Spatial variability of tight oil well productivity and the impact of technology

Justin Montgomery

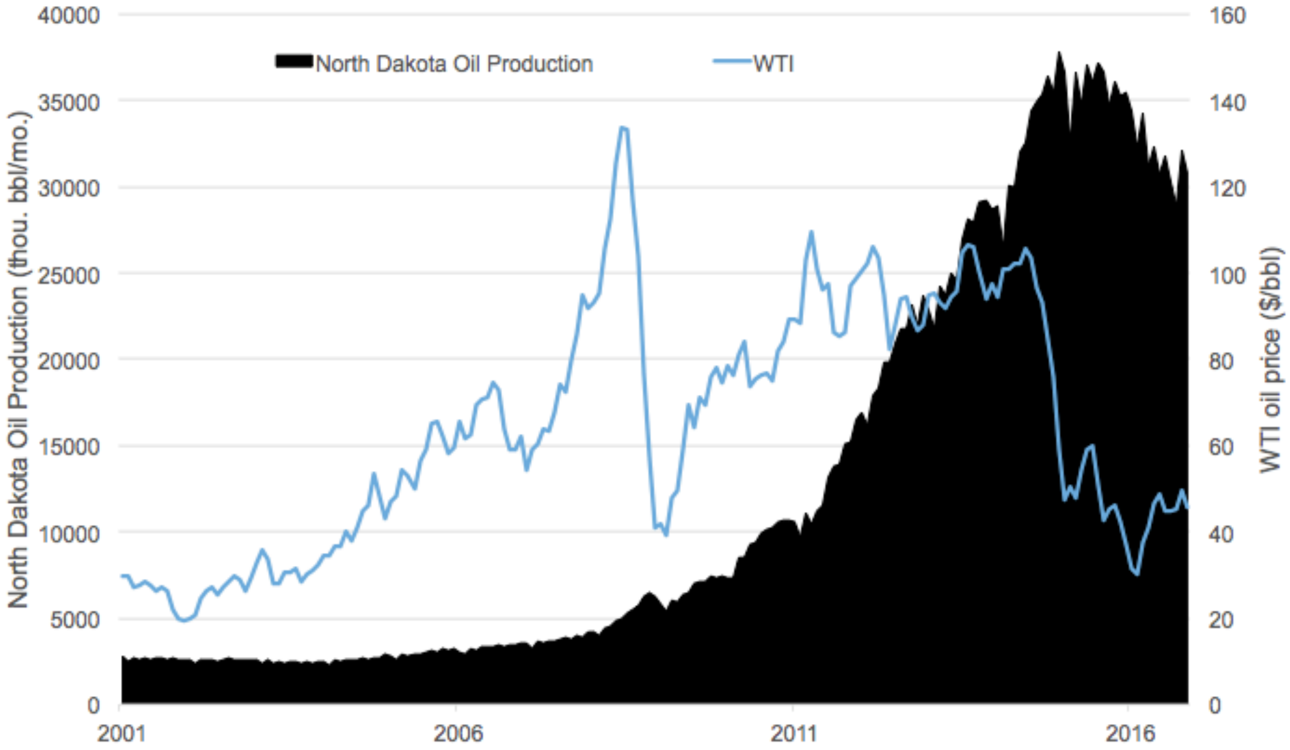
PhD Candidate
Department of Civil and Environmental Engineering
In collaboration with Francis O'Sullivan and John Williams

MIT Earth Resources Laboratory
2017 Annual Founding Members Meeting
May 31, 2017

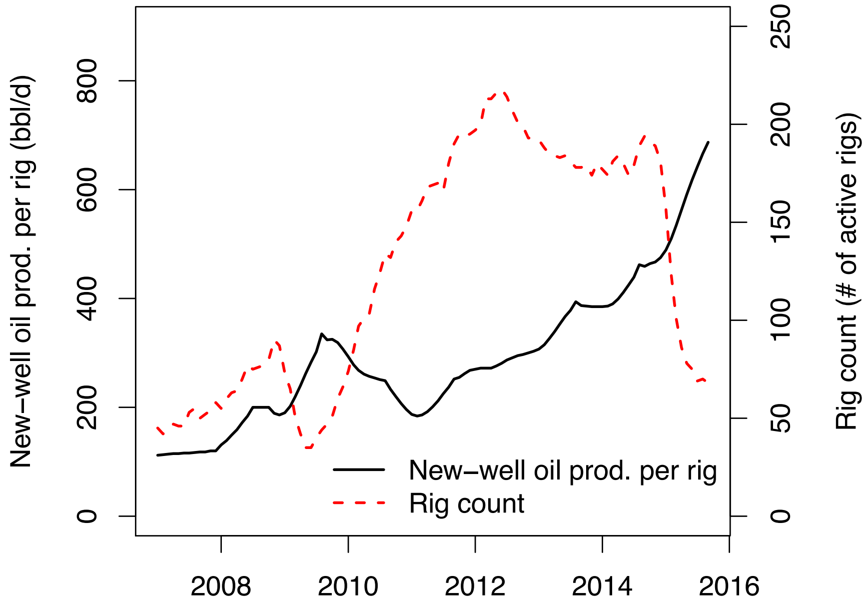
Massachusetts Institute of Technology

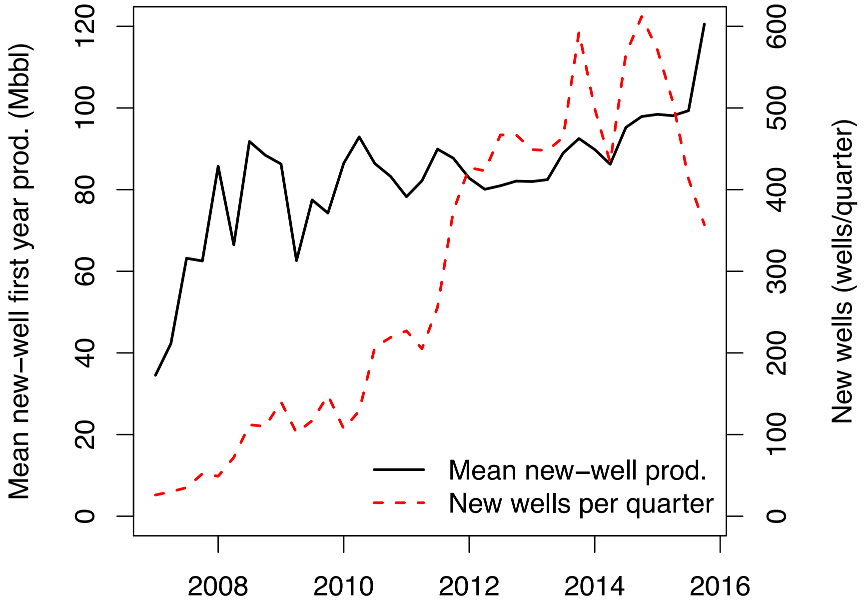
Earth Resources Laboratory

Williston Basin of North Dakota was at the forefront of tight oil extraction but now faces economic uncertainty



Rising rig and well productivity suggest greater resilience than expected

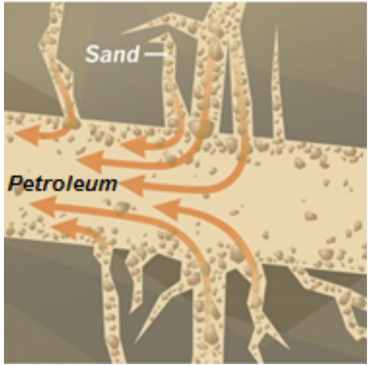




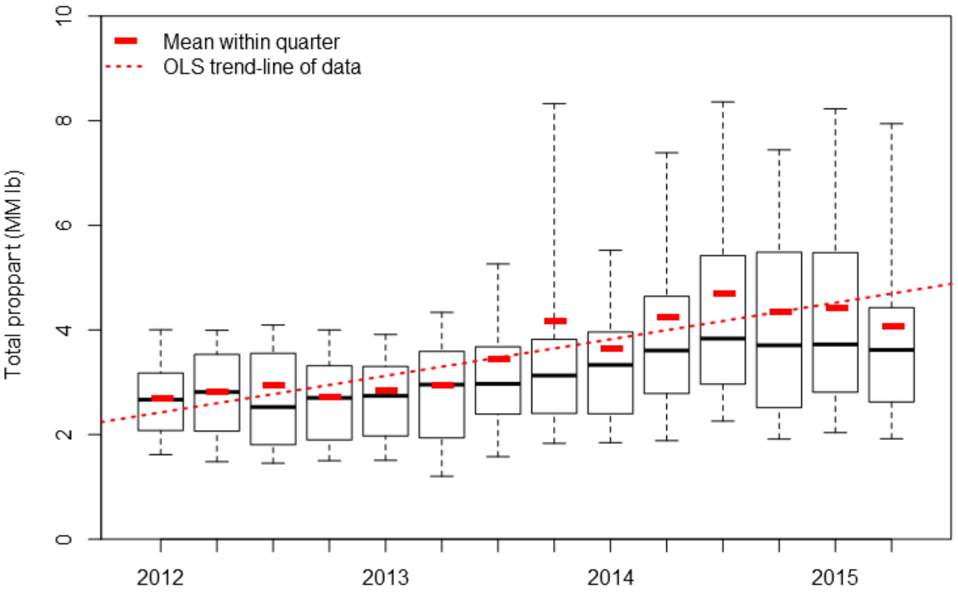
Improvement of well productivity has been driven in part by changes in well and stimulation design

- Trends toward longer wells and larger stimulations (hydraulic fracturing)

- Motivation for identifying impact:
  1. **Forecast well productivity based on anticipated changes**
  2. **Optimize wells**



Increase in proppant (sand) per well over time



Massachusetts Institute of Technology

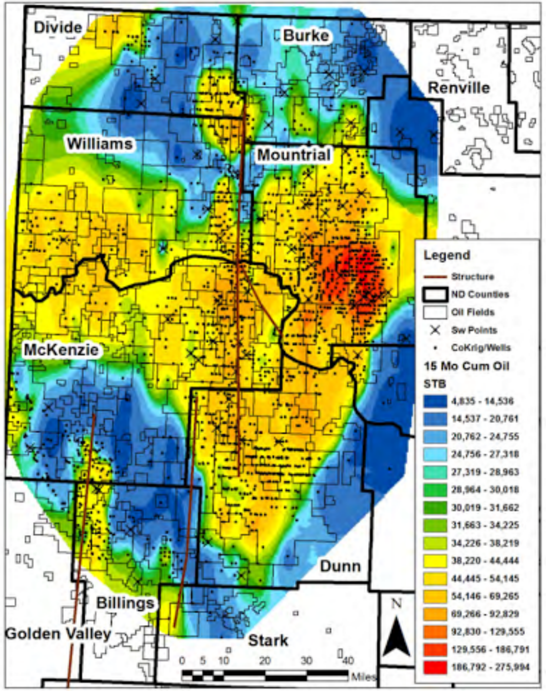
Earth Resources Laboratory

Another important dynamic is where wells are being drilled –
"sweet-spotting" or "high-grading"

- Activity continuing to cluster in high productivity areas

- Motivation for identifying location influence:
  1. **Need to control for this to accurately understand impact of design changes**
  2. **Assess well portfolios and resource economics based on location in field**

Well productivity heat map



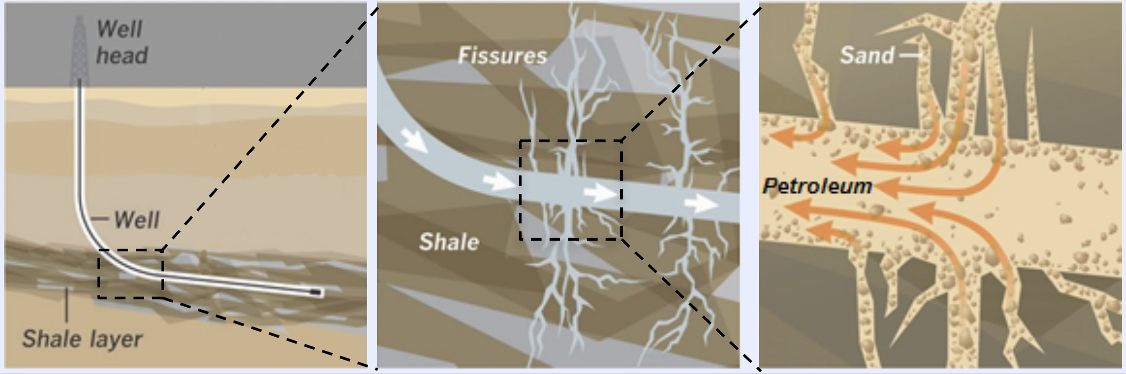
Source: Schmidt, 2011

How much of the improvement in well productivity is due to technology (design changes) vs location (sweet spotting)?

- Big public datasets available (Frac Focus, North Dakota Mineral Resources)
- Can we use **econometrics/machine learning** to understand and make predictions?

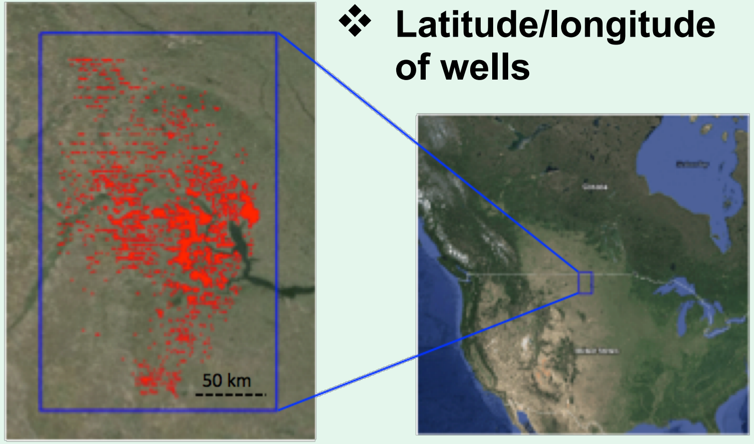
Technology

- Horizontal length of completed well
- Amount of water injected to create fissures
- Amount of proppant (sand) injected to "prop" fissures open



Location

- Target formation (Bakken/Three Forks)
- **Latitude/longitude of wells**

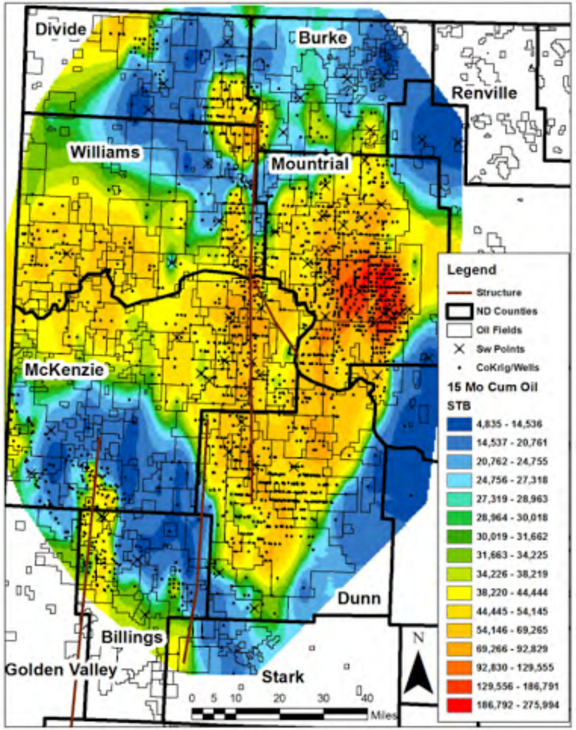


Current regression models to understand the influence of technology on productivity

- Nonspatial linear regression (NS)
- Fixed Effects (FE), such as county-level used by EIA

- Issues:
  - Not spatially granular enough
  - Residuals are spatially autocorrelated
    - → Omitted variable bias

$$Y = X\beta + \epsilon$$

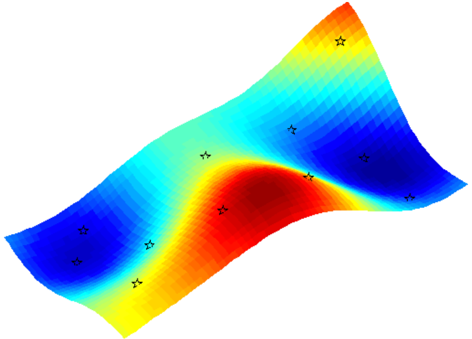


Regression-kriging provides an appropriate tool for distinguishing between impact of location and technology

$$Y = X\beta + \lambda e + u$$

Technology trend

Spatial component



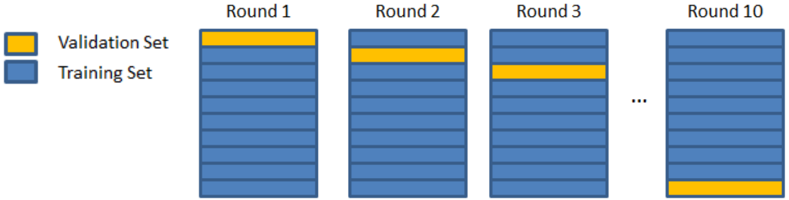
Start

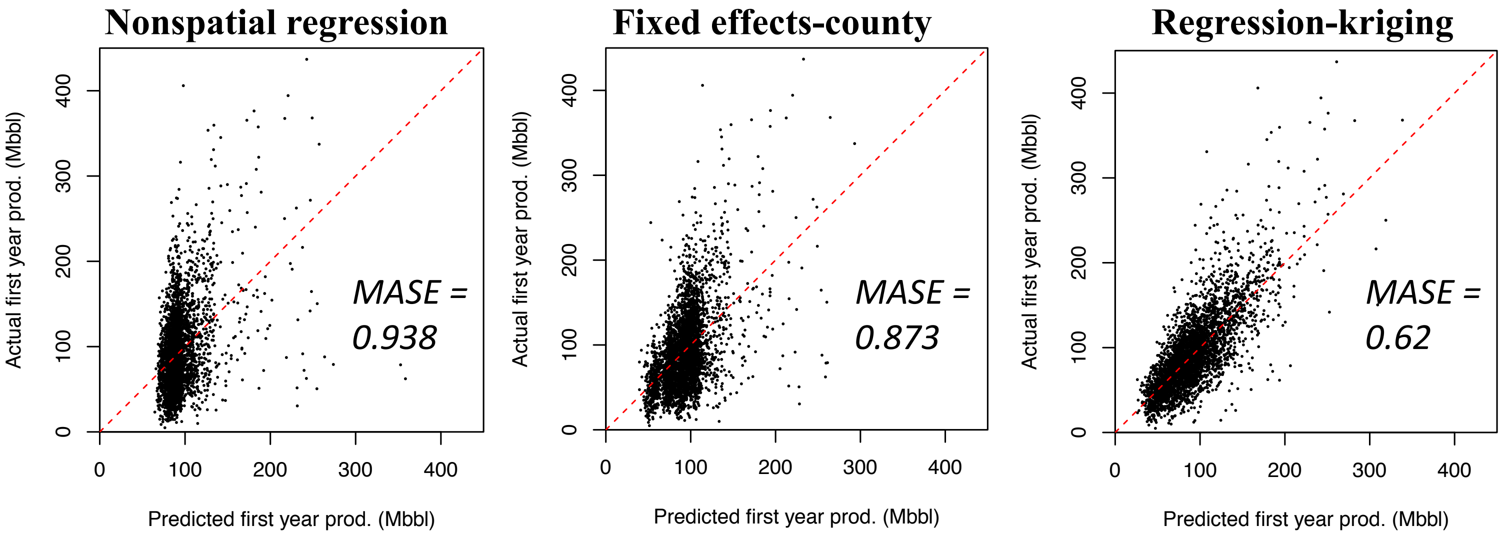
Estimate trend with linear regression → *Estimate spatial correlations by kriging residuals* → *Converged?*

No → *Remove interpolated spatial component* → *Estimate trend with linear regression*

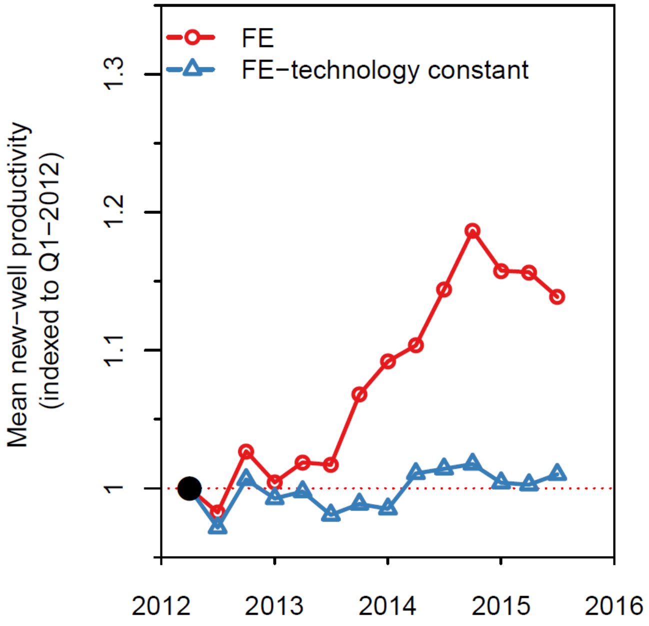
Yes → **Estimates for model parameters**

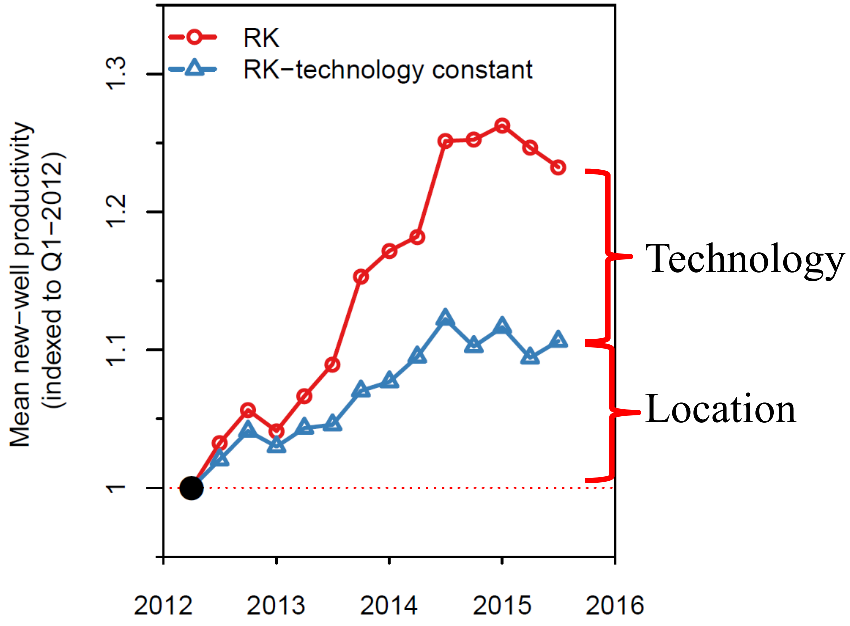
RK improves accuracy (in 10-fold cross validation) compared to currently used regression models





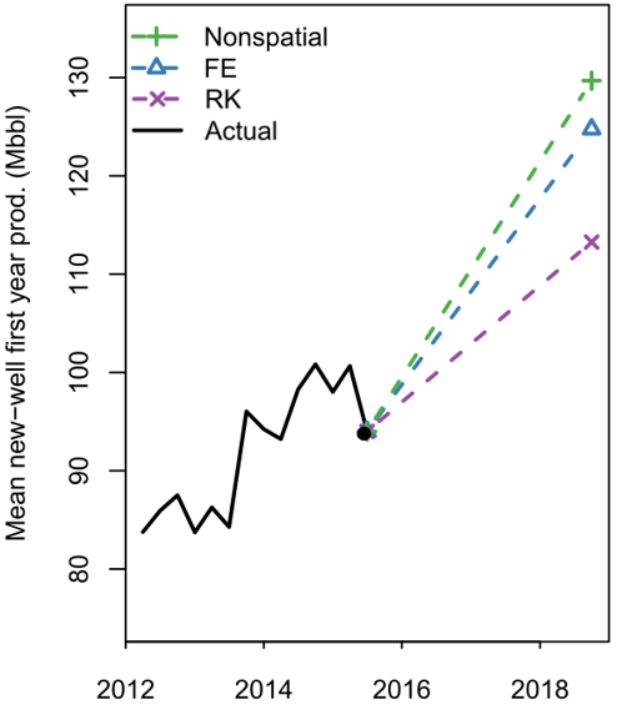
Existing regression models overestimate the role of technology relative to location





Overestimating the impact of technology leads to overoptimistic forecasts and poor design choices for wells

Forecasts for 2018 designs

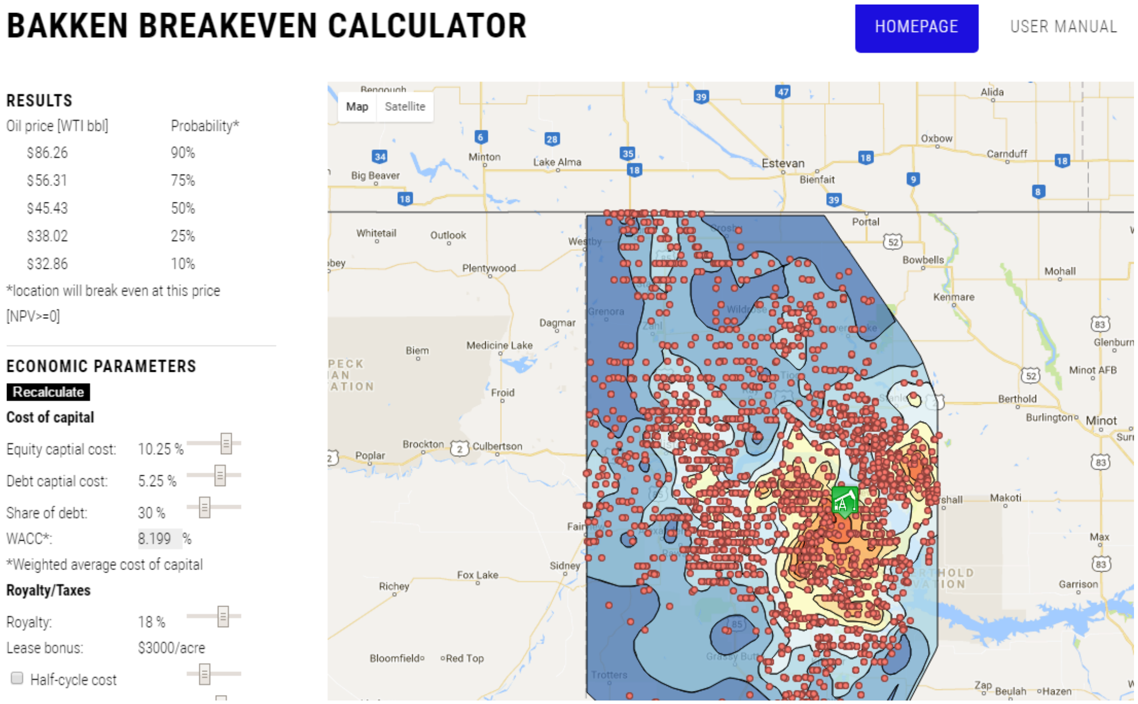


Key findings

1. Regression-kriging improves prediction accuracy
2. Shifts in well design and drilling location have **contributed equally** in recent years
3. County-level fixed effects inadequate to detect sweet-spotting → EIA forecast is likely overoptimistic
4. Current models encourage **over-stimulation** of wells

Future work

- Apply to other unconventional fields
- Predict decline rates
- Use to develop improved field-scale economic models



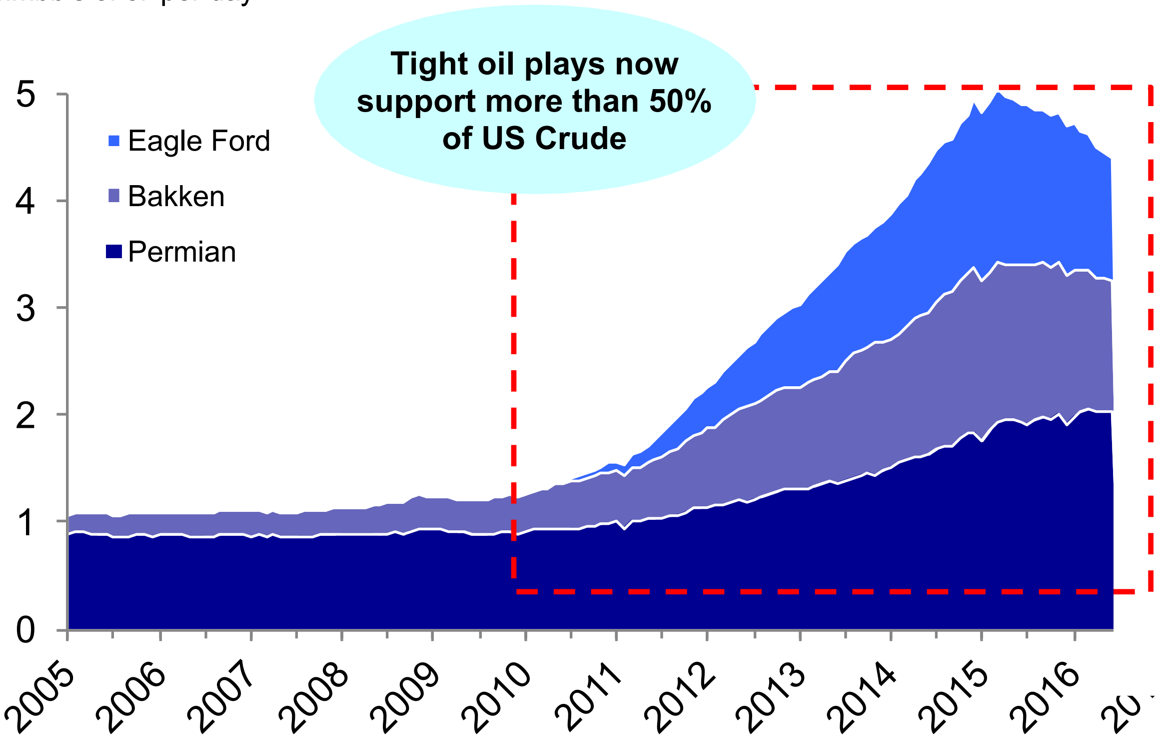
Thank you! Questions?

- Thank you to MIT Energy Initiative for supporting this research

- Full paper is:
  Montgomery, J. B., & O'Sullivan, F. M. (2017). Spatial variability of tight oil well productivity and the impact of technology. Applied Energy, 195, 344-355.

US tight oil production growth has demonstrated the potential of shale and other unconventional formations – Combined output from three of the main US plays alone is now equivalent to the total output of China or Canada

Illustration of crude oil production growth from some select major U.S. unconventional oil plays since 2005
MMbbls of oil per day

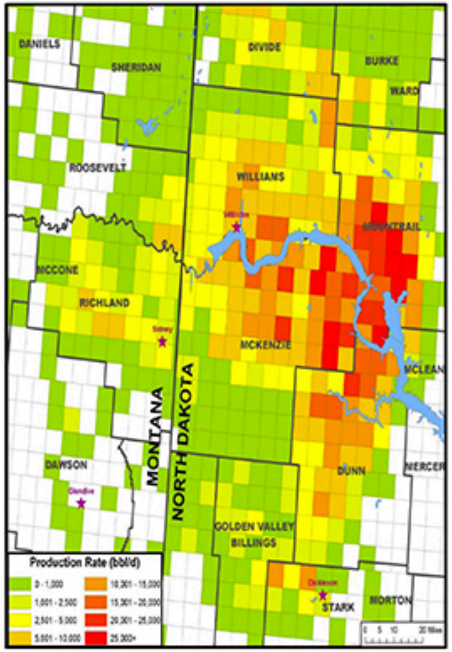


Combined, the Bakken, Permian and Eagle Ford have added more than 4 MMbbls per day of production to US output over the past 5 years

Source: F. O'Sullivan, United States Energy Information Administration, HPDI Production Database

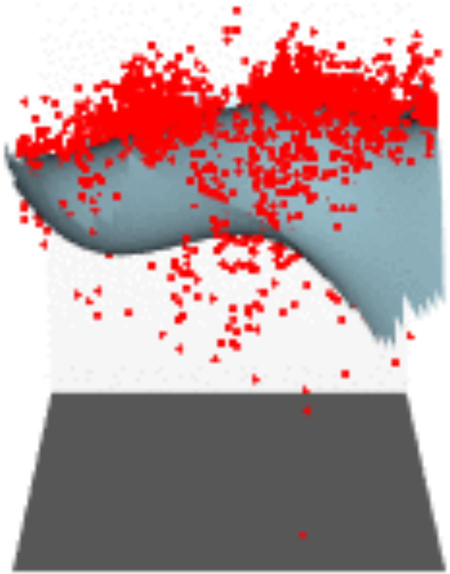
Some other approaches that have been used to control for location

Fixed effects – county or township level

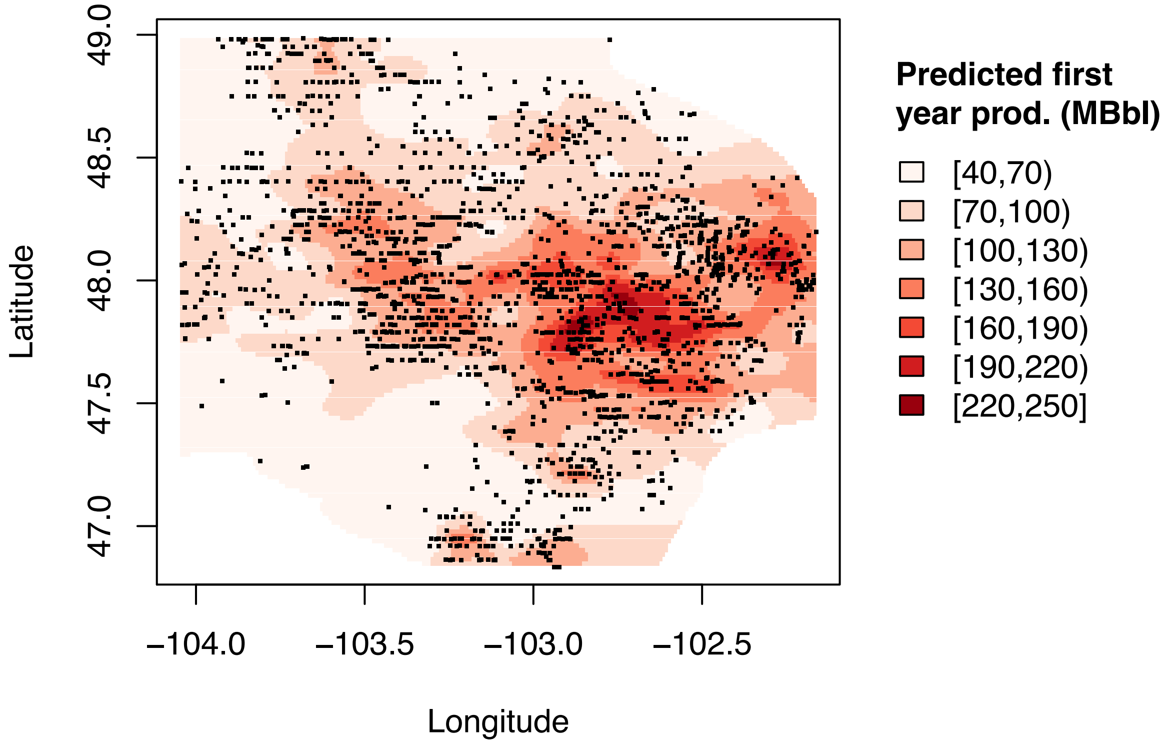


Source: RBC Capital Markets "Bakken Heat Map" – May 19, 2014

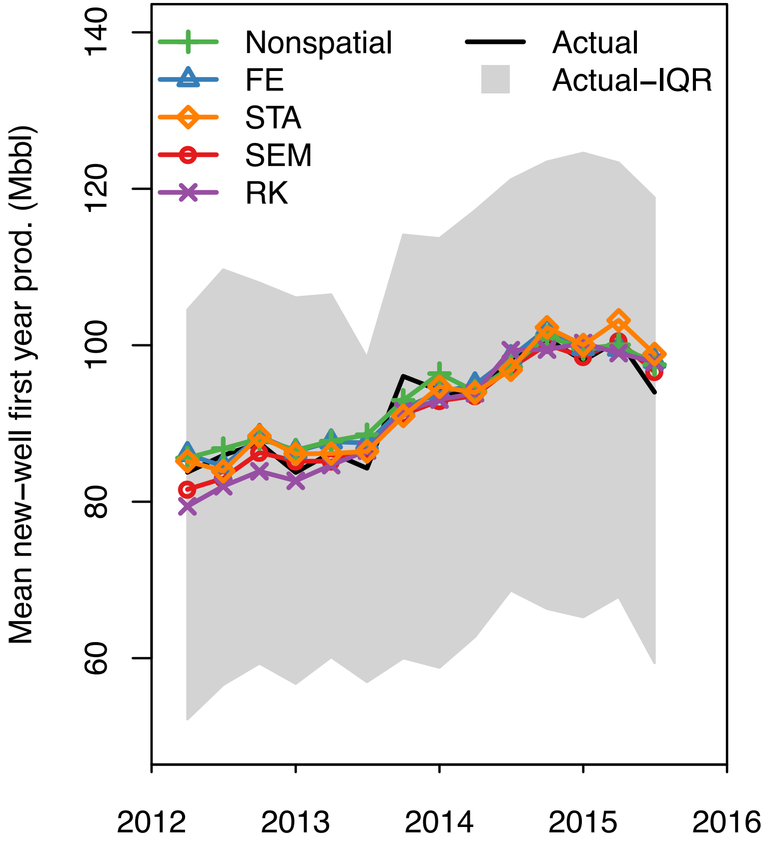
Surface trend analysis (productivity fit to polynomial of coordinates)



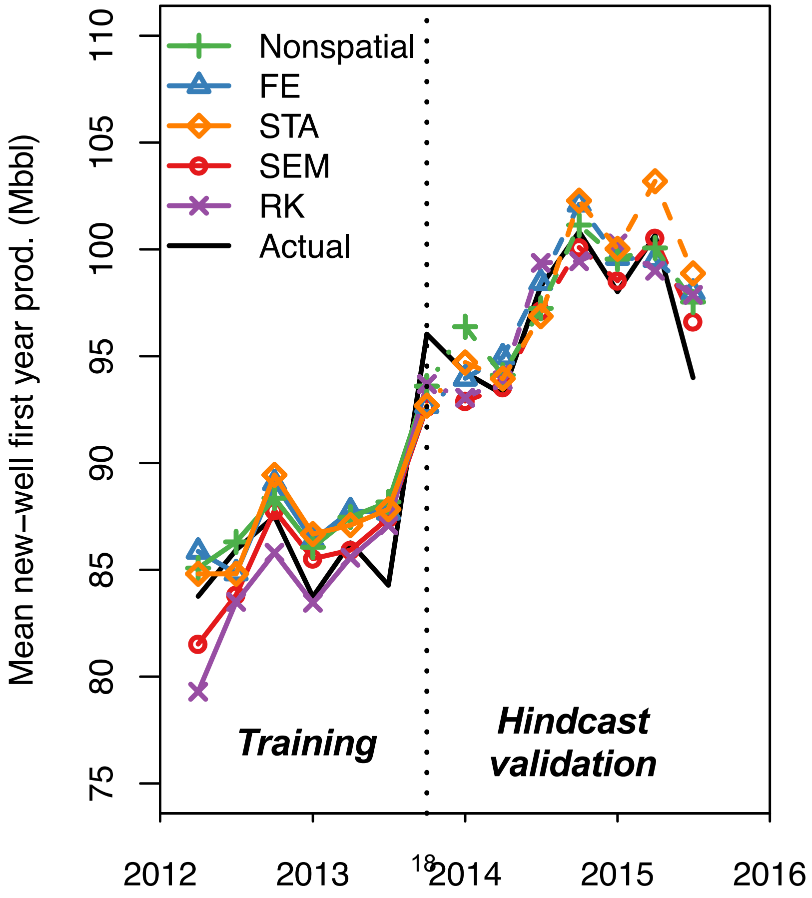
Results of regression kriging – Productivity forecast with typical well designs for 2018



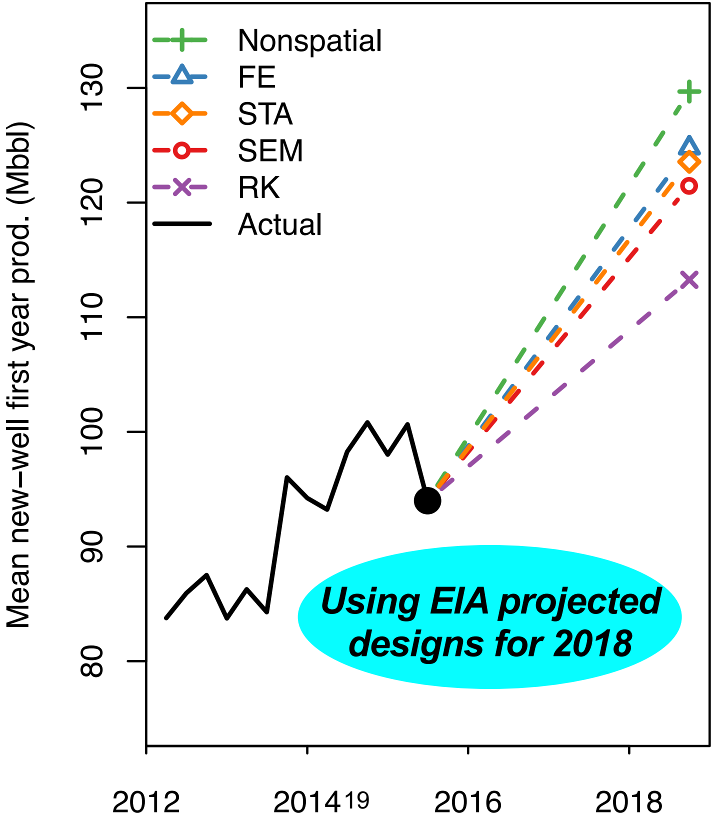
Each model provides a good fit to the mean productivity over time



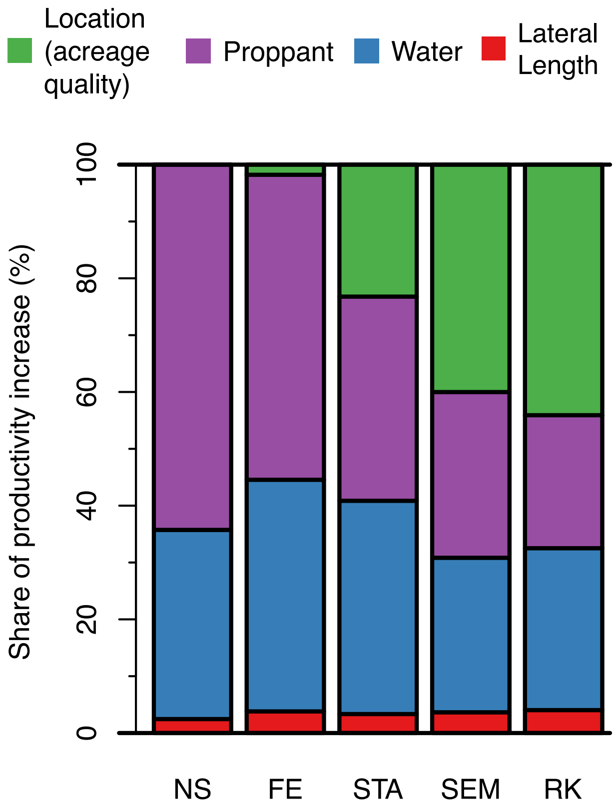
Training only with data from early wells shows that mean production can be reliably forecasted based on changes in location and technology



These models are useful for forecasting production and economics of future wells – Important differences between RK and existing approaches such as FE become clear

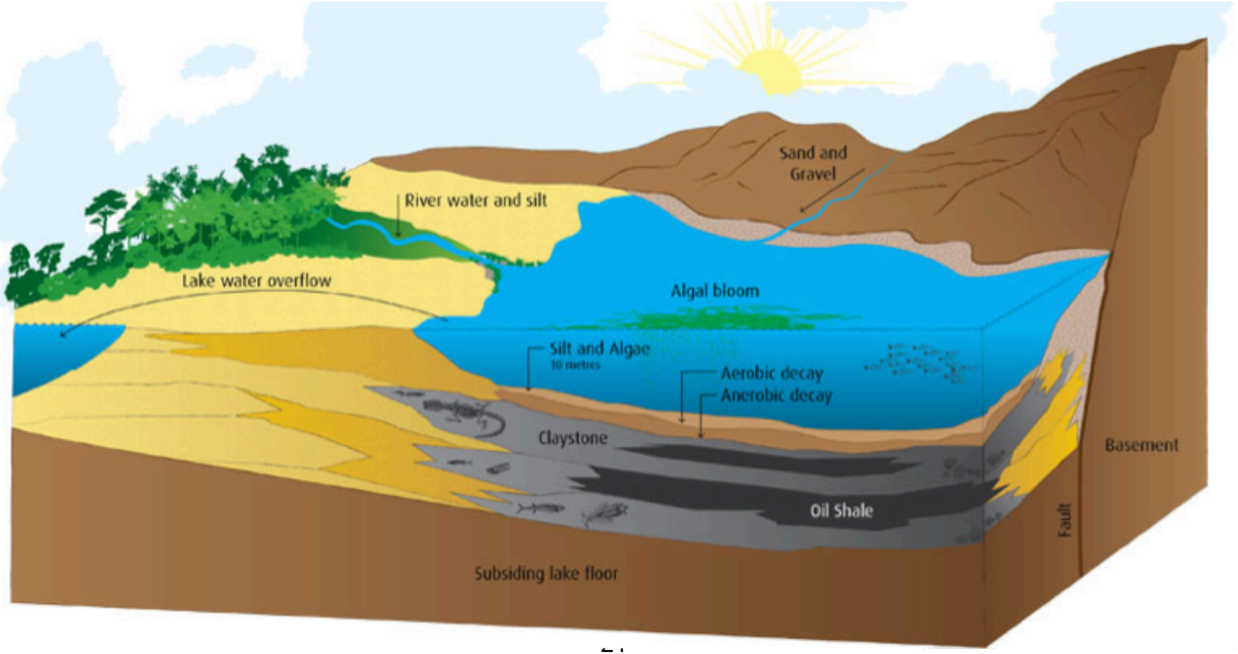


Differences in impact attributed to different parameters

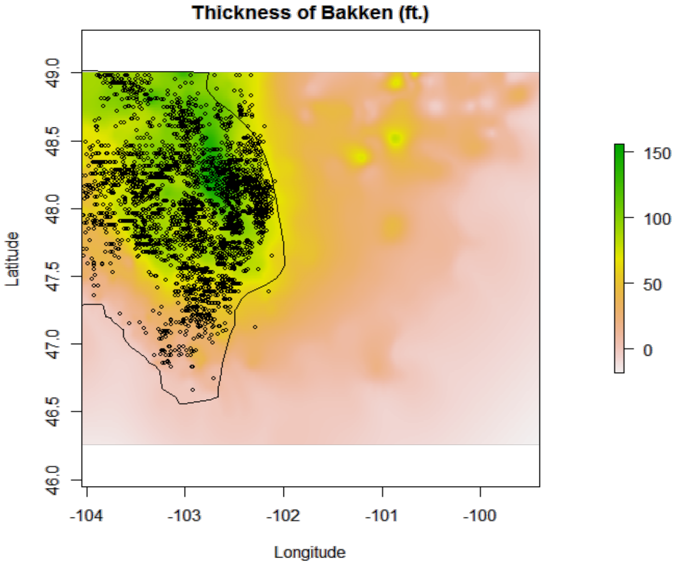


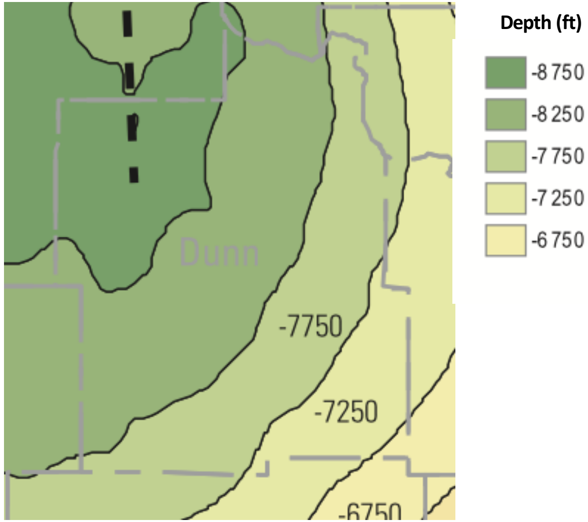
Location is important because key geological controls on production vary spatially across basin

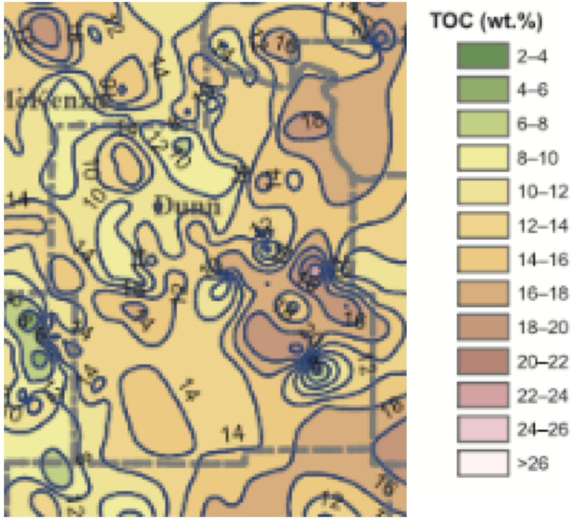
- Spatial trends and patterns result from physical processes over long lengths of time
  - Occur at various scales (e.g. macro: formation thickness, grain size/porosity, thermal maturity; micro: natural fractures)
- Geological controls may be poorly understood or hard to quantify



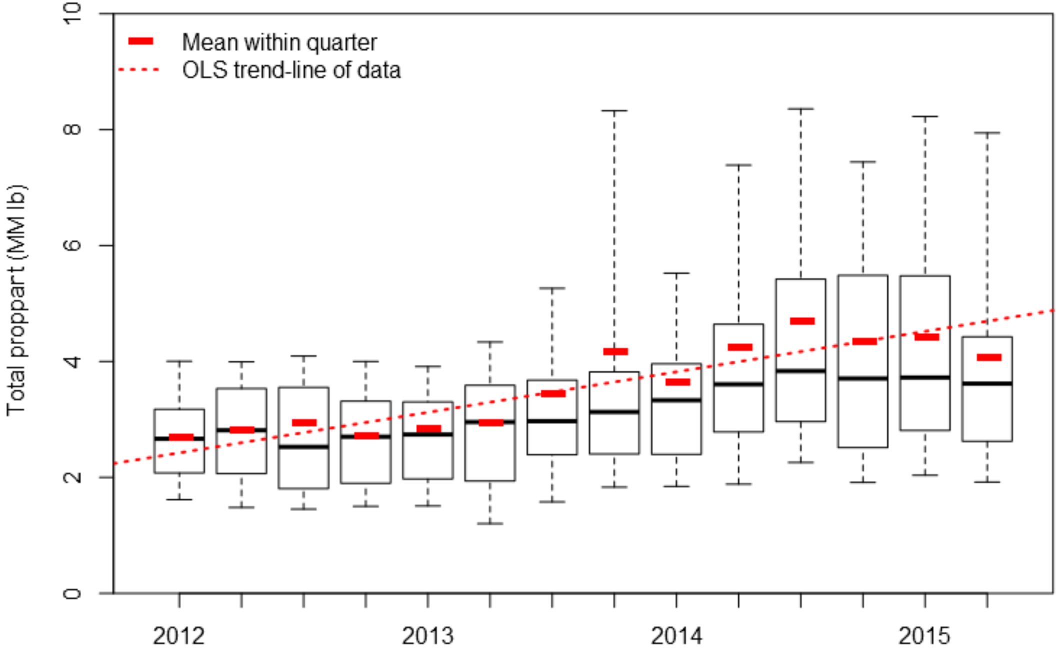
Location is important because key geological controls on production vary spatially across basin

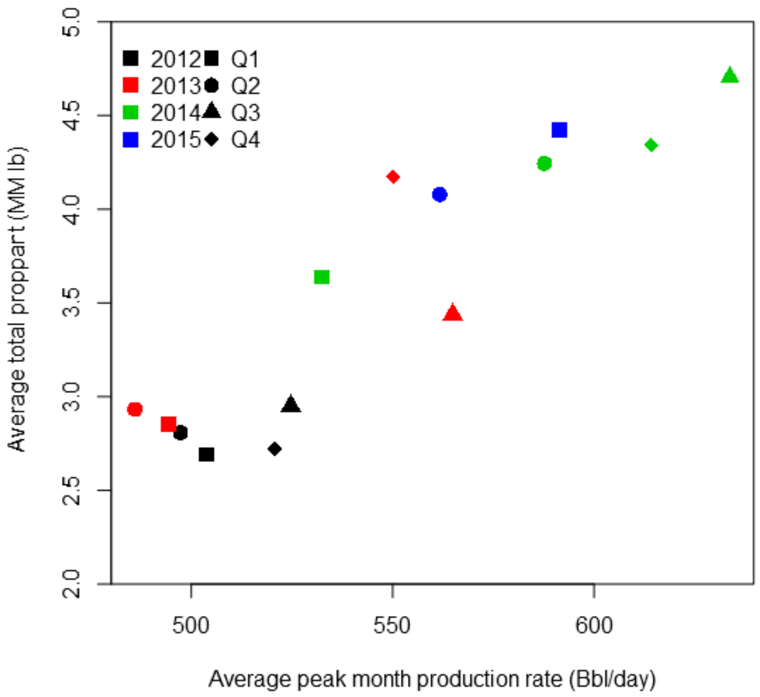


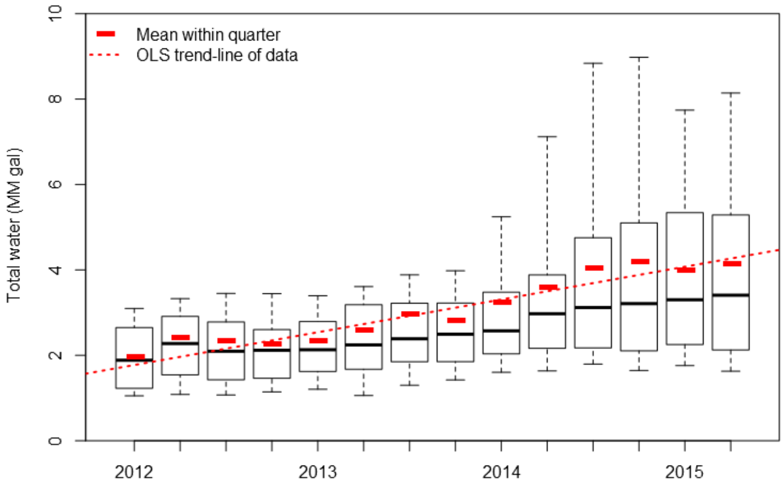




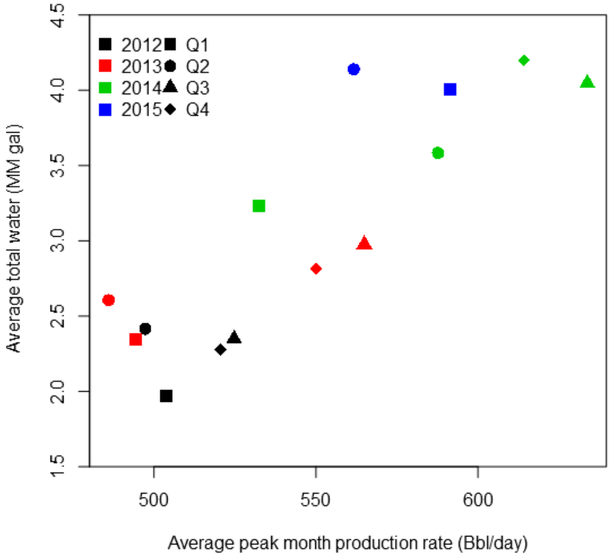
Amount of proppant has been increasing over time and is correlated with productivity

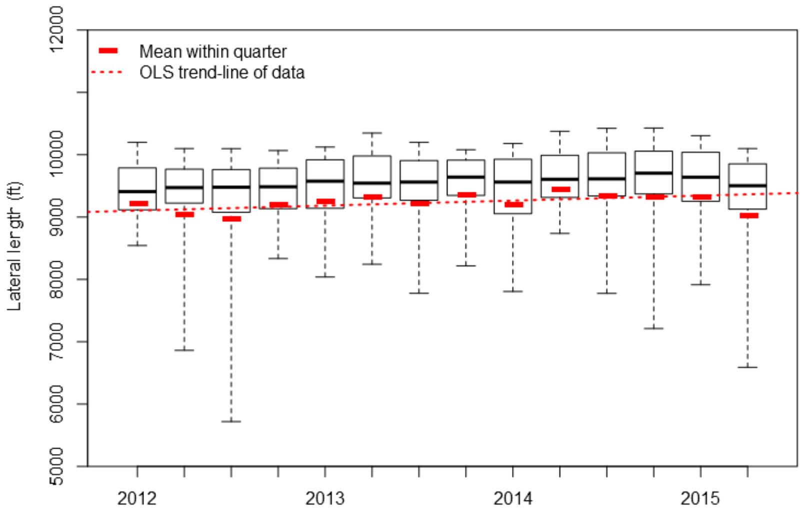




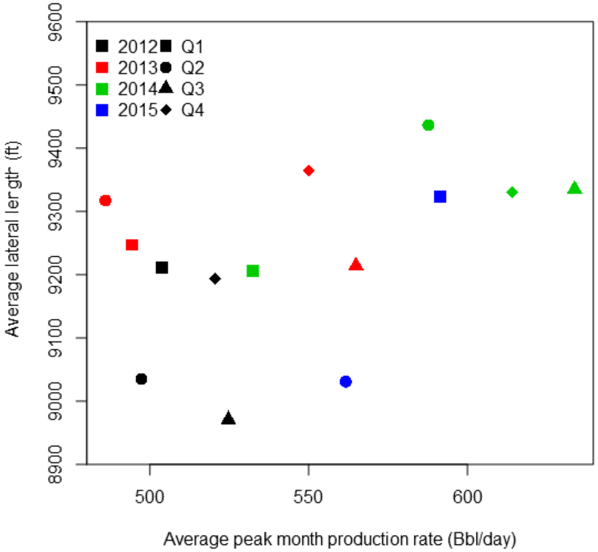


Water trends





Lateral length trends



Definition of models:

Multiple linear regression model:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I_n})$$

Ordinary least squares:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

Multiple linear regression model with variance-covariance matrix:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Omega})$$

Generalized least squares:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\Omega^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\Omega^{-1}\boldsymbol{Y}$$

	NS	FE	STA	SEM	RK
Form	$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$	$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$	$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$	$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \rho \boldsymbol{W}\boldsymbol{e} + \boldsymbol{u}$	$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\lambda}\boldsymbol{e} + \boldsymbol{u}$
Technology variables in $\boldsymbol{X}$	Lateral length, water volume, proppant mass	Lateral length, water volume, proppant mass	Lateral length, water volume, proppant mass	Lateral length, water volume, proppant mass	Lateral length, water volume, proppant mass
Additional variables in $\boldsymbol{X}$ to control for location	N/A	County indicators, formation indicator	Second order polynomial of coordinates, formation indicator	Second order polynomial of coordinates, formation indicator	Second order polynomial of coordinates, formation indicator
Fitted parameters to control for spatial autocorrelation	N/A	N/A	N/A	ρ	τ^2, σ^2, ϕ
Decay of spatial autocorrelation assumed	N/A	N/A	N/A	Inverse distance weighting, first 50 neighbors only	Exponential

Table 1: Summary of the regression models used.

One approach to estimating the effect of technology on productivity is linear regression with ordinary least squares – Omitted-variable bias is a problem though

Technology

Productivity $\boxed{Y = X\beta + \epsilon}$

Geological controls (omitted variable)

More realistically: $\boxed{Y = X\beta + Z\theta + \epsilon}$

Bias of Estimate:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$
$$= (X^T X)^{-1} X^T (X\beta + Z\theta + \epsilon)$$

$$E[\hat{\beta}|X] = \beta + \underbrace{(X^T X)^{-1} E[X^T Z|X]\theta}$$

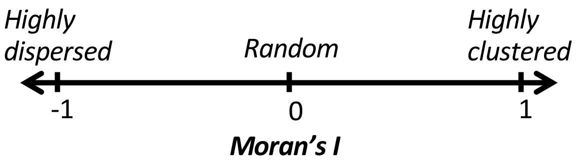
Bias is introduced if:

$$\theta \neq 0$$
$$\mathrm{cov}(X,Z) \neq 0$$

Evaluating the models

Moran's I to measure spatial autocorrelation:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(\epsilon_i - \bar{\epsilon})(\epsilon_j - \bar{\epsilon})}{\sum_i (\epsilon_i - \bar{\epsilon})^2}$$



Back transformation:

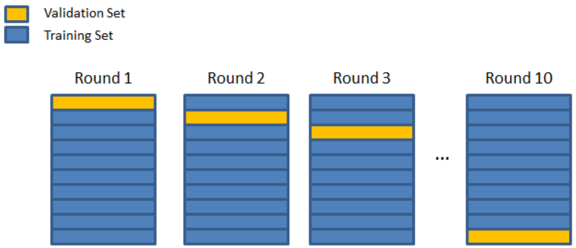
$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{n} ((Y_i - Y_i^*)^2)$$

$$Q_i = \exp(Y_i + \hat{\sigma}^2/2) \quad i = 1, \ldots, n$$

Model accuracy:

$$\mathrm{MASE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|Y_i - Y_i^*|}{|Y_i - \bar{Y}|}$$

10-fold cross validation:

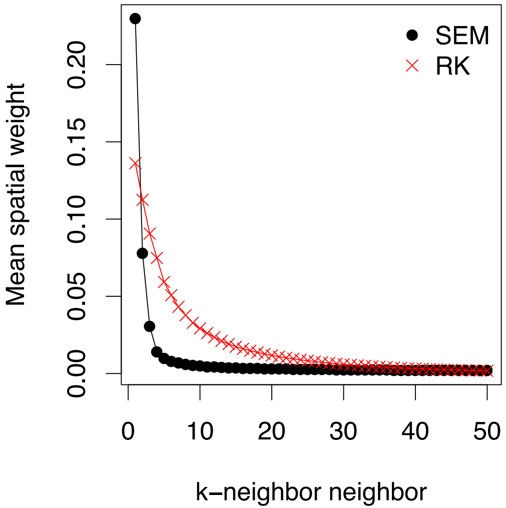


Comparison of models:

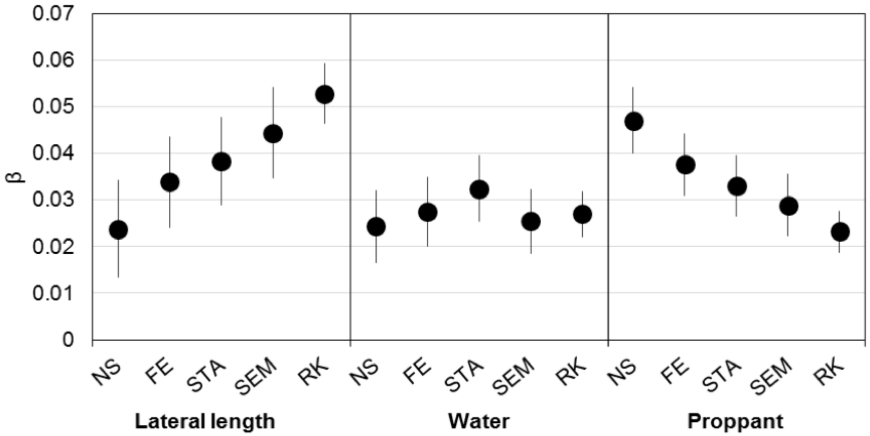
	NS	FE	STA	SEM	RK
Time to run (s)	0.0041	0.00804	0.00828	3.7	478.54
MASE	0.938	0.871	0.815	0.662	0.532
10-fold CV MASE	0.938	0.873	0.816	0.669	0.62
Moran's I ($\boldsymbol{W}$)	0.512	0.443	0.403	-0.00895	-2.26E-04
Moran's I ($\boldsymbol{\lambda}$)	0.548	0.482	0.444	0.245	0.102

Table 2: Comparison of performance for the regression models. Moran's I was calculated with both the inverse distance weighted matrix $\boldsymbol{W}$ and the kriging weights matrix $\boldsymbol{\lambda}$.

Spatial weights matrix W:



Coefficient estimates:



Comparison of models' estimates of technology and location driven improvement in productivity