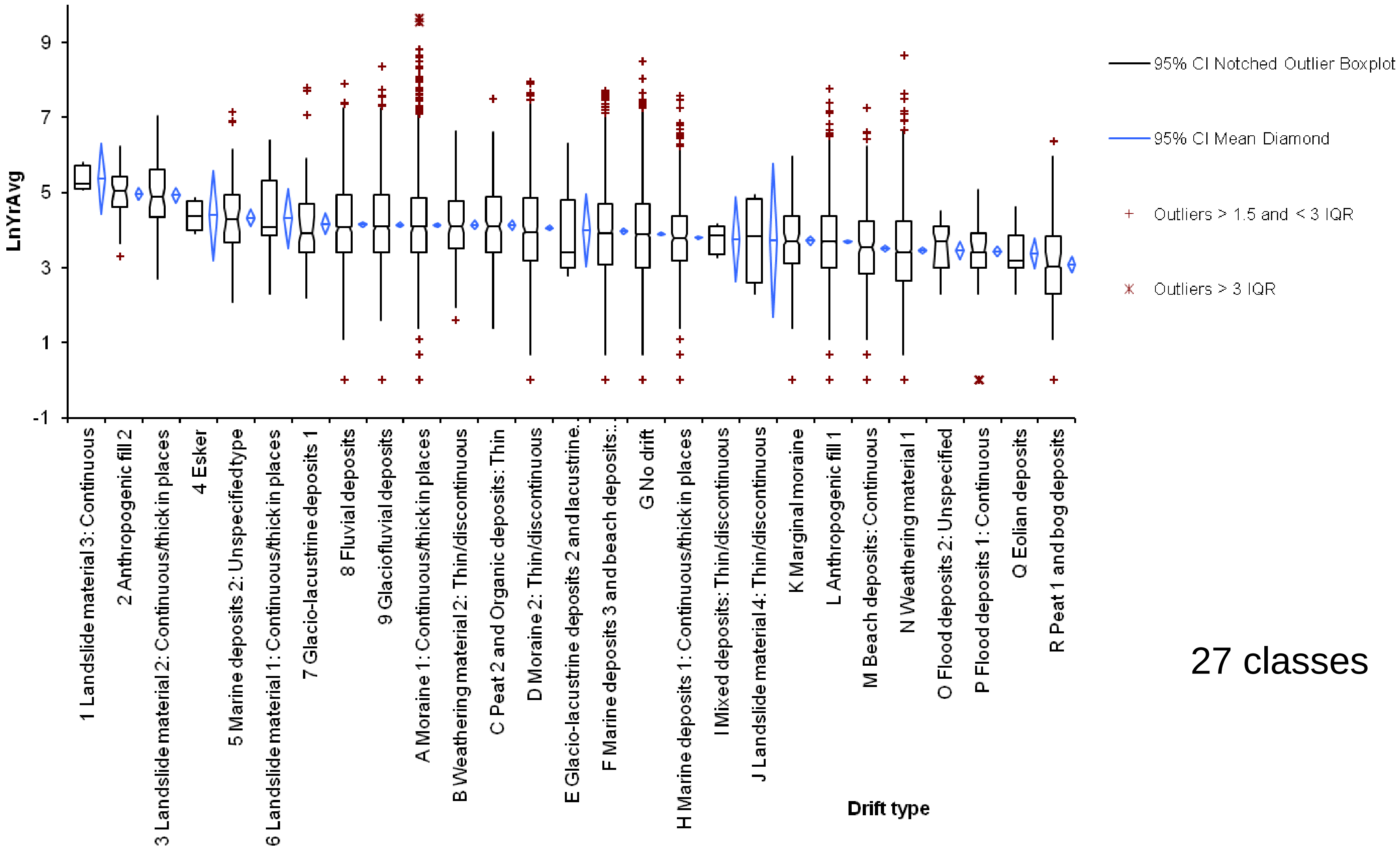

33 classes

# Radon concentrations on simplified bedrock types (no drift)



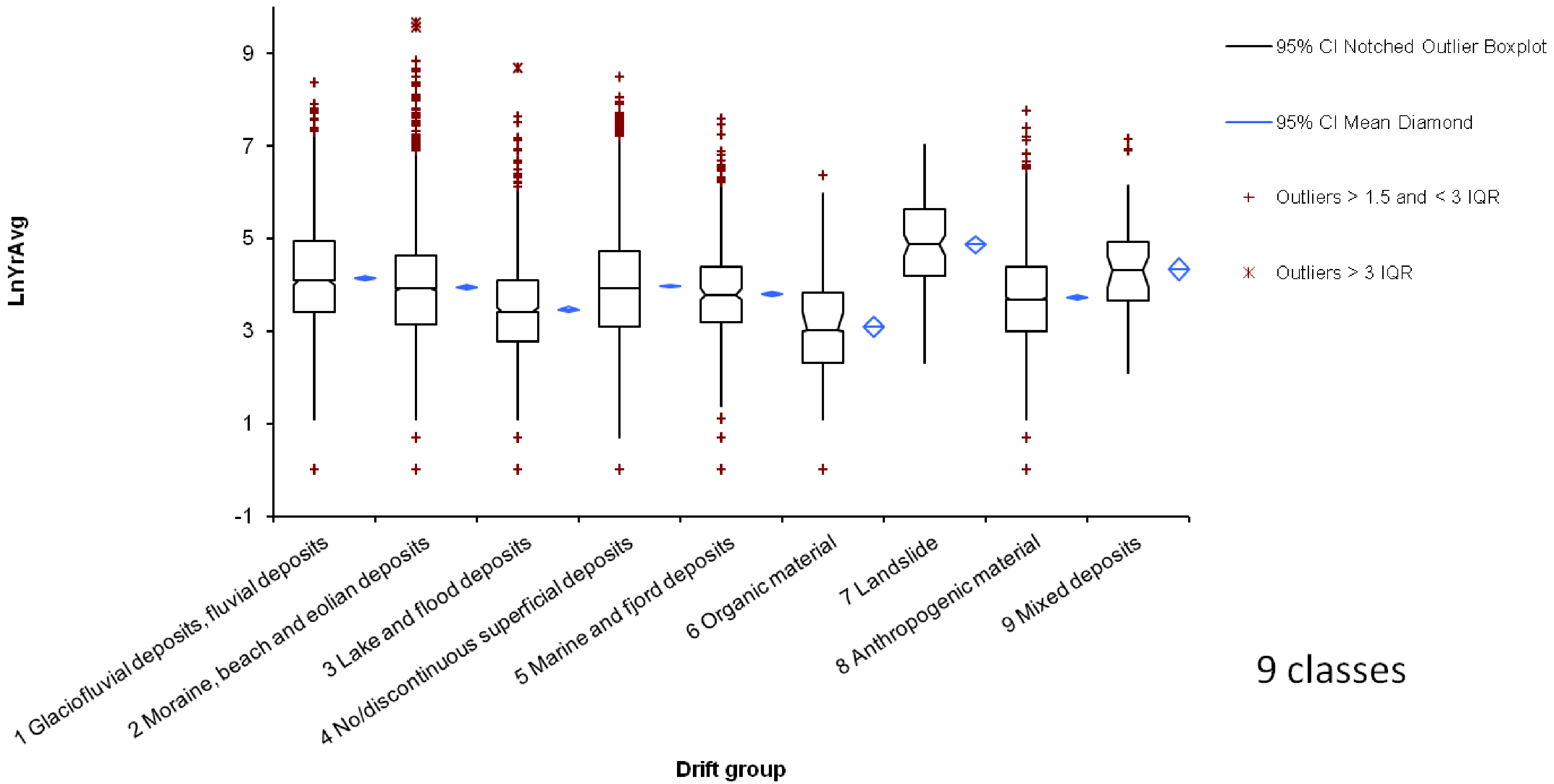

18 classes

# Radon concentrations on drift types




27 classes

# Radon concentrations on simplified drift types




9 classes

# 1-way ANOVA results

*What portion of the observed variance in radon concentrations can be attributed to different sources of variation?*

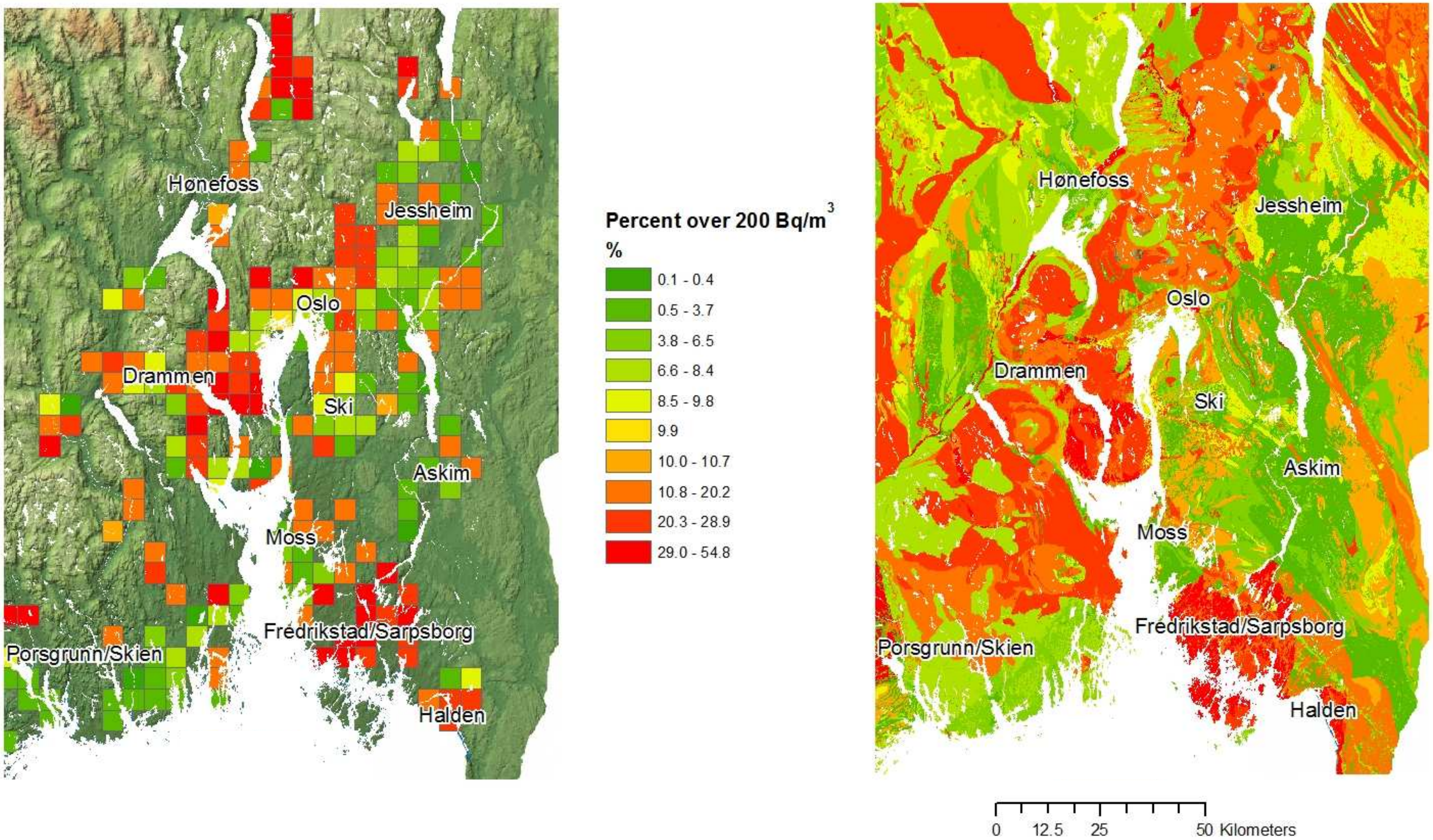
| **Source of variation** | Drift variety: Continuous | Drift variety: Thin/none | Classes | Rn meas. | Meas./class | **Portion %** |
|---|---|---|---|---|---|---|
| Rock type and drift combinations | ✓ | | 164 | 10669 | 65 | 18.8 |
| Rock type and drift combinations | ✓ | ✓ | 198 | 16406 | 83 | 17.6 |
| Generalised rock type and drift combinations | ✓ | | 86 | 10661 | 124 | 16.0 |
| Rock type | | ✓ | 33 | 5734 | 174 | 15.1 |
| Generalised rock type and drift combinations | ✓ | ✓ | 104 | 16395 | 158 | 14.9 |
| Generalised rock type | | ✓ | 18 | 5734 | 319 | 12.9 |
| Generalised rock type and drift combinations | | ✓ | 18 | 5734 | 319 | 12.9 |
| Rock type | ✓ | ✓ | 34 | 16413 | 483 | 12.4 |
| Generalised rock type | ✓ | ✓ | 18 | 16413 | 912 | 11.1 |
| Drift drift | ✓ | | 20 | 10679 | 534 | 6.4 |
| Drift drift | ✓ | ✓ | 27 | 16415 | 608 | 4.3 |
| Generalised drift | ✓ | | 8 | 10681 | 1335 | 4.2 |
| Generalised drift | ✓ | ✓ | 9 | 16418 | 1824 | 2.8 |

Increasing portion of the observed variance

# Learned from 1-way ANOVA

- Strength of relation to radon increases: 1 drift alone, 2 bedrock alone, 3 bedrock and drift combinations
- More detailed drift and bedrock classifications explain more of the variation in radon values
- Ignoring thin/discontinuous drift improves the explaining power of all classifications
- Bedrock-only classifications perform best in areas of thin and discontinuous drift
- Interaction between bedrock and thick drift classes means that the combined classifications function better

# Percentage > 200 Bq/m³



# 2-way ANOVA results

| n | 1120 |
|---|---|

| Year Average by Simplified BEDROCK type | n | Mean | SE | Pooled SE | SD |
|---|---|---|---|---|---|
| 4 | 160 | 3.99148 | 0.091683 | 0.086207 | 1.15971 |
| 5 | 160 | 3.58207 | 0.092410 | 0.086207 | 1.16891 |
| 6 | 160 | 3.75715 | 0.109766 | 0.086207 | 1.38844 |
| 7 | 160 | 3.83138 | 0.077329 | 0.086207 | 0.97815 |
| 9 | 160 | 3.84320 | 0.089164 | 0.086207 | 1.12785 |
| 12 | 160 | 3.75266 | 0.072335 | 0.086207 | 0.91497 |
| 15 | 160 | 3.89490 | 0.092424 | 0.086207 | 1.16908 |

| Year Average by Simplified DRIFT type | n | Mean | SE | Pooled SE | SD |
|---|---|---|---|---|---|
| 1 | 280 | 3.91123 | 0.065669 | 0.065167 | 1.09886 |
| 2 | 280 | 3.62779 | 0.068429 | 0.065167 | 1.14504 |
| 5 | 280 | 3.85876 | 0.065158 | 0.065167 | 1.09029 |
| 8 | 280 | 3.83242 | 0.072538 | 0.065167 | 1.21379 |

| Source of variation | Sum squares | DF | Mean square | F statistic | p | S |
|---|---|---|---|---|---|---|
| BEDROCK | 15.95113 | 6 | 2.65852 | 2.24 | 0.0377 | 1.1% |
| D%RIFT | 12.96513 | 3 | 4.32171 | 3.63 | 0.0125 | 0.9% |
| BEDROCK x DRIFT | 130.98256 | 18 | 7.27681 | 6.12 | <0.0001 | 9.0% |
| Residual | 1298.46275 | 1092 | 1.18907 | | | |
| Total | 1458.36157 | 1119 | | | | 11.0% |

# Learned from 2-way ANOVA

- The presence of different drift types change the relationships between bedrock types and indoor radon
- More work is required to identify the principal interactions with confidence
- If true, classifiers based on independent bedrock and drift factors could be flawed

# Naive Bayes Classifier

- 3753 instances (geological units)
- 2 categorical features: Bedrock (37 categories), Drift (9 categories)
- 2 output classes:
  - Based on GM of indoor radon in geological unit
  - Percent over 200 Bq/m3
  - LOW (0-20 %)
  - HIGH (> 20%)

Naive Bayes Classifier

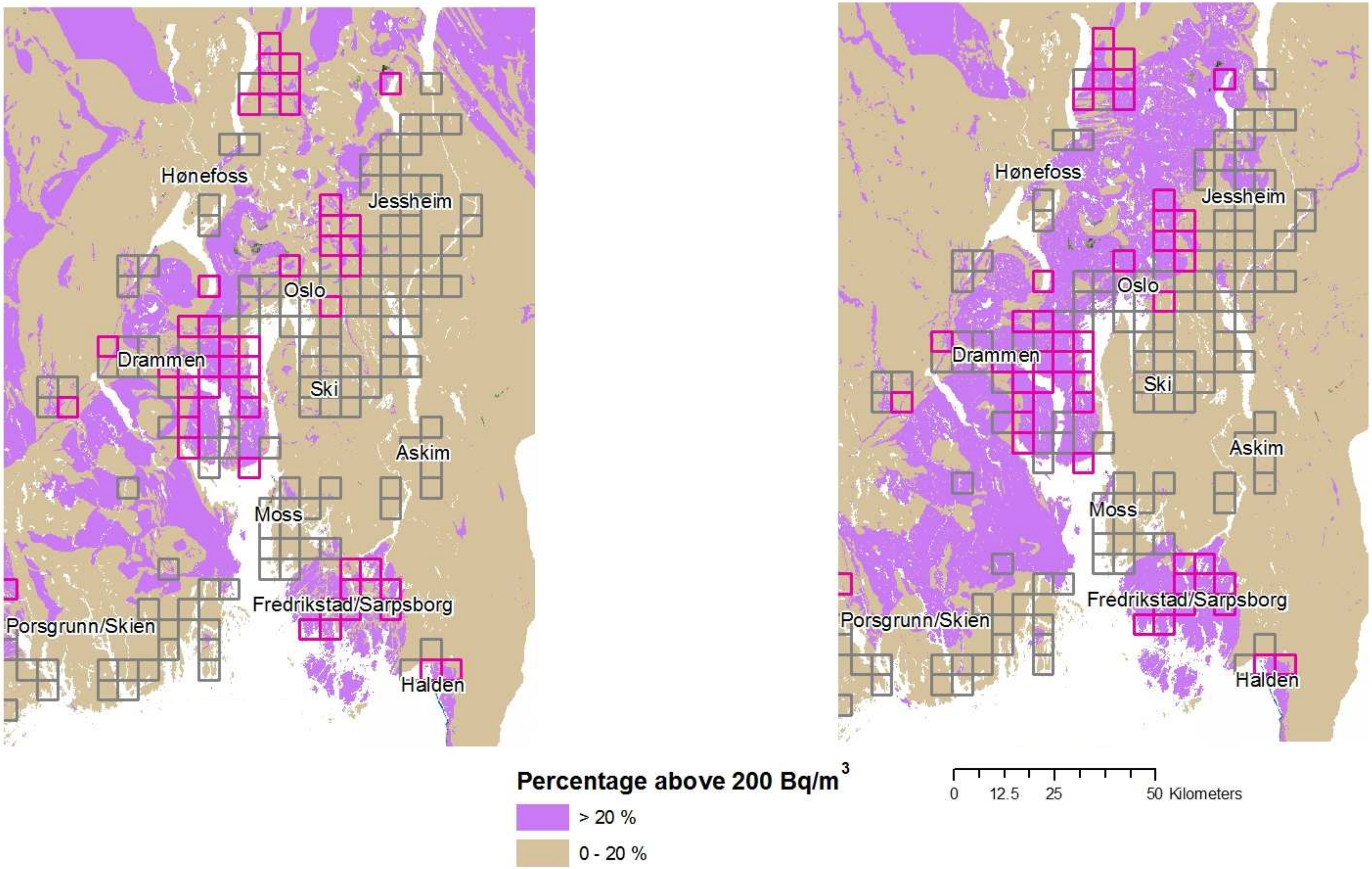
Using equal prior probabilities and 5-fold cross-validation:

| | | Predicted class | | | |
|---|---|---|---|---|---|
| Actual class | | HIGH | LOW | Total | Sensitivity |
| | HIGH | 349 | 243 | 592 | 0.59 |
| | LOW | 889 | 2272 | 3161 | 0.71 |
| | Total | 1238 | 2515 | 3753 | |
| | Predictive value | 0.28 | 0.90 | | |

| | |
|---|---|
| Classifier accuracy CA | 0.70 |
| Matthewsd Correlation Coefficient MCC | 0.24 |

ANOIVA/direct assignment

Bayes classifier



# Testing geological classifications using ANOVA and "Tukey Contrast"

Simplified Bedrock - Drift absent – 9.5 % of variation explained

| | |
|---|---|
| n | 3715 |

| LnYrAv by BEDCODE | n | Mean | SE | Pooled SE | SD |
|---|---|---|---|---|---|
| BAS | 41 | 5.573 | 0.2046 | 0.1898 | 1.310 |
| BG | 1688 | 3.816 | 0.0287 | 0.0296 | 1.177 |
| BGR | 926 | 4.707 | 0.0423 | 0.0399 | 1.287 |
| BML | 243 | 3.939 | 0.1162 | 0.0779 | 1.812 |
| BS | 689 | 4.029 | 0.0368 | 0.0463 | 0.966 |
| BST | 128 | 4.277 | 0.0783 | 0.1074 | 0.886 |

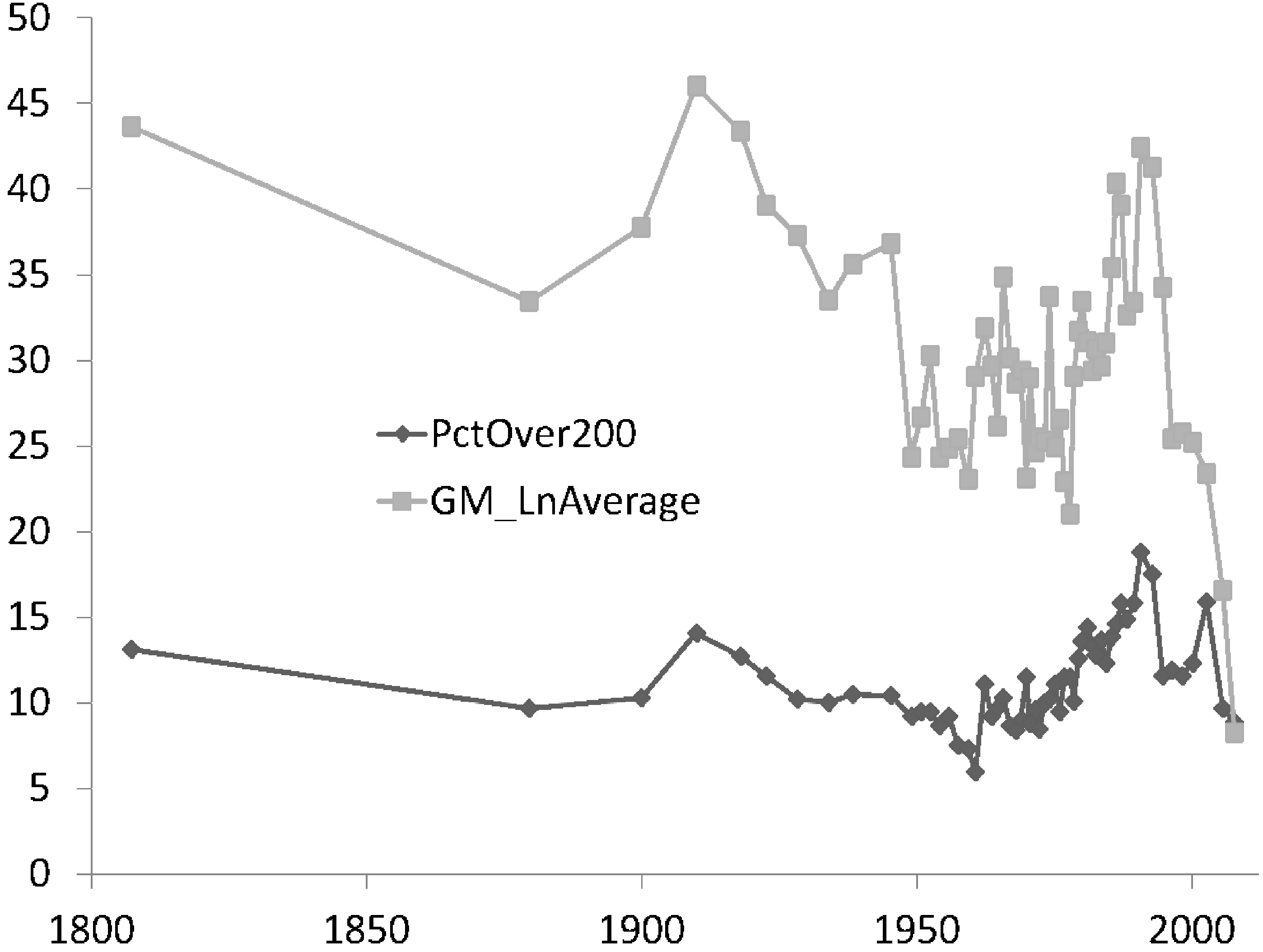
| Source of variation | Sum squares | DF | Mean square | F statistic | p |
|---|---|---|---|---|---|
| BEDCODE | 578.717 | 5 | 115.743 | 78.39 | <0.0001 |
| Residual | 5476.312 | 3709 | 1.476 | | |
| Total | 6055.029 | 3714 | | | |

| Tukey Contrast | Difference | 95% CI | |
|---|---|---|---|
| BAS v BG | 1.757 | 1.209to 2.304 | (significant) |
| BAS v BGR | 0.866 | 0.313to 1.418 | (significant) |
| BAS v BML | 1.634 | 1.049to 2.219 | (significant) |
| BAS v BS | 1.543 | 0.987to 2.100 | (significant) |
| BAS v BST | 1.296 | 0.675to 1.918 | (significant) |
| BG v BGR | -0.891 | -1.033to -0.750 | (significant) |
| BG v BML | -0.123 | -0.360to 0.115 | |
| BG v BS | -0.213 | -0.370to -0.057 | (significant) |
| BG v BST | -0.461 | -0.778to -0.143 | (significant) |
| BGR v BML | 0.769 | 0.519to 1.018 | (significant) |
| BGR v BS | 0.678 | 0.504to 0.852 | (significant) |
| BGR v BST | 0.431 | 0.104to 0.757 | (significant) |
| BML v BS | -0.091 | -0.349to 0.168 | |
| BML v BST | -0.338 | -0.716to 0.040 | |
| BS v BST | -0.247 | -0.581to 0.086 | |

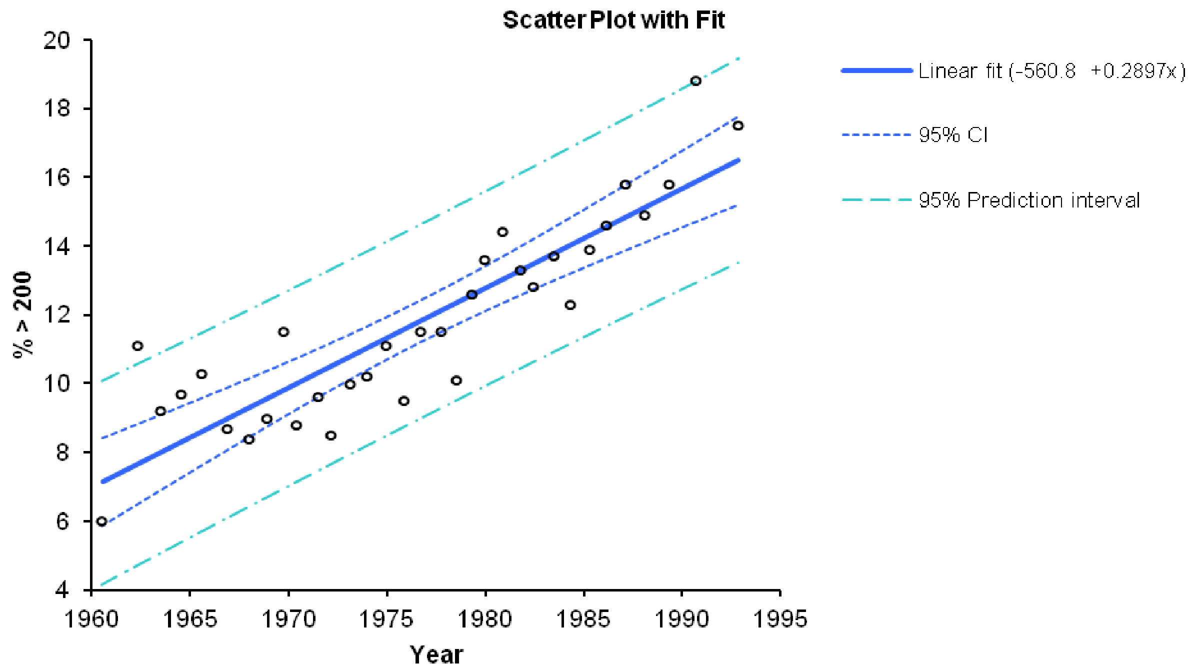
**Alum Shale > Granite > [All olthers] > Gneiss**

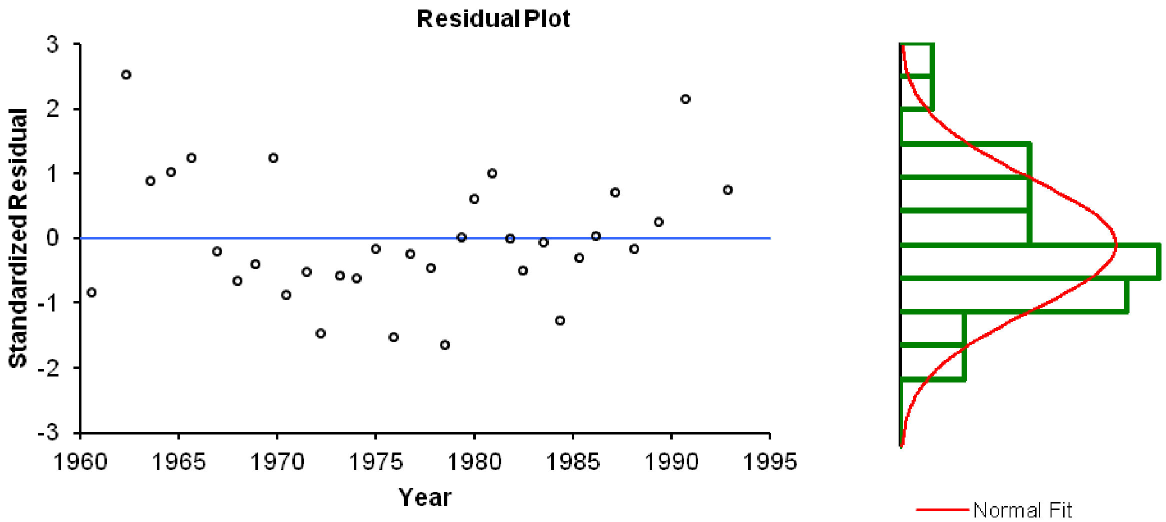
What is responsible for the other 80 % of the observed variance?

# Construction Year vs. Geometric Mean and % > 200 (1 point = 1000 dwellings)



# Linear Regression



# Learned from analysis of dwelling properties

Partly explains why the geological classes explain a relatively small portion of the total variance in radon measurements

What is the relative importance of geology and dwelling properties?

# Naive Bayes Classifier

- Input features $f_i$ (Bedrock category, Drift category)
- Output classes $c_j$ (High, Low)

Determine class probabilites:

$$p_j = \Pi_i P(f_i \mid c_j) * P(c_j)$$

Conditional probabilities

Prior probabilities

Assign instance to class with highest probability