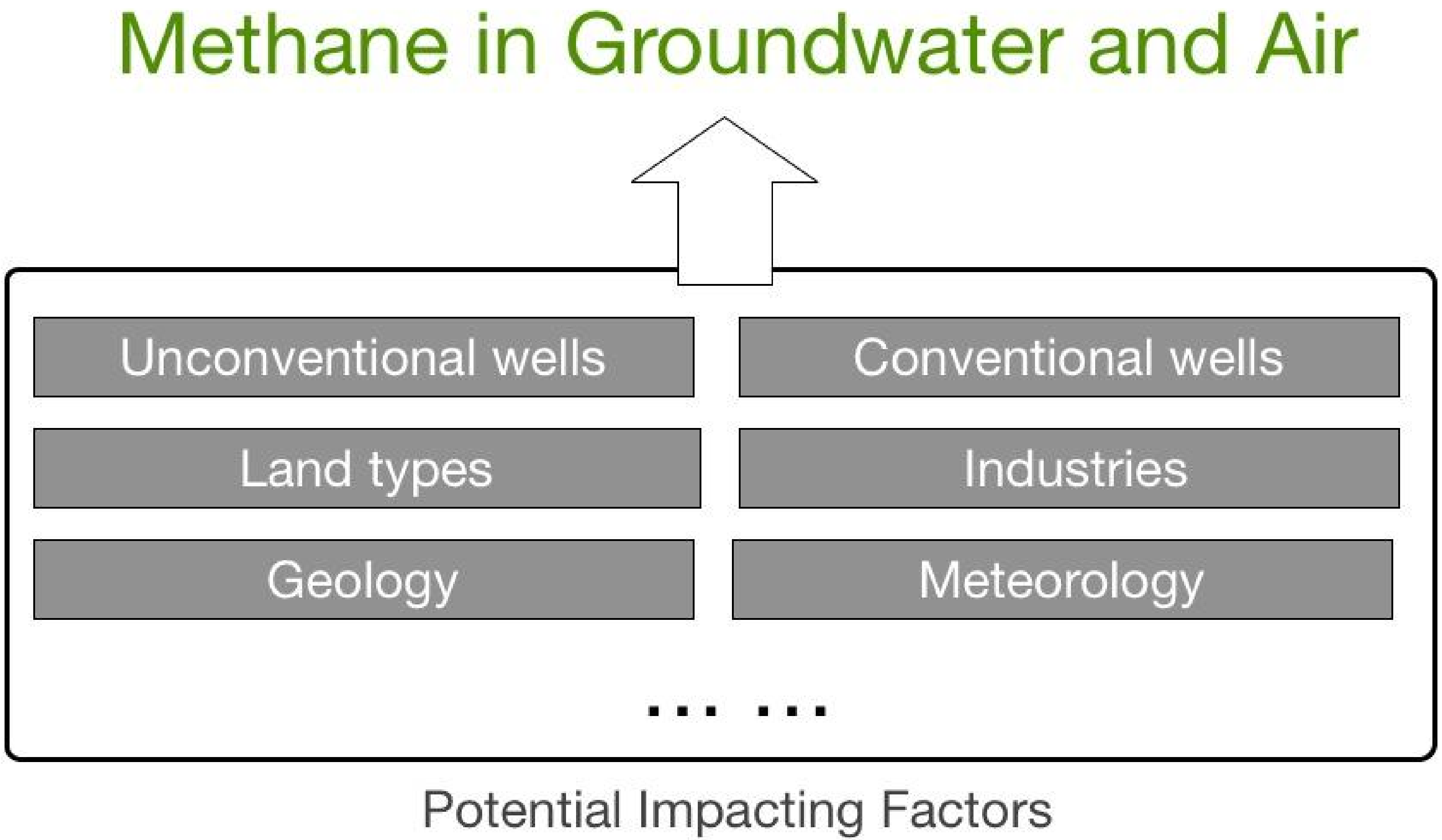
A data-driven approach to evaluate environmental impacts of shale-gas drilling

Jessie Li[1], Guanjie Zheng[1], Fei Wu[1],
Anna K. Wendt[2], Matt S. Gonzales[2], Susan L. Brantley[2]

[1]College of Information Sciences and Technology
[2]Department of Geosciences

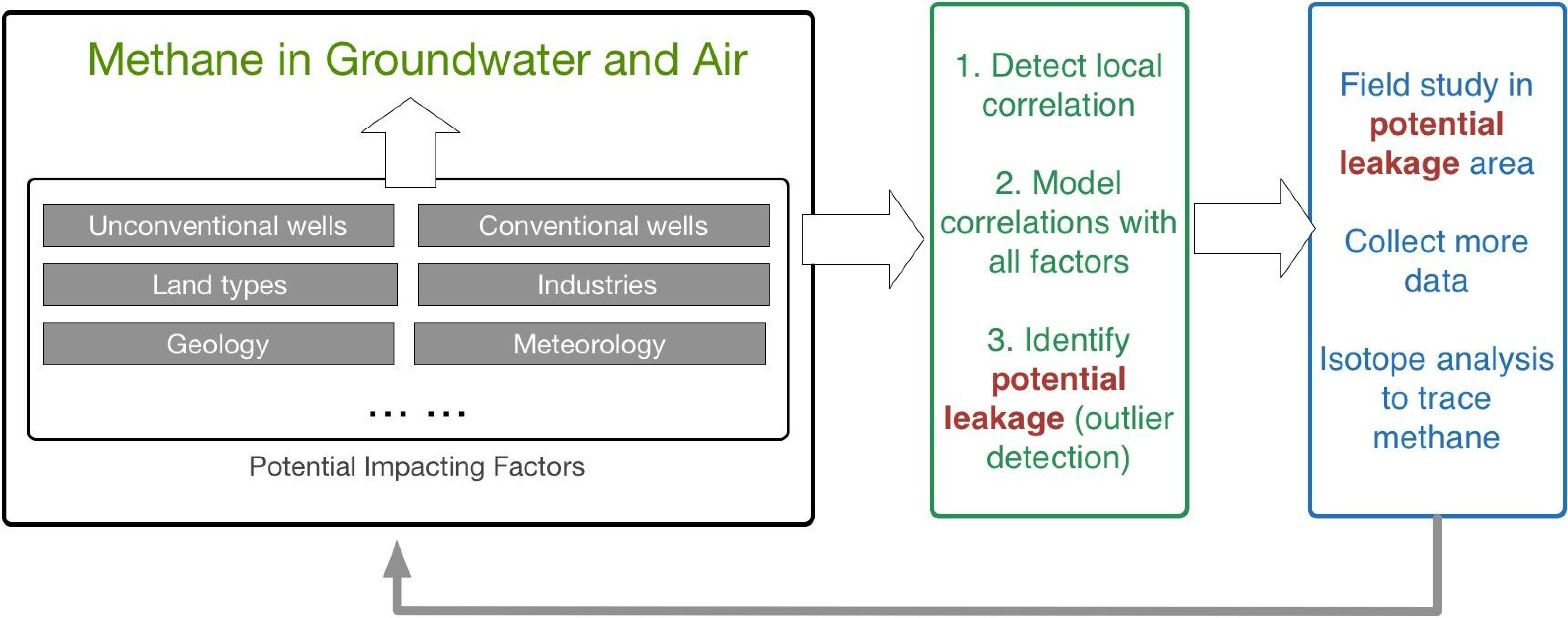
Penn State University

Objective: Analyze methane in groundwater/air to assess impacts of shale gas developments

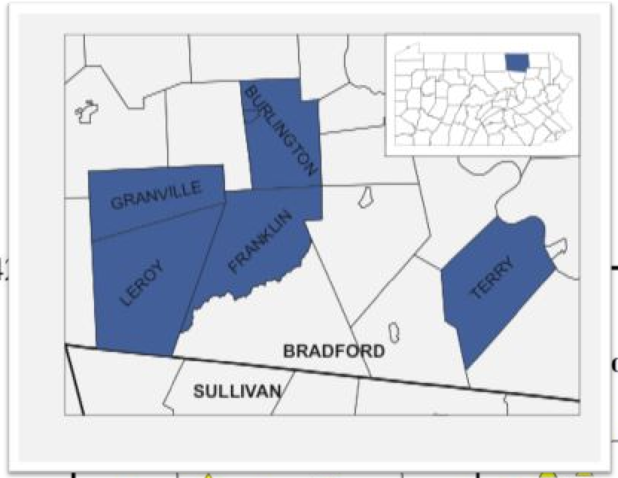


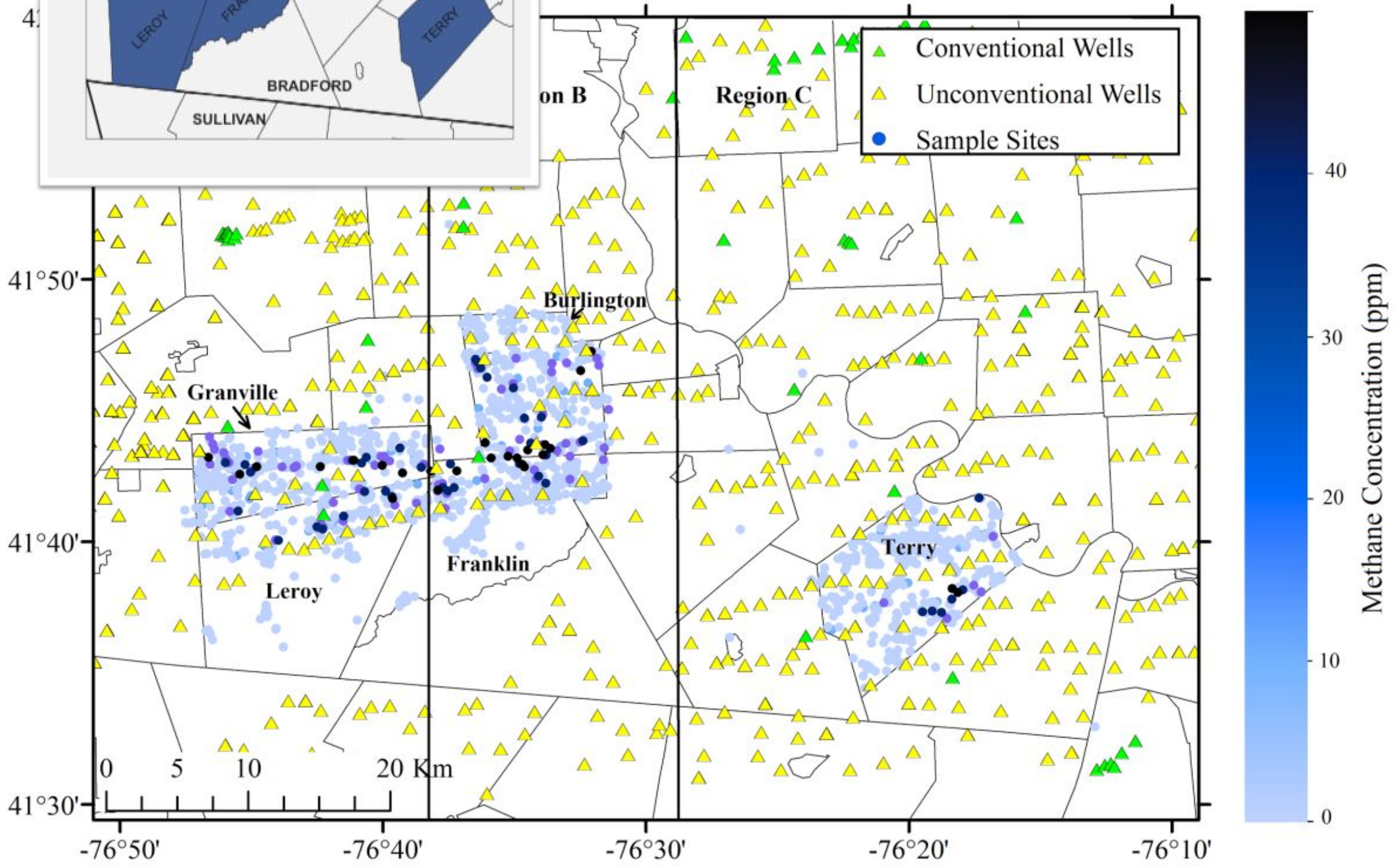
How can we efficiently assess the environmental impacts of shale-gas development in Pennsylvania, given that there are almost 10,000 shale gas wells + >300,000 conventional oil/gas wells + many other impacting factors

Objective: Analyze methane in groundwater/air to assess impacts of shale gas developments: **a bigger picture**



1,684 groundwater samples in Bradford

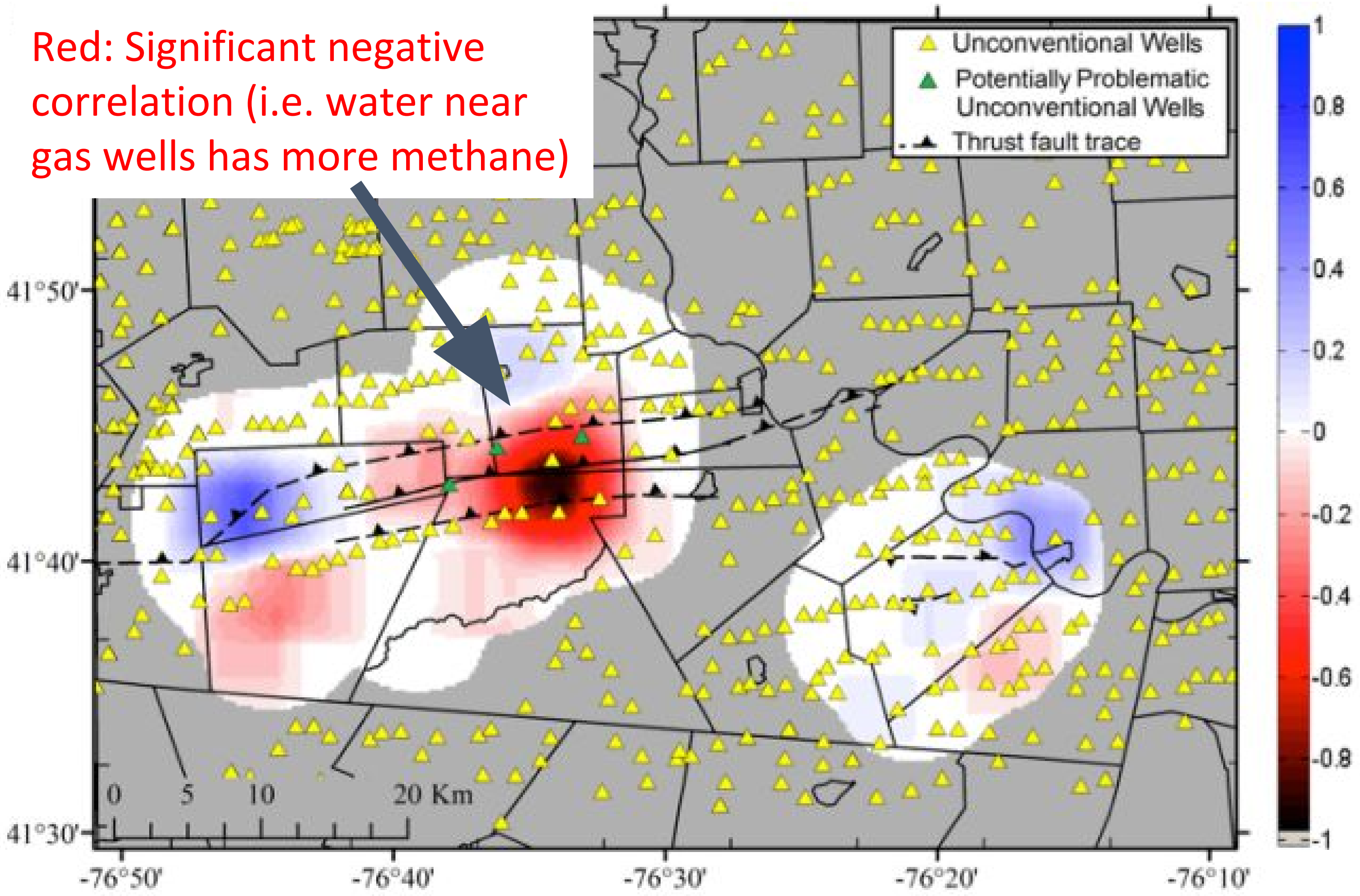




- Data from PA DEP, public on shale network database now
- 30~50 chemical analytes (e.g., methane, barium, ...)
- Blue dots: methane in water (darker → higher)
- Yellow: unconv. wells

Last year's workshop: **a sliding window technique** to **search for locations** with significant correlations between methane concentration and distance to unconventional wells

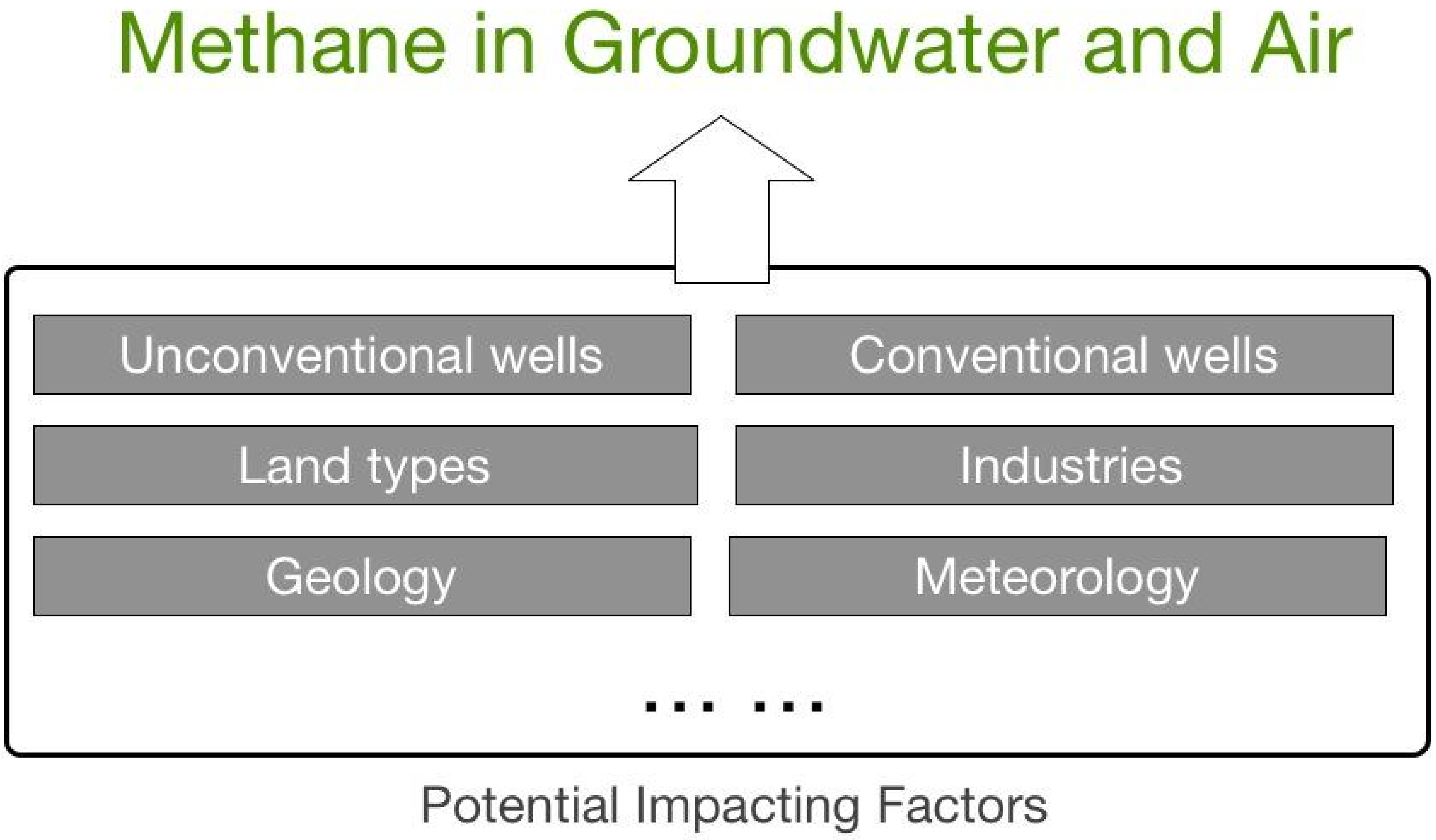
Red: Significant negative correlation (i.e. water near gas wells has more methane)



Observations:

1. Significant correlations in **local** areas
2. Major **faults** going through the hot spot
3. Three wells not have **intermediate casing** at the depth they intersect with fault

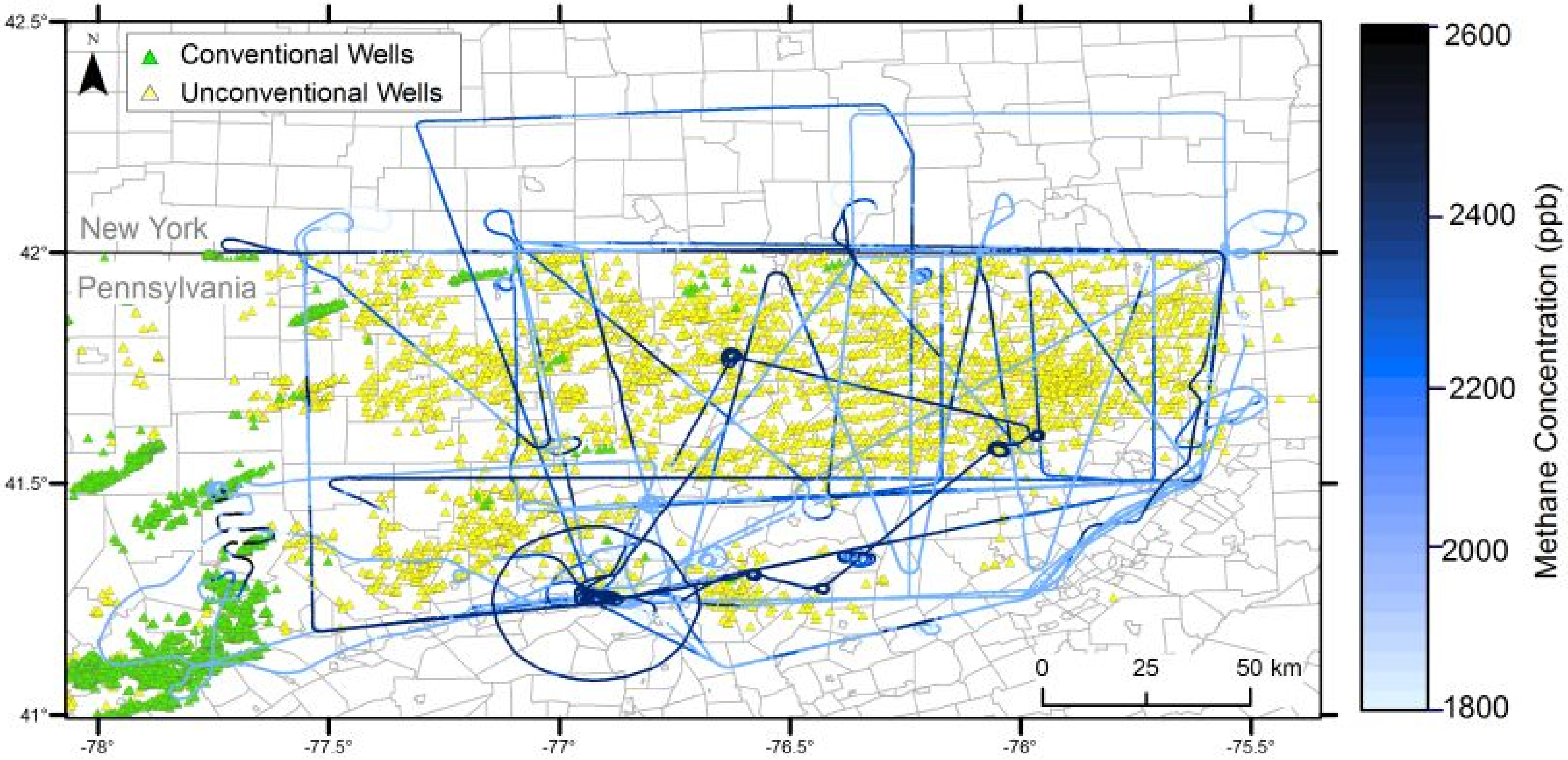
What's new this year?



- Include air quality data

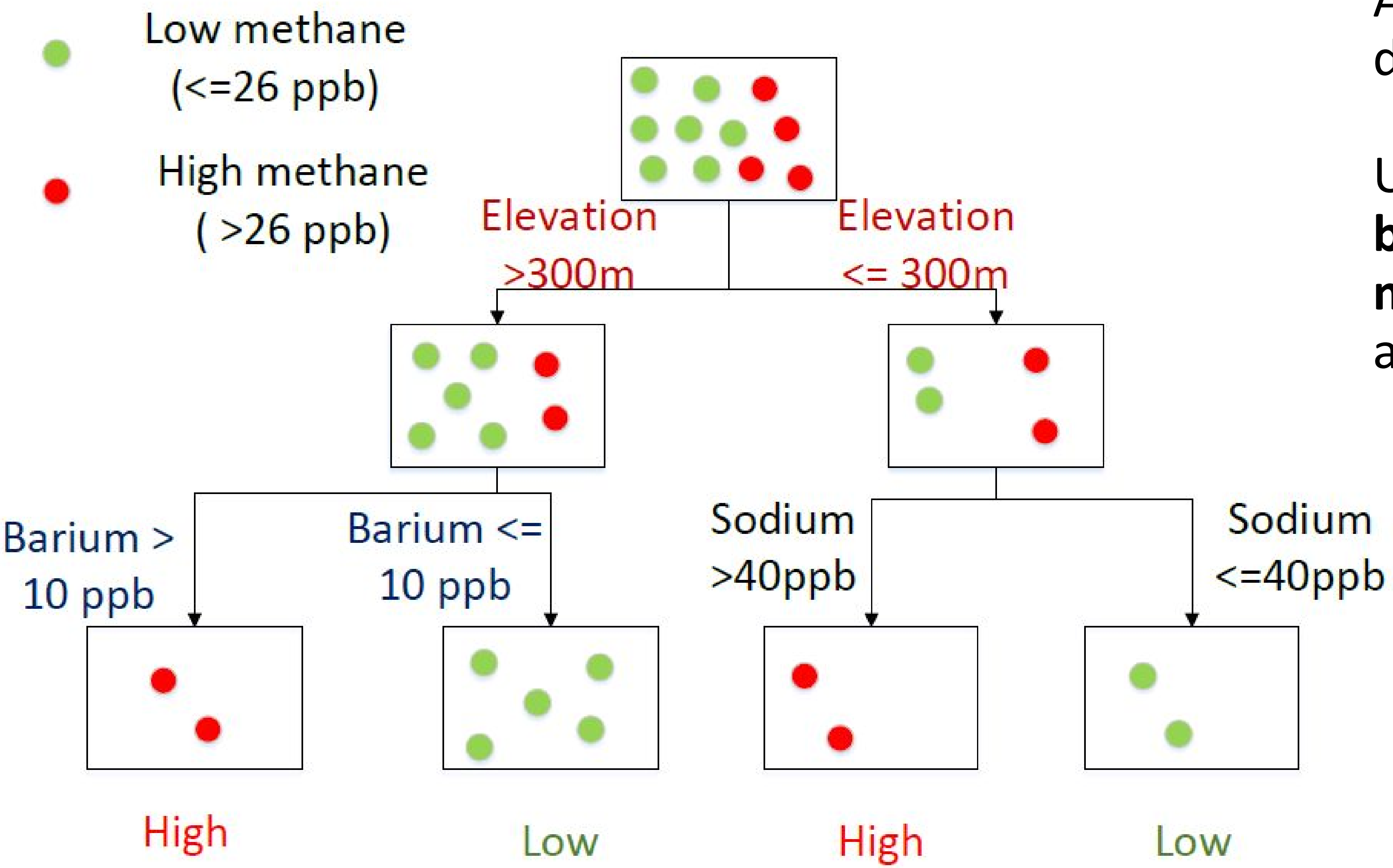
- Consider more factors simultaneously in the analysis

Air measurements made 1 sample/s by T. Lauvaux (meteorologist) along flight paths



- -Blue: methane concentration
- -Yellow/green: wells
- -**64,261** air samples
- -May to June 2015
- -Range: 1788~3326 ppb
- -More data collected by vehicles and towers

Question 1: Can we predict methane values by using impacting factors?



A simple example model: decision tree

Uses **elevation, sodium, barium** to predict **methane concentration** as **low** or **high**

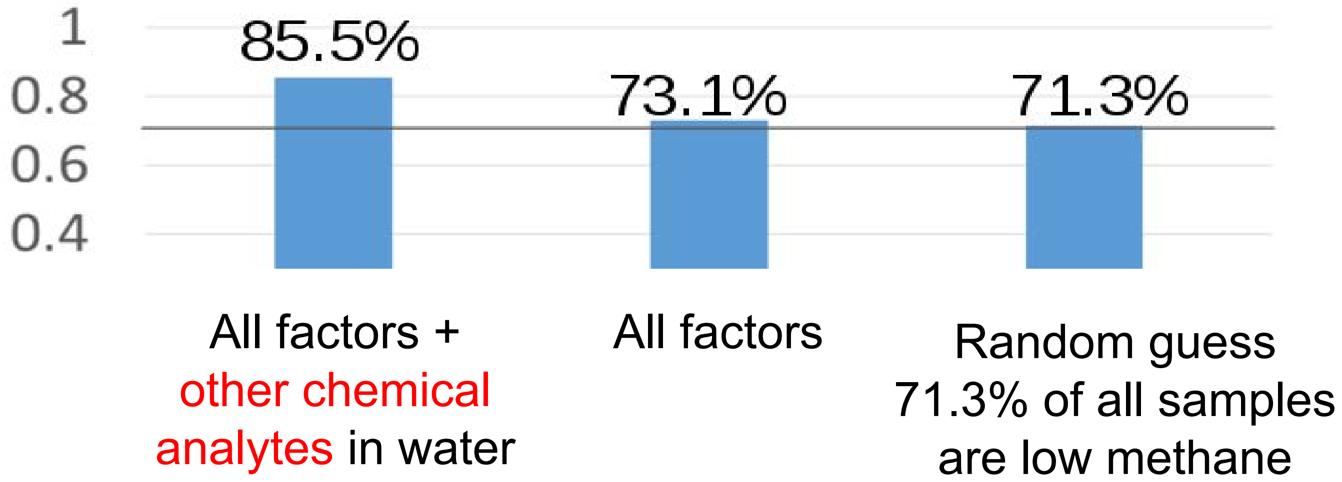
Factors we considered in the model

- Gas wells (conventional , unconventional): distance and density
- Geological features (elevation, faults)
- Land use (e.g., wetland)
- Industries
- Meteorology

Prediction result by the best model

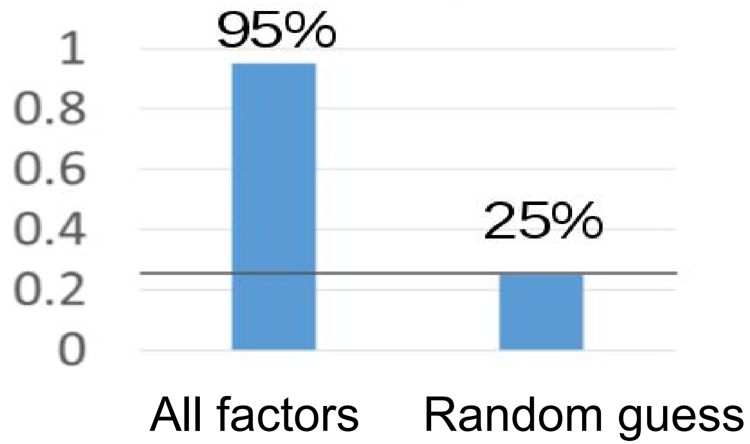
Methane in groundwater

'High' (> 26 ppb), 'Low' (≤ 26 ppb)



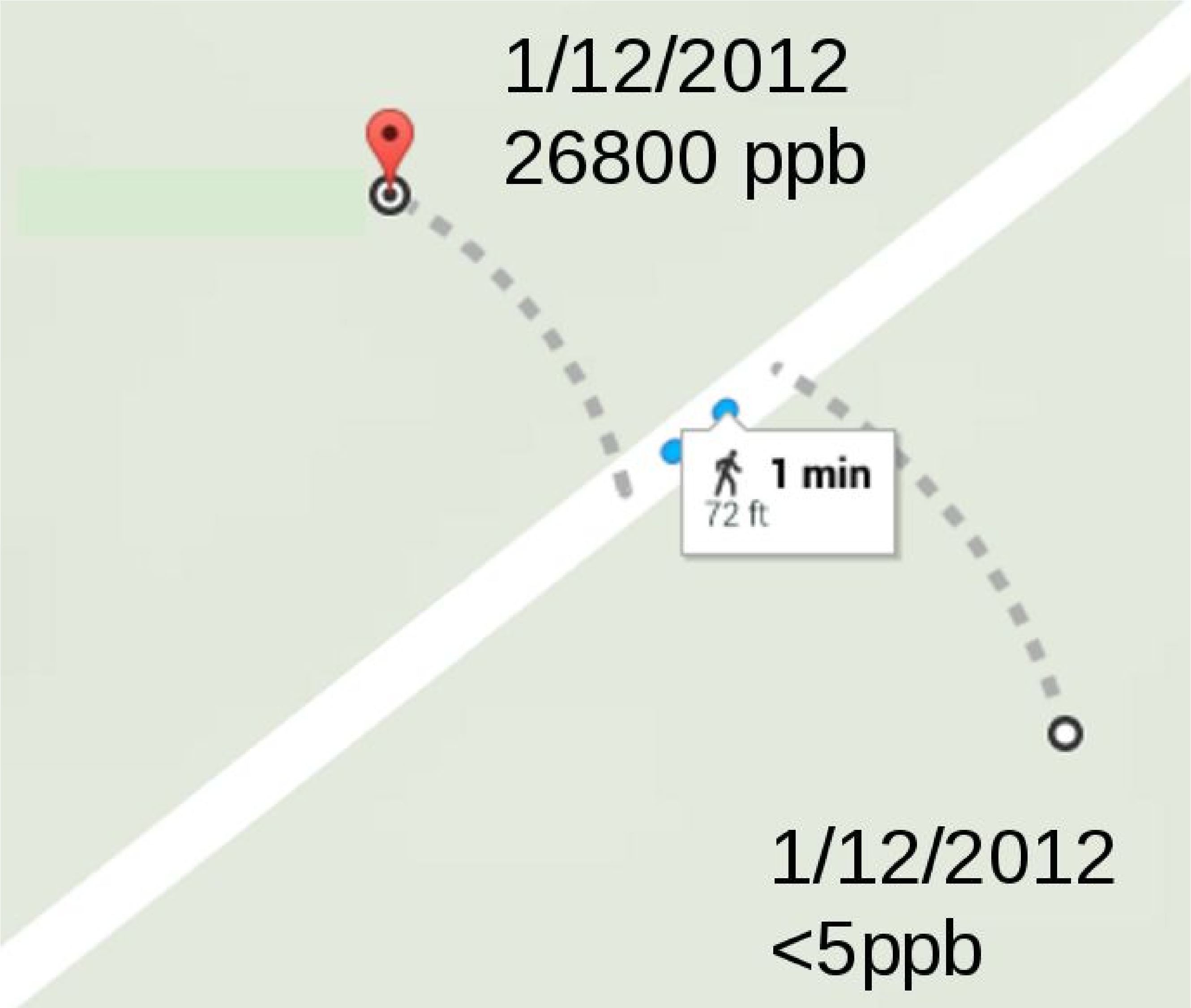
Methane in air

4 quartiles: 0, 1911, 1921, 1939 ppb



1. The water chemistry is **hard to predict** (73.1% vs. 71.3% random guess)
2. Using other chemical analytes improves the accuracy (85.5%). Some **chemical concentrations (e.g., Barium) are highly correlated with methane concentrations** in groundwater in Bradford county (likely because natural gas moves with **brines** in the area)
3. We can predict air chemistry (95%) much **better** than water chemistry

Why predicting water is hard?



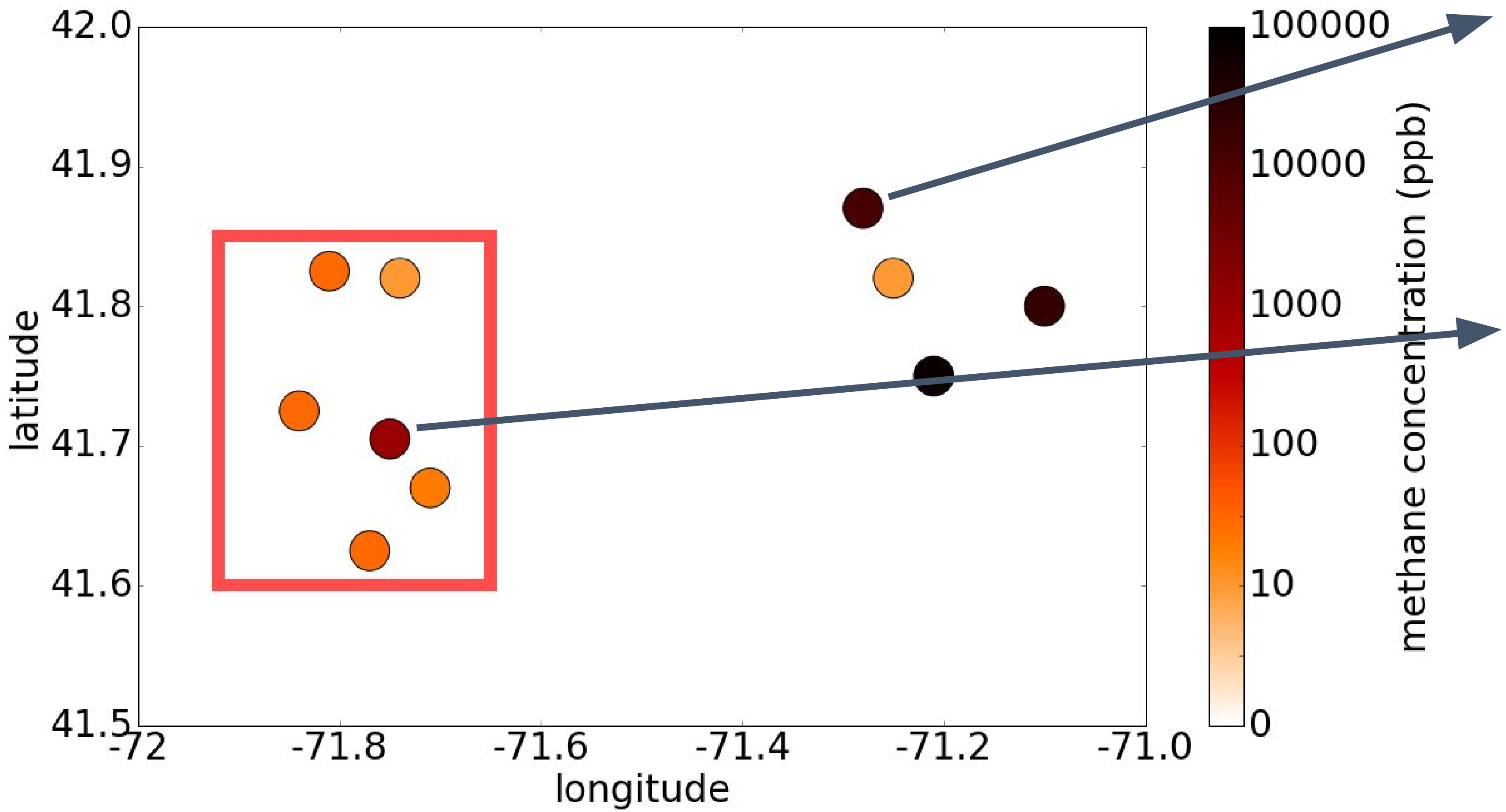
- High spatial **heterogeneity** of methane concentration

- Not enough predictors

These two water samples are only 72 feet from each other, but differ by more than 20,000 ppb methane

Question 2: Detecting **non-trivial outliers** defined by **local region**

A toy example. Darker color: higher methane. Lighter color: lower methane

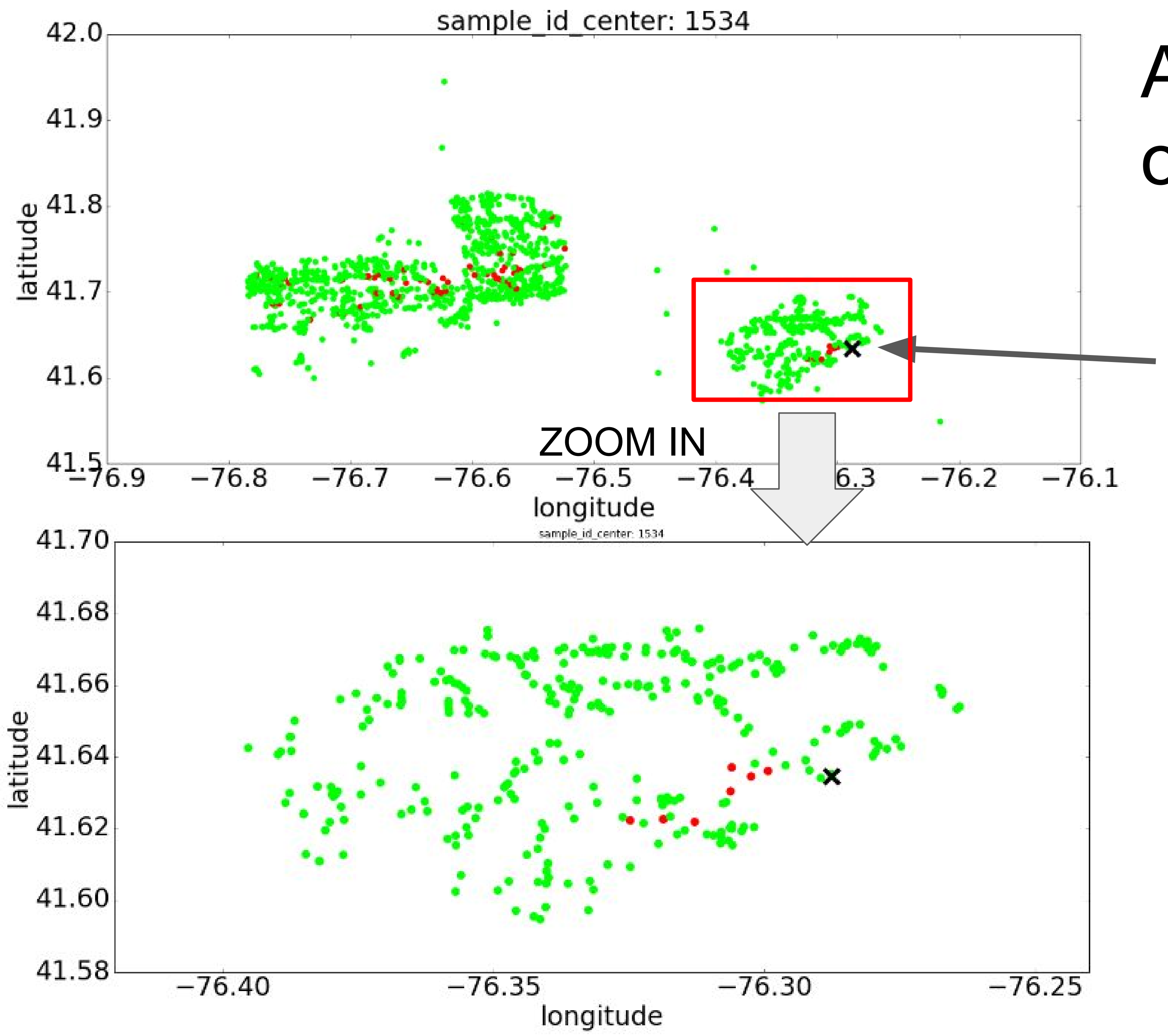


"Trivial" outliers: top-3 samples with highest methane

Non-trivial outlier:

Rank **4th** (globally)
Rank **1st** (in this region)

Anomalous site 1 on real data



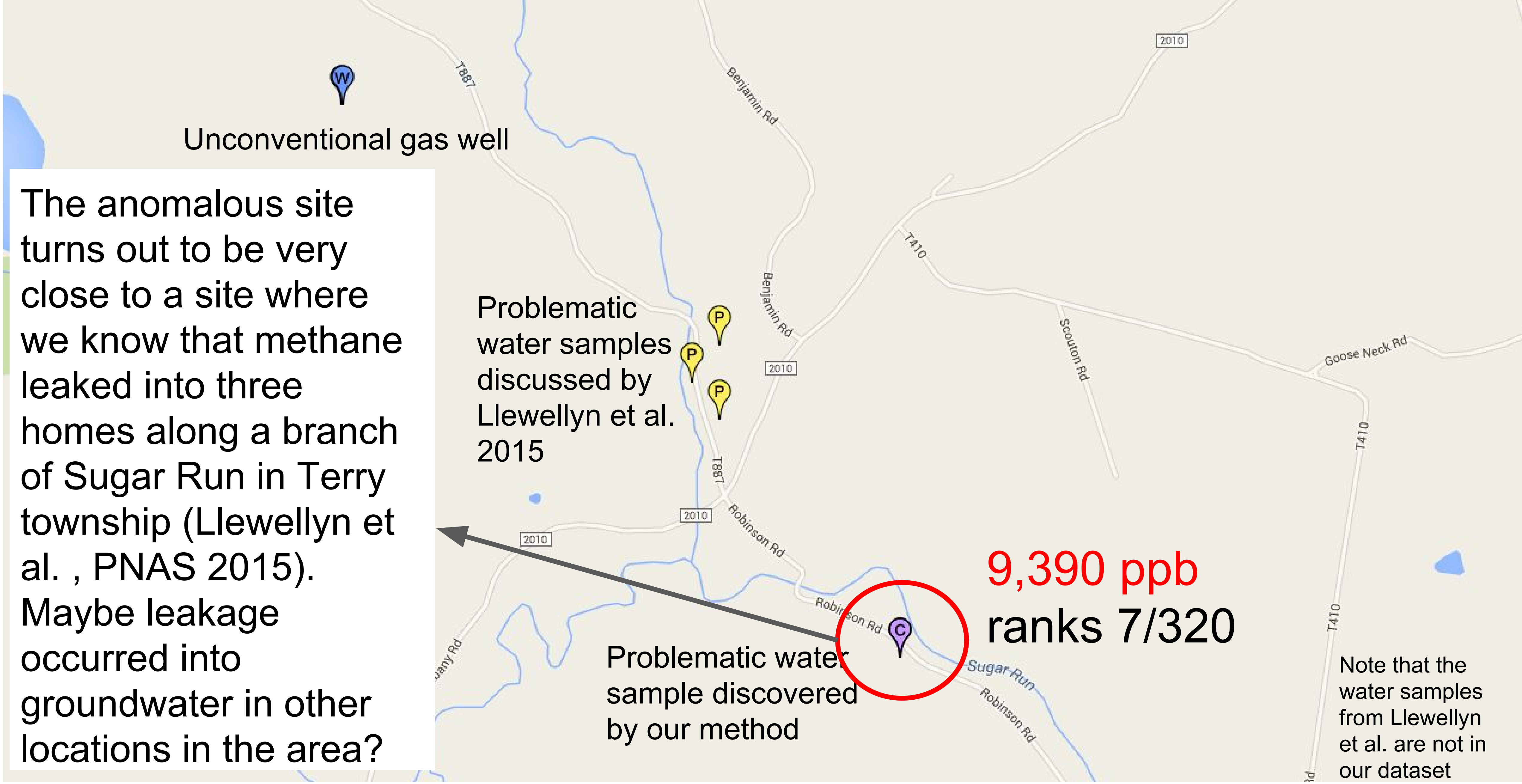
Methane at X: 9,390 ppb

green: lower than outlier
red: higher than outlier

In the global region, this point ranks **68/1545 (top 4.2%)**

In this local region, this point ranks **7/320 (top 2.1%)**

The anomalous site turns out to be very close to a site where we know that methane leaked into three homes along a branch of Sugar Run in Terry township (Llewellyn et al. , PNAS 2015). Maybe leakage occurred into groundwater in other locations in the area?

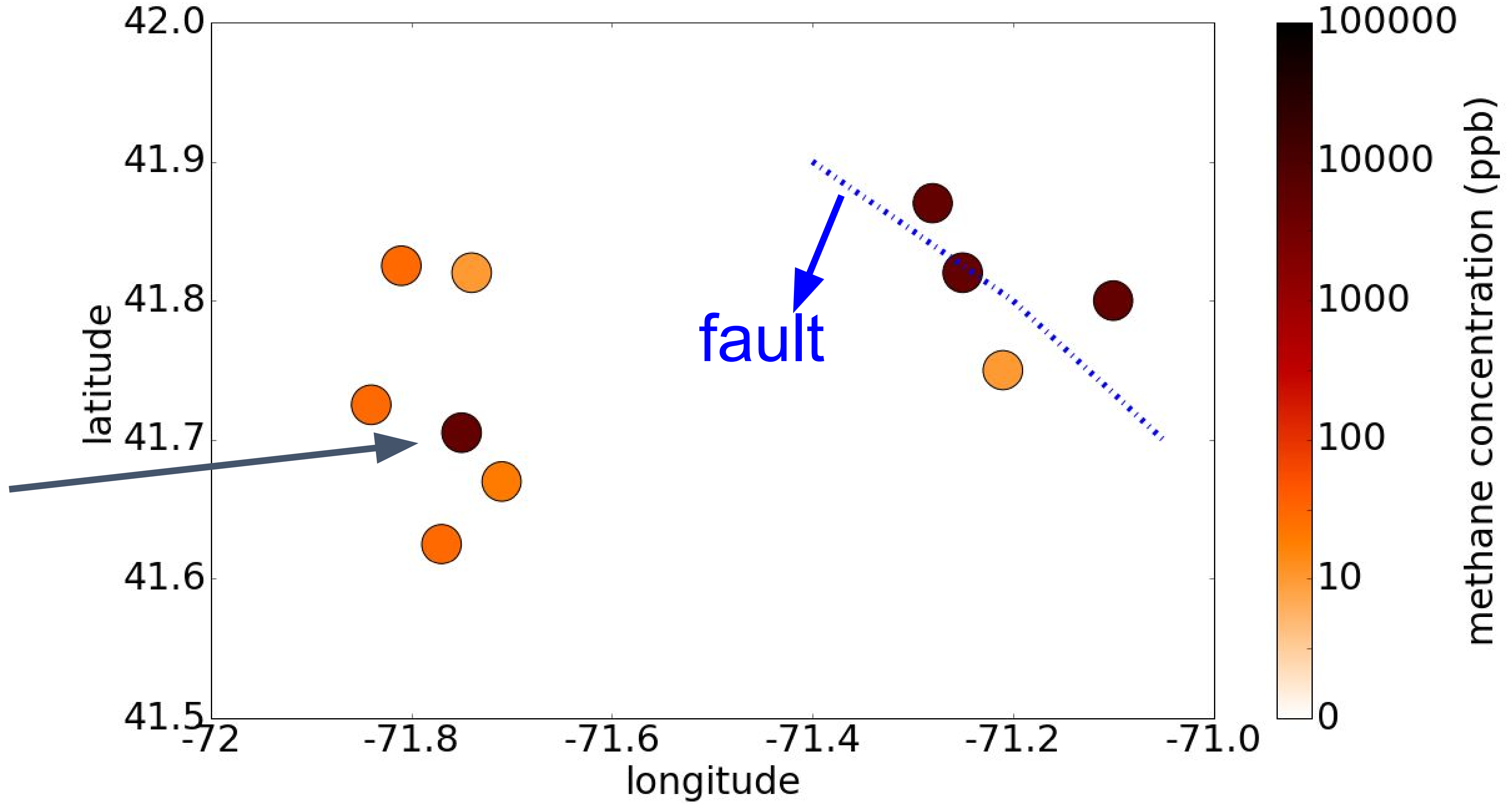


focal point Uncon Well Llewellyn

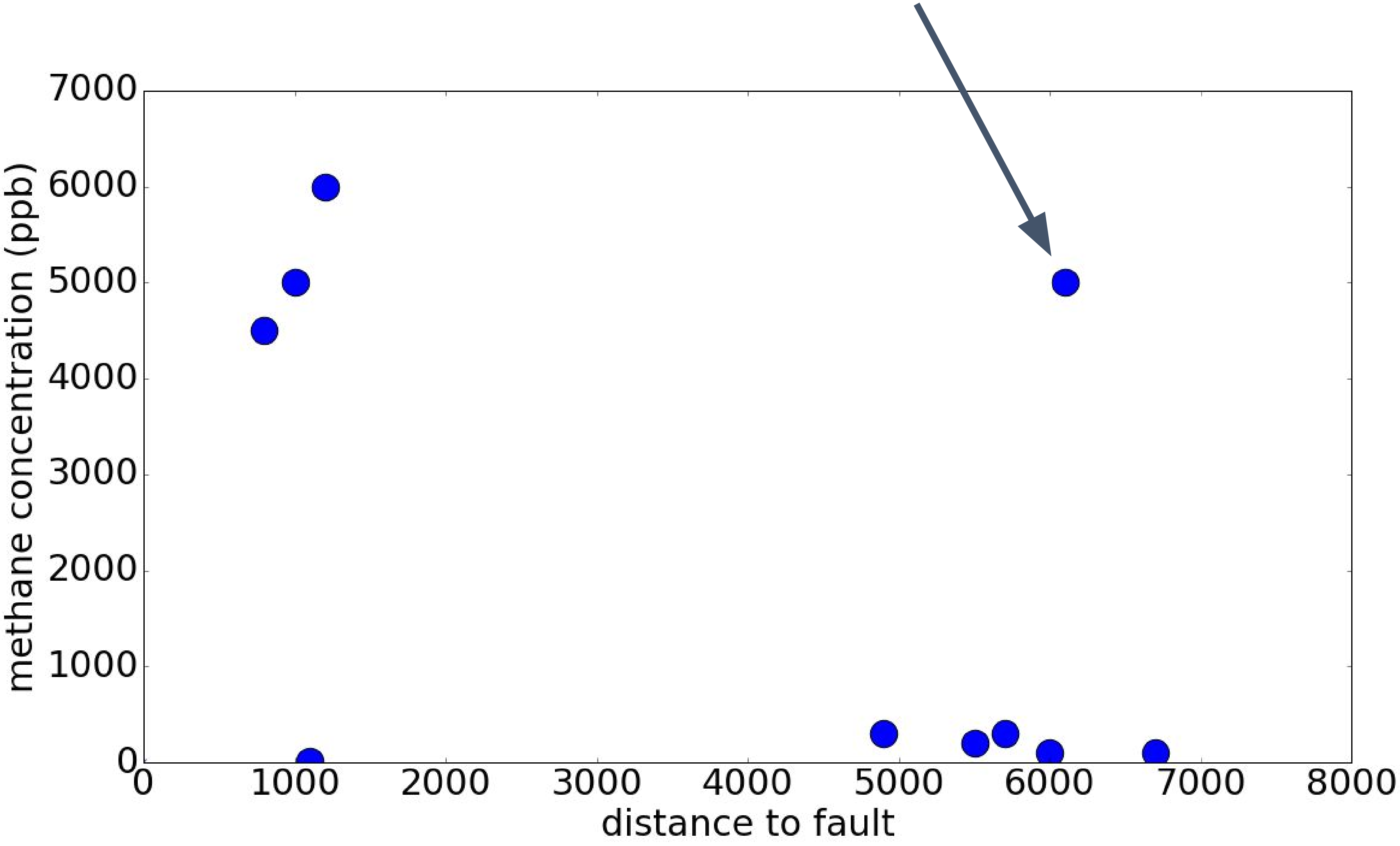
Question 2: Detecting non-trivial outliers defined by **other factors (e.g., fault, well)**

A toy example. Darker color: higher methane. Lighter color: lower methane

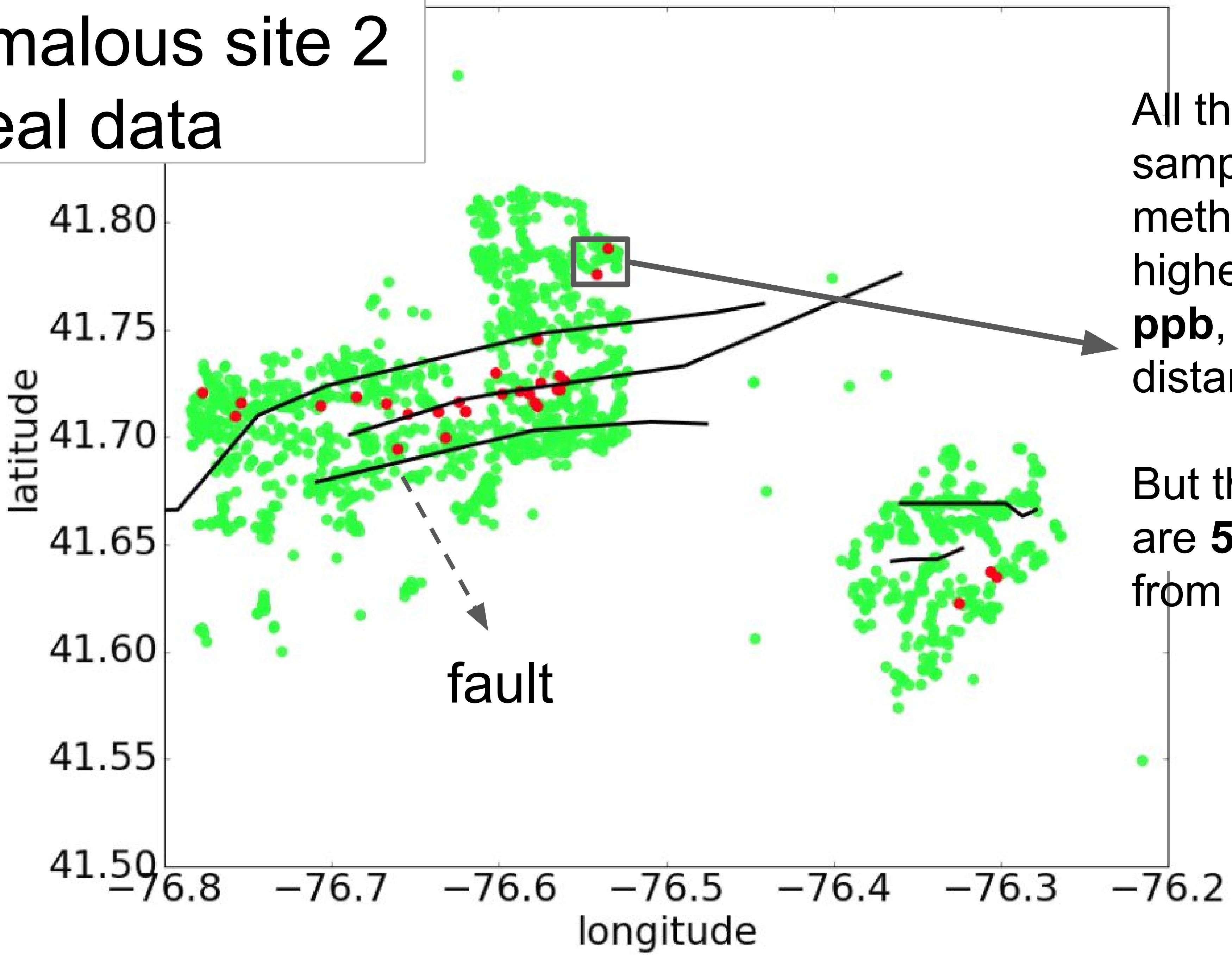
Most of the high-methane samples are along the fault, but **this point** is far from fault



Most of the high methane concentrations are along the fault, but this point is far from fault



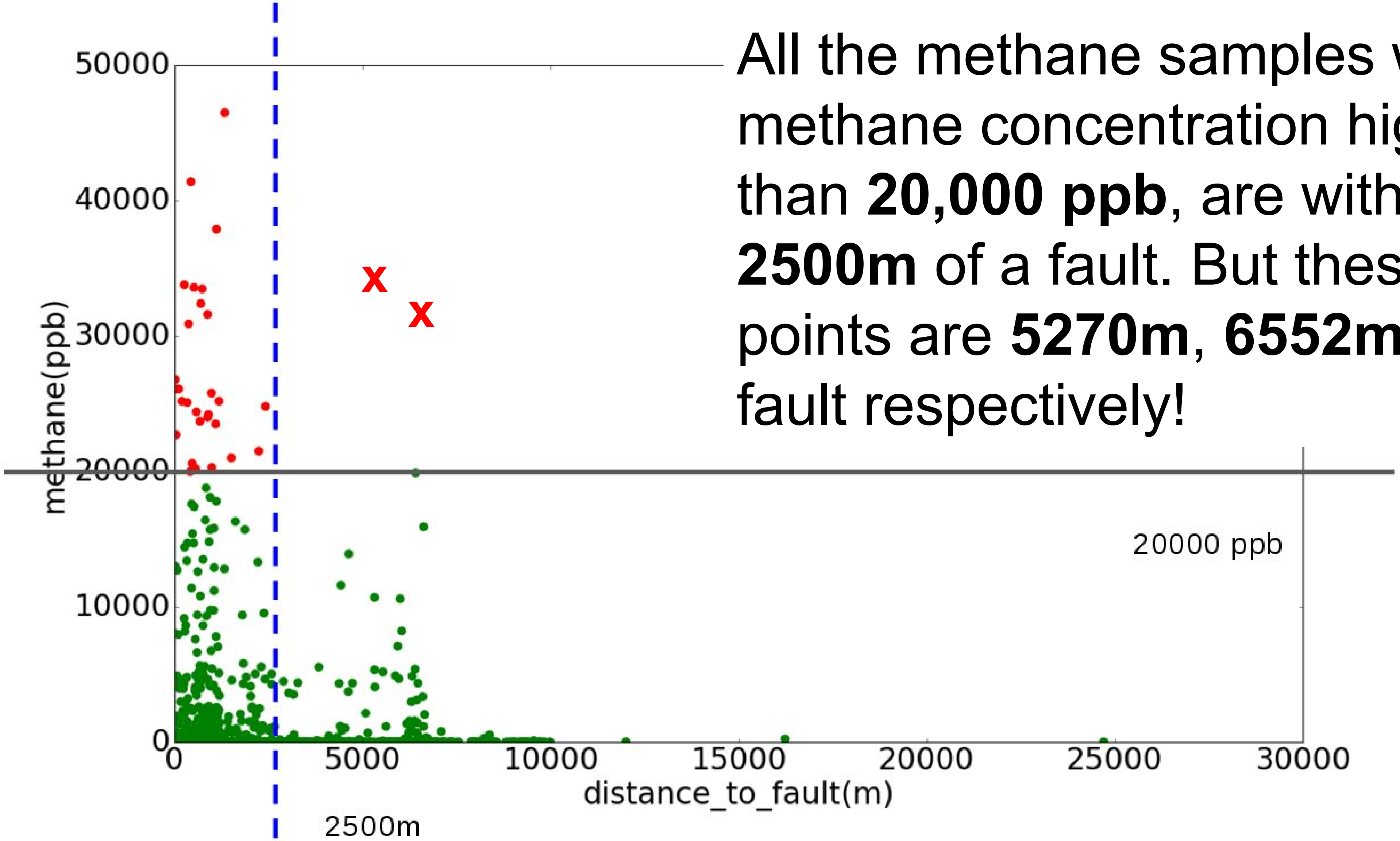
Anomalous site 2 on real data



All the methane samples with a methane concentration higher than **20,000 ppb**, are within **2500m** distance to fault.

But these two points are **5270m, 6552m** from fault respectively!

Anomalous site 2 on real data



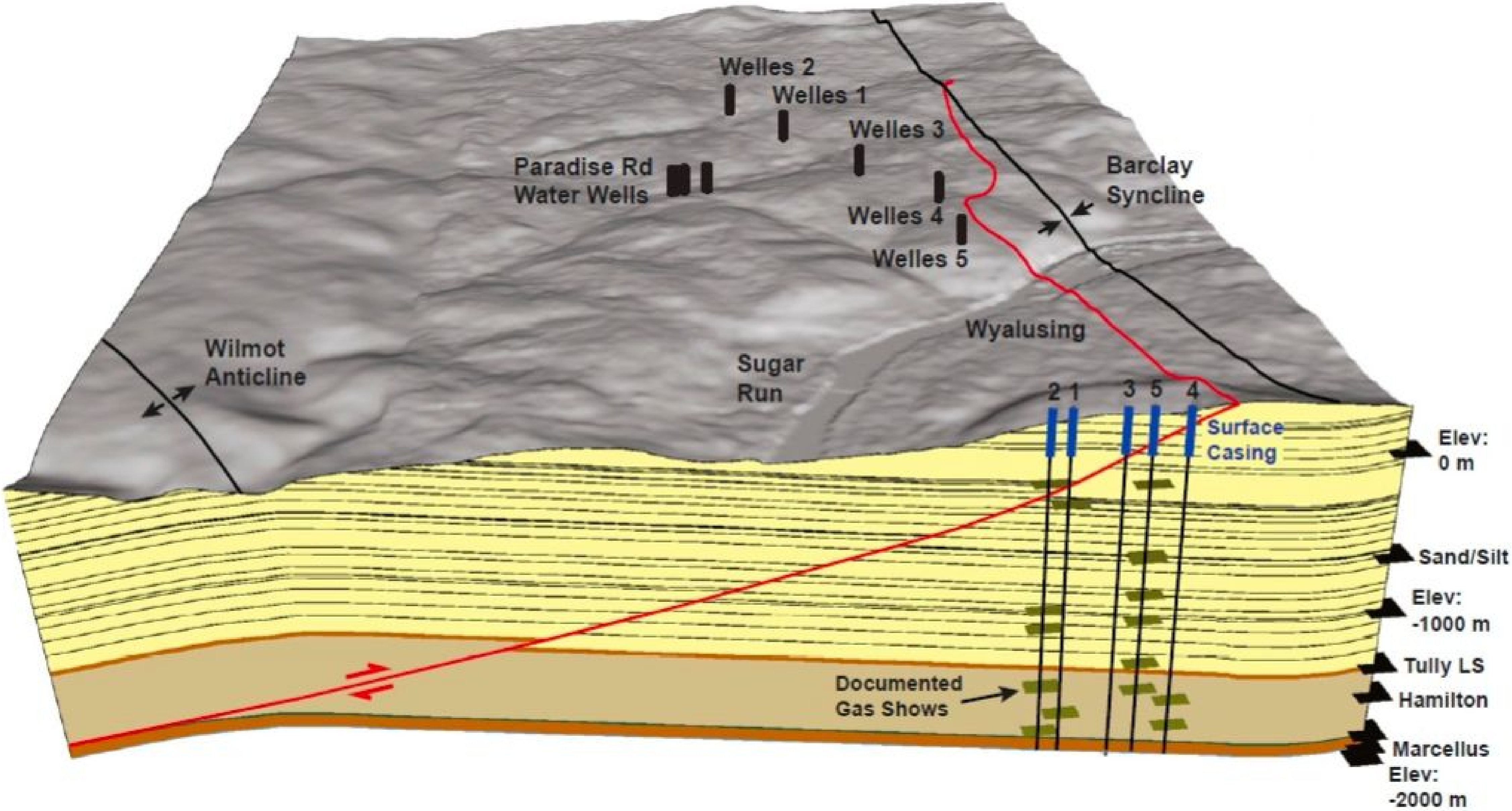
All the methane samples with a methane concentration higher than **20,000 ppb**, are within **2500m** of a fault. But these two points are **5270m**, **6552m** from fault respectively!

Acknowledgements: GE Gift Fund to Penn State (for J. Li, SLB, GZ, FW, MG); NSF funding to SLB for Shale Network data; Jon Pollak of CUAHSI for help with data from HydroDesktop; data provision by PA DEP (working with Seth Pelepko, Bill Kosmer).

Conclusion

- **Modeling** methane in water/air with correlations to all potential impacting factors
  - The **water chemistry is hard to predict** (73.1% vs. 71.3% random guess)
  - Some chemical concentrations (e.g., **Barium**) are **highly correlated with methane** concentrations in groundwater in Bradford county (likely because natural gas moves with **brines** in the area) → 85.5% accuracy when using other analytes
  - We can predict **air** chemistry much **better** than water chemistry
- When large groundwater datasets are made public, we can develop tools such as **outlier detection** to locate potentially problematic regions
  - Anomalous site 1 defined by **local area**, anomalous site 2 defined by **distance to fault**
  - Such an approach may be necessary to find problems in a state with **10,000 new shale gas wells** and more than **300,000 conventional wells**

Supplementary materials



Lewellyn et al., 2015 (PNAS)