# Bayesian hierarchical modeling: application towards production results in the Eagle Ford Shale of South Texas

Se Yoon Lee

Texas A & M University

February 27, 2021

# Motivation of Research: World Map of Shale Oil and Gas

The United States Energy Information Administration (US EIA) estimated that technically recoverable shale gas resources in the world are **7201 trillion cubic feet** (Annual Energy Outlook 2018).

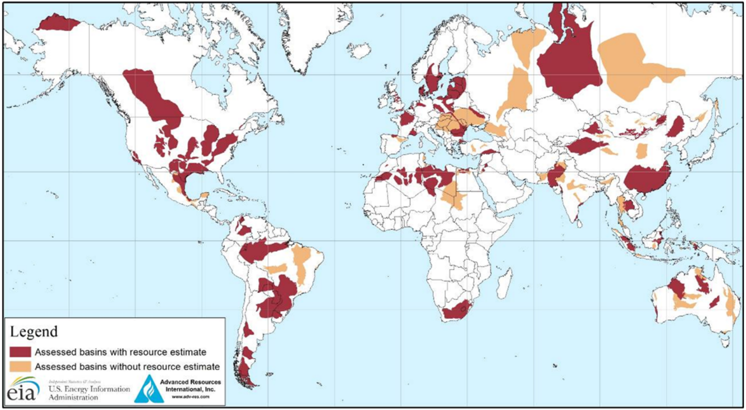
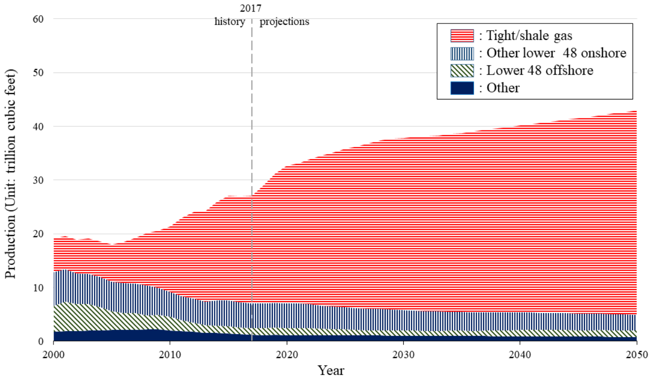



Figure: Basins with assessed shale oil and gas formations in 41 other countries.

# United States energy independence: Historic and expected natural gas production by sources in the United States




- **Shale oil production remains the leading source of US crude oil production from 2017 to 2050 (US EIA).**

# The Shale Revolution: horizontal drilling and hydraulic-fracturing technologies

- The increase of oil and gas production in the mid-2000s was possible due to widespread utilization, advance, and conjunction of horizontal drilling and hydraulic-fracturing technologies.
- In 2016, hydraulically fractured horizontal wells accounted for 69% of all oil and gas wells drilled in the US.
- The US has been a net energy importer since 1953.
- The US is expected to be a net energy exporter by the early 2020s.
- The Shale Revolution: American path to energy independence.

# Hydraulically fractured horizontal well

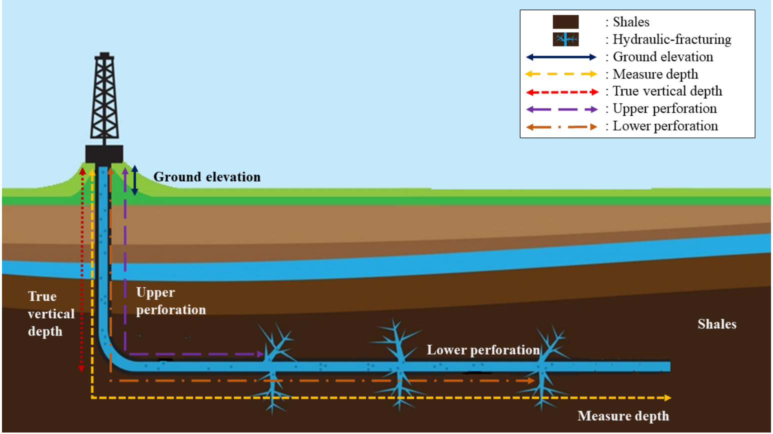



Figure: A pictorial example of a hydraulically fractured horizontal well. Architecture of hydraulic-fracturing design components is important in well productivity.

# Petroleum Industry & Big Data

The oil and gas industry has faced the Era of Big Data in their quest to better understand a well productivity of unconventional reservoirs.

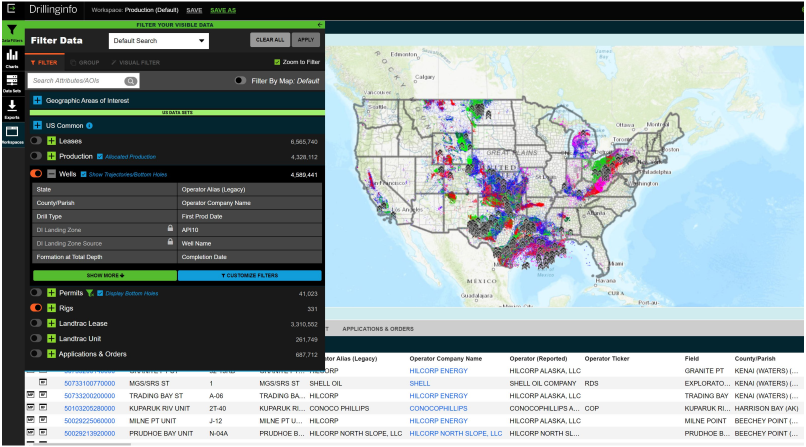
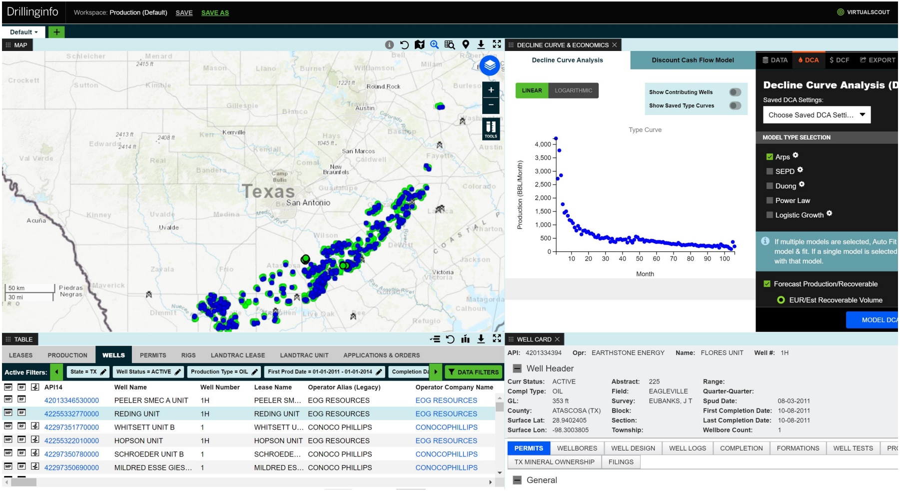
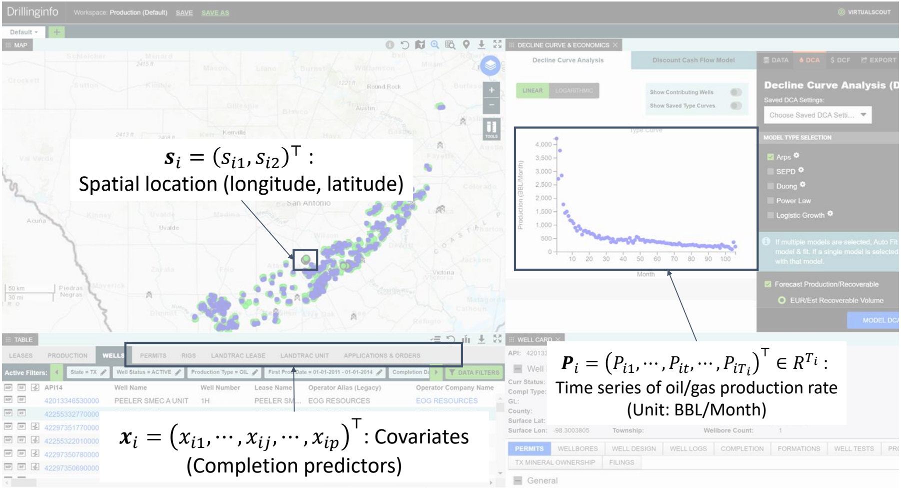



Figure: Screenshot of Drilling Info Database System.

# Shale oil wells data in Drilling Info (Drillinginfo.com)



# Shale oil wells data in Drilling Info (Drillinginfo.com)




*Key idea*: in petroleum engineering, **'well completion'** means the process of transforming a well ready for the initial production.

Shale oil wells data can be called **'spatial functional data'**.

# Characteristics of shale oil wells data: $\{(\mathrm{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$

Most of the characteristics are investigated by Lewis et al (1918) through empirical studies:

## Oil production rate time series data: $\{\mathrm{P}_{\boldsymbol{i}}\}_{i=1}^N$

- $\mathrm{P}_i = (\mathrm{P}_{i1}, \cdots, \mathrm{P}_{iT_i})^\top \in \mathbb{R}^{T_i}$
- All the wells are rarely drilled at once. (If $i \neq j$, then $\mathrm{P}_{i1}$ and $\mathrm{P}_{j1}$ can be measured at different months).
- Production period of each well can be different. (If $i \neq j$, then $T_i$ and $T_j$ can be different).
- Ordinarily a well reaches its maximum production, $\max_{t=1,\cdots,T_i} \mathrm{P}_{it}$, $i = 1, \cdots, N$, within a few months after its completion.
- Eventually the production $\mathrm{P}_{it}$, $t = 1, \cdots, T$, becomes so small over time and it becomes unprofitable.

## Completion predictors: $\{\mathbf{x}_i\}_{i=1}^N$

- $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^\top \in \mathbb{R}^p$
- Completion take places before the initial production, $\mathrm{P}_{i1}$, $i = 1, \cdots, N$. Formally, completion means the process of making a well ready for the first production from the well. Therefore, the completion data, $\mathbf{x}_i^\top$, can used as predictors while $\mathrm{P}_i$ used as response.
- The number of completion data, $p$, can be as large as 100.
- Identifying a subset of completion data, $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^\top$, which is significant in well productivity or future production is an important task.

## Spatial location: $\{\mathbf{s}_i\}_{i=1}^N$

- $\mathbf{s}_i = (s_{i1}, s_{i2})^\top \in \mathbb{R}^2$ : (longitude, latitude)
- Locations of wells, $\{\mathbf{s}_i\}_{i=1}^N$, usually form a band where oil or gas reserves exists underground.

# A Shale Reservoir Problem

Based on the characteristics of shale oil wells data $\{(\mathrm{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^{N}$, we suggest a '*shale reservoir problem*' useful at upstream petroleum industry.

## Three goals of a shale reservoir problem

1. **(a) Curve fitting**: uncover a hidden pattern from the time series data $\mathrm{P}_i = (\mathrm{P}_{i1}, \cdots, \mathrm{P}_{iT_i})^\top$ for each well $i$ ($i = 1, \cdots, N$) through some parametric curve $q(t; \boldsymbol{\vartheta}_i)$ involving the subject-specific parameter $\boldsymbol{\vartheta}_i$. The $\boldsymbol{\vartheta}_i$ can be multidimensional, and its component has its own interpretation explaining well productivity;
2. **(b) Covariates analysis**: identify significant completion predictors among the $p$ predictors $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^\top$ explaining a well productivity described by each component of the vector $\boldsymbol{\vartheta}_i$;
3. **(c) Spatial prediction**: predict beforehand the temporal profile of oil production rate of a new well during $T^*$ months, denoted by $\mathrm{P}^* = (\mathrm{P}_1^*, \cdots, \mathrm{P}_{T^*}^*)^\top$, provided that $p$ completion predictors $\mathbf{x}^* = (x_1^*, \cdots, x_p^*)^\top$ and a new location $\mathbf{s}^* = (s_1^*, s_2^*)^\top$ are specified.

# Some literature review on a *shale reservoir problem*

Consider shale oil wells data $\{(\mathrm{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$ from a shale reservoir.

1. Sub-goal **(a)** based on $\{\mathrm{P}_i\}_{i=1}^N$: widely called decline curve analysis (DCA);
   Recently, there arises a need to quantify uncertainty associated with DCA. This task is even more important in analyzing unconventional reservoir.
   1. Bootstrapping method: Jochen (1996) and Cheng (2010).
   2. Bayesian method: Gong (2011, 2014) and Zhang (2015).

# Some literature review on a *shale reservoir problem*

Consider shale oil wells data $\{(\mathrm{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$ from a shale reservoir.

1. Sub-goal **(a)** based on $\{\mathrm{P}_i\}_{i=1}^N$: widely called decline curve analysis (DCA);
   Recently, there arises a need to quantify uncertainty associated with DCA. This task is even more important in analyzing unconventional reservoir.
   1. Bootstrapping method: Jochen (1996) and Cheng (2010).
   2. Bayesian method: Gong (2011, 2014) and Zhang (2015).
2. Sub-goals **(a)** – **(b)** based on $\{(\mathrm{P}_i, \mathbf{x}_i)\}_{i=1}^N$: Vyas Aditya (2017) researched on statistical modeling to accommodate $\{(\mathrm{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$, but treated $\mathbf{s}_i$ as covariates, implying that $\mathbf{s}_i \in \mathbf{x}_i$.

# Some literature review on a *shale reservoir problem*

Consider shale oil wells data $\{(\mathrm{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^{N}$ from a shale reservoir.

1. Sub-goal **(a)** based on $\{\mathrm{P}_i\}_{i=1}^{N}$: widely called decline curve analysis (DCA);
   Recently, there arises a need to quantify uncertainty associated with DCA. This task is even more important in analyzing unconventional reservoir.
    1. Bootstrapping method: Jochen (1996) and Cheng (2010).
    2. Bayesian method: Gong (2011, 2014) and Zhang (2015).
2. Sub-goals **(a)** – **(b)** based on $\{(\mathrm{P}_i, \mathbf{x}_i)\}_{i=1}^{N}$: Vyas Aditya (2017) researched on statistical modeling to accommodate $\{(\mathrm{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^{N}$, but treated $\mathbf{s}_i$ as covariates, implying that $\mathbf{s}_i \in \mathbf{x}_i$.
3. Sub-goals **(a)** – **(c)** based on $\{(\mathrm{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^{N}$: Jaeyoung Park (2020) researched on forecasting $\mathrm{P}_i$ at new well location, but the location data $\mathbf{s}_i$ is not treated as index of a geo-statistical stochastic process such as Gaussian process, lacking fully Bayesian sense.

# Goal of this research

1. We develop a fully Bayesian model to fulfill the sub-goals **(a)** – **(c)** to accommodate the shale oil wells data $\{(\mathrm{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^{N}$.
2. One of the challenges is about how we can treat the location set $S = \{\mathbf{s}_i\}_{i=1}^{N}$ as an index set of a geo-statistical stochastic process to achieve the sub-goal **(c)**.
3. At the same time, we want a proposed model to be consistent with decline curve analysis applications to be aligned with industrial needs.
4. Eventually, we will propose a Bayesian hierarchical model that
   (i) canonically extends the decline curve analysis,
   (ii) covariates analysis is possible by a sparse favoring prior,
   (iii) and spatial prediction is enabled by a spatial random effect induced by a Gaussian process.

# Research region: Eagle Ford Shale Play

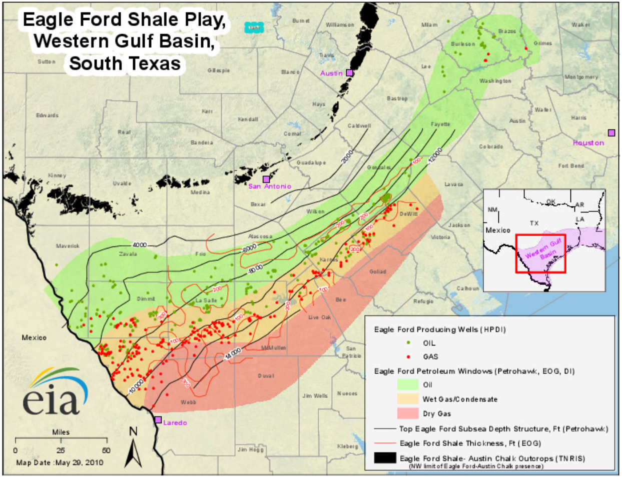
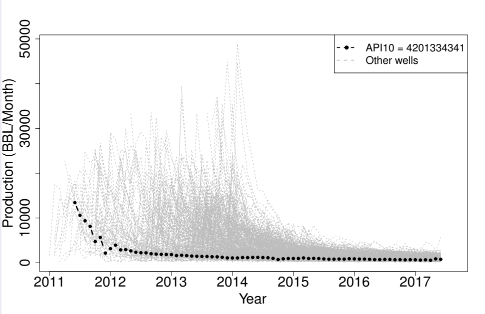



Figure: Eagle Ford region with three types of petroleum windows.

# The well data: $\{\mathrm{P}_{\boldsymbol{i}}, \mathbf{x}_i^\top, \mathbf{s}_i\}_{i=1}^N$

- $N = 360$, $p = 11$
- The time frame of the production time series data, $\{\mathrm{P}_i = (\mathrm{P}_{i1}, \cdots, \mathrm{P}_{iT_i})\}_{i=1}^{N=360}$, spans from January 2011 through June 2017.
- Minimum production period of 360 wells is 42 months, that is, $\min_{i=1,\cdots,360} T_i = 42$, and maximum production period of 360 wells is 78 months, that is, $\max_{i=1,\cdots,360} T_i = 78$.



## 11 well completion predictors $\mathbf{x} = (x_1, \cdots, x_j, \cdots, x_{11})^\top$

| $j$ | Predictors (Unit) | Explanation |
|---|---|---|
| 1 | First test flowing tubing pressure (psi) | Pressure at the wellhead during the initial potential test in an oil well |
| 2 | Ground elevation (ft) | Elevation in respect to ground level |
| 3 | Measured depth (ft) | Total depth the property was drilled to |
| 4 | True vertical depth (ft) | Total true vertical depth of intended bottom hole |
| 5 | Upper perforation (ft) | Upper perforation of the producing property |
| 6 | Lower perforation (ft) | Lower perforation of the producing property |
| 7 | Perforated interval length (ft) | Length of perforated interval |
| 8 | Completion count (integer) | Number of completions performed on that well |
| 9 | First test oil volume (bbl) | Calculated oil production during initial potential test |
| 10 | First test oil gravity | Measurement of the relative density of petroleum liquid to water during the first test |
| 11 | Abstract | Abstract number |

Table: Unit of abstract is blocks of land units in Texas. It applies to Texas only.

## Well locations $\{\mathbf{s}_i = (s_{i1}, s_{i2})^\top\}_{i=1}^N$

Spatial location of the $i$-well $(s_{i1}, s_{i2})^\top$ represents (longitude, latitude) coordinate, recorded in decimal degrees used in WGS84 coordinate reference system.

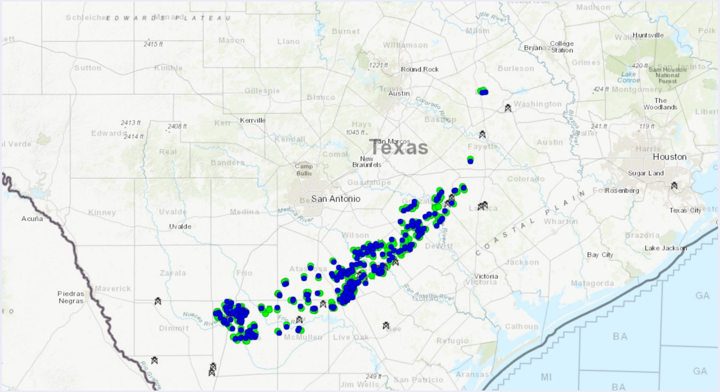



Figure: Spatial locations of the 360 wells

# Non-spatial Bayesian hierarchical models

Consider $\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^N$ where $\mathbf{y}_i = (y_{i1}, \cdots, y_{iT_i})^\top$ is a log-scaled oil production rate trajectory for the $i$-th well. ($y = \log\ \mathrm{P}$)

- Individual-level model

$$y_{it} = \mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2).$$

- Population model

$$\begin{gathered}\theta_{li} = \alpha_l + \mathbf{x}_i^\top \boldsymbol{\beta}_l + \epsilon_{li}, \quad \epsilon_{li} \sim \mathcal{N}(0, \sigma_l^2), \\ \beta_{lj}|\lambda_{lj}, \tau_l, \sigma_l \sim \mathcal{N}(0, \lambda_{lj}^2 \tau_l^2 \sigma_l^2),\ \lambda_{lj} \sim \mathcal{C}^+(0,1),\ \tau_l \sim \mathcal{C}^+(0,1), \\ \sigma \sim \mathcal{C}^+(0,1),\ \alpha_l \sim \pi(\alpha_l) \propto 1,\ \sigma_l^2 \sim \pi(\sigma_l^2) \propto 1/\sigma_l^2.\end{gathered}$$

The indices $i$, $t$, $l$, and $j$ take $i \in \{1, \cdots, N\}$, $t \in \{1, \cdots, T_i\}$, $l \in \{1, 2, 3\}$, and $j \in \{1, \cdots, p\}$, respectively.

# Rate decline curve (RDC)

Mean part of the individual-level model is a log-scaled rate decline curve

$$\mu(t;\theta_{1i},\theta_{2i},\theta_{3i}) = \log\ q(t;M_i,b_i,k_i), \theta_{1i} = \log\ M_i, \theta_{2i} = \log\ b_i, \theta_{3i} = \log\ k_i.$$

We consider four choices:

$$\begin{aligned}
\text{Weibull } (\mathcal{M}_1): &\quad q(t|M,b,k) = M \cdot \text{Weibull}(t|b,k) = M \cdot bkt^{k-1}\exp(-bt^k) \\
\text{Arps } (\mathcal{M}_2): &\quad q(t|M,b,k) = M \cdot \text{GPD}(t|b,k) = M \cdot (1/b)(1+kt/b)^{-1/k-1} \\
\text{SEDM } (\mathcal{M}_3): &\quad q(t|M,b,k) = M \cdot \exp(-bt^k) \\
\text{Duong } (\mathcal{M}_4): &\quad q(t|M,b,k) = M \cdot \text{Fréchet}(t;b,k) = M \cdot bkt^{-1-k} exp(-bt^{-k})
\end{aligned}$$

- Parameters $M$, $b$ and $k$ are positive real numbers.
- Roughly speaking, statistically, they are interpreted as maximum value, scale parameter (or rate), and shape parameter, respectively.
- In petroleum engineering, they have their own interpretations in terms of well productivity; refer to Mishra (2012), Arps (1945), Valko (2010), and Duong (2011).

# Example: Bayesian decline curve analysis

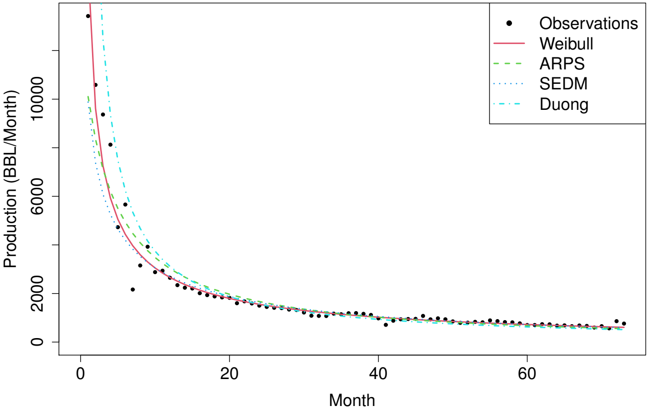



Figure: Bayesian decline curve analysis based on the four models.

# Model choices

To choose the best model among the four non-spatial models, we used diviance information criterion (DIC), widely applicable information criterion (WAIC), and posterior predictive loss criterion (PPLC). The smaller numbers are indications for better predictive performance.

Table: $\mathrm{DIC}_{\mathrm{G}}$, WAIC, and PPLC for the four non-spatial models

| | $\mathrm{DIC}_{\mathrm{G}}$ | WAIC | PPLC |
|---|---|---|---|
| Weibull ($\mathcal{M}_1$) | 18977 | 18587 | 4465 |
| Arps ($\mathcal{M}_2$) | 19554 | 19069 | 4574 |
| SEDM ($\mathcal{M}_3$) | 22014 | 19338 | 4634 |
| Duong ($\mathcal{M}_4$) | 45819 | 31707 | 8398 |

- Weibull model ($\mathcal{M}_1$) is selected as the best model with respect to the three model comparison criteria.
- The results from next slide are based on the Weibull model.

# Spatial Bayesian hierarchical models

Consider the shale oil wells data $\{\mathbf{y}_i, \mathbf{x}_i, \mathbf{s}_i\}_{i=1}^N$.

- Individual-level model

$$y_{it} = \mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2).$$

- Population model

$$\begin{aligned}
&\theta_{li} = \theta_l(\mathbf{s}_i) = \alpha_l + \mathbf{x}_i^\top \boldsymbol{\beta}_l + \epsilon_l(\mathbf{s}_i) + \eta_l(\mathbf{s}_i), \quad \epsilon_l(\cdot) \sim \mathcal{GWN}(\sigma_l^2),\\
&\eta_l(\cdot) \sim \mathcal{GP}(0, \mathcal{K}_{\gamma_l}(\cdot,\cdot)), \quad \mathcal{K}_{\gamma_l}(\mathbf{s}_i, \mathbf{s}_j) = \gamma_l^2 \exp(-e^{\rho_l} \|\mathbf{s}_i - \mathbf{s}_j\|_2^2),\\
&\beta_{lj} | \lambda_{lj}, \tau_l, \sigma_l \sim \mathcal{N}(0, \lambda_{lj}^2 \tau_l^2 \sigma_l^2),\\
&\sigma, \gamma_l, \lambda_{lj}, \tau_l \sim \mathcal{C}^+(0,1),\\
&\alpha_l \sim \pi(\alpha_l) \propto 1, \quad \sigma_l^2 \sim \pi(\sigma_l^2) \propto 1/\sigma_l^2.
\end{aligned}$$

The indices $i$, $t$, $l$, and $j$ take $i \in \{1, \cdots, N\}$, $t \in \{1, \cdots, T_i\}$, $l \in \{1, 2, 3\}$, and $j \in \{1, \cdots, p\}$, respectively. $\rho_l$ is hyper-parameter.

# Some literature review on modeling framework

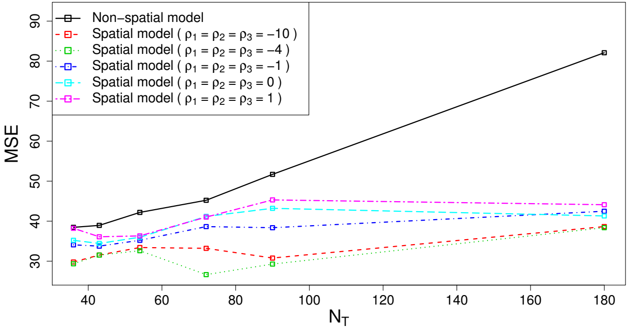
1. The platform of the proposed spatial is called the nonlinear mixed effect model, or hierarchical nonlinear model (Davidian and Gitinan, 2003).
2. This framework has been widely used in biological, agricultural, environmental, and medical applications, including COVID-19 outbreak problem (Lee, Lei, and Mallick, 2020).
3. Typically, one of the main motivations to employ this framework is to fulfill the sub-goals (a) (curve fitting) and (b) (covariates analysis) across diverse industrial problems.
4. A prominent feature of the proposed spatial model is the introduction of the spatial random effect on the population level to achieve the sub-goal (c) (spatial prediction).

# Results:Spatial prediction versus non-spatial prediction

- The ranges $\rho_l$ ($l = 1, 2, 3$) are identically set by -10, -4, -1, 0, and 1.
- Number of test wells among 360 wells: 36, 43, 54, 72, 90, and 180.
- We report mean squared error (MSE)

$$\mathrm{MSE}_{\mathrm{N_{test}}} = \frac{1}{\mathrm{N}_T} \sum_{i=1}^{\mathrm{N_{test}}} \|\widehat{\mathbf{y}}_{i,\mathrm{median}}^{\mathcal{P}} - \mathbf{y}_i\|_2^2, \quad , N_{\mathrm{test}} = 36, 43, 54, 72, 90, \&180$$

$\widehat{\mathbf{y}}_{i,\mathrm{median}}^{\mathcal{P}}$ : Posterior predictive median, $\mathbf{y}_i$ : Observation



# Bayesian decline curve analysis at individual/reservoir level

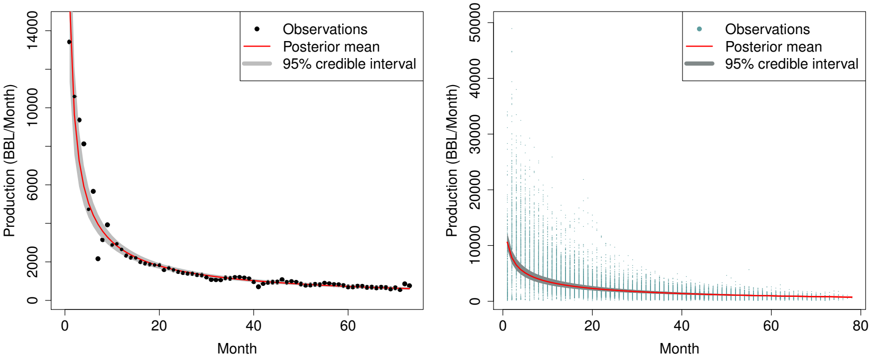



Figure: Estimation results of Bayesian decline curve analysis at an individual well (left) and reservoir levels (right). Shaded regions on both panels represent the pointwise posterior 95% credible intervals.

# Covariates analysis via horseshoe prior

On the second stage, we used the sparse horseshoe prior on the $\boldsymbol{\beta}_l$

$$\theta_{li} = \theta_l(\mathbf{s}_i) = \alpha_l + \mathbf{x}_i^\top \boldsymbol{\beta}_l + \epsilon_l(\mathbf{s}_i) + \eta_l(\mathbf{s}_i), \quad l = 1, 2, 3.$$

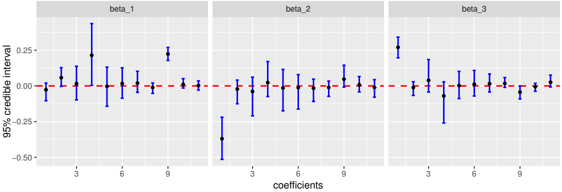



Figure: Posterior 95 % credible intervals for the coefficients: $\boldsymbol{\beta}_1$ (left), $\boldsymbol{\beta}_2$ (middle), and $\boldsymbol{\beta}_3$ (right). The symbol $\bullet$ represents the posterior mean.

As for the carrying capacity, **true vertical depth** and **the first test oil volume** are selected as important well completion predictors. Commonly for the scale and shape parameters, the **first test flowing tubing pressure** has been found to be the most important completion predictor.

# Spatial predictions at two new locations

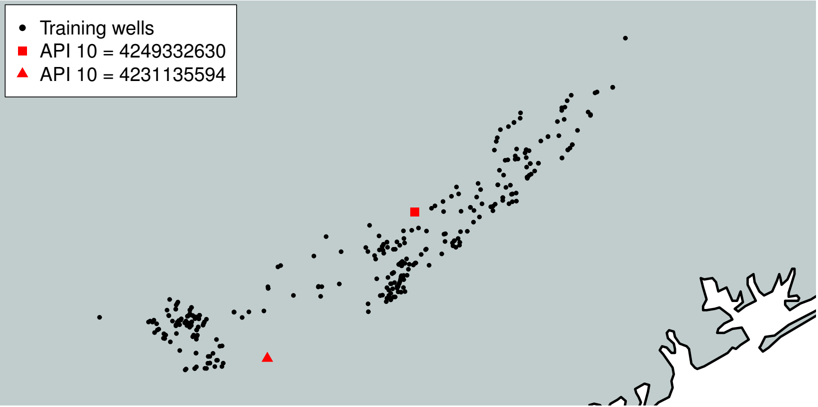



Figure: Locations of 324 training wells and two test wells. The symbols ■ and ▲ represent the locations of well-1 (API10=4249332630) and well-2 (API10=4231135594), respectively.

# Spatial predictions at two new locations

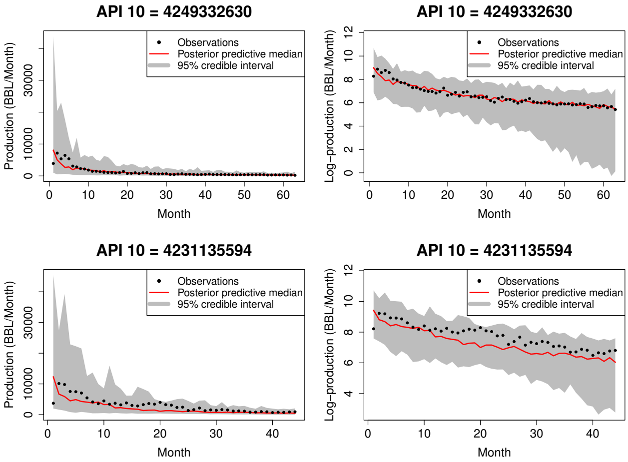



Figure: Spatial predictions for two test wells: well-1 (top) and well-2 (bottom). The left and right panels are displayed in the original and log-transformed production rate scales, respectively.

# Spatial predictions for Estimated Ultimate Recovery

Estimated Ultimate Recovery (EUR) is defined as a sum of monthly oil production rate over 30 year period (360 months).

Table: Summaries of EUR$^*$s for the two test wells.

| API10 | Posterior mean | Posterior median | 95% credible interval |
|---|---|---|---|
| 4249332630 | 128013.5 | 126391.9 | (58340.3, 226077.5) |
| 4231135594 | 167384.9 | 165041.3 | (46443.1, 317699.5) |

# Conclusion & future researches

- The proposed spatial model outperformed its non-spatial counterpart.
- This elucidates that a reasonable exploitation of spatial information can bring advantage to better forecast a future behavior of oil or gas production before actual drilling takes place.
- Some future researches are as follow:
  (i) use correlated linear regressions on the second stage instead of separate regressions;
  (ii) fully Bayesian estimation for the range parameters;
  (iii) use various correlation functions such as the Matérn correlation function or an anisotropic covariance function;
  (iv) use variational methods to enhance computational speed;
  (v) accommodate time-varying covariates (time series of water/gas production rate);

# Core references


1. Lewis, J.O. and Beal, C.H., 1918. Some New Methods for Estimating the Future Production of Oil Wells. Transactions of the AIME, 59(01), pp.492-525.
2. Davidian, M. and Giltinan, D.M., 2003. Nonlinear models for repeated measurement data: an overview and update. Journal of agricultural, biological, and environmental statistics, 8(4), p.387.

Thank you very much