Sankhyā : The Indian Journal of Statistics
2022, Volume 84-B, Part 1, pp. 1–43

Bayesian Hierarchical Modeling: Application Towards Production Results in the Eagle Ford Shale of South Texas

Se Yoon Lee and Bani K. Mallick
Texas A & M University, College Station, USA

Abstract

Recently, the petroleum industry has faced the era of data explosion, and many oil and gas companies resort to data-driven approaches for unconventional field development planning. The objective of this paper is to analyze shale oil wells in a shale reservoir and develop a statistical model useful for upstream. Shale oil wells dataset comprises three aspects of information: oil production rate time series data; well completion data; and well location data. However, traditional decline curve analysis only utilizes the temporal trajectory of the production rates. Motivated by this, we propose a Bayesian hierarchical model that exploits the full aspects of the shale oil wells data. The proposed model provides the following three functionalities: first, estimations of a production decline curve at an individual well and entire reservoir levels; second, identification of significant completion predictors explaining a well productivity; and third, spatial predictions for the oil production rate trajectory of a new well provided completion predictors. As a fully Bayesian approach has been adopted, the functionalities are endowed with uncertainty quantification which is a crucial task in investigating unconventional reservoirs. The data for this study come from 360 shale oil wells completed in the Eagle Ford Shale of South Texas.

AMS (2000) *subject classification.* Primary 62F15; Secondary 62H11, 62M20.
Keywords and phrases. Bayesian hierarchical modeling, Decline curve analysis, Shale oil wells, Latent kriging

1 Introduction

1.1. Motivation: the Era of Big Data and the Shale Boom One of the global trends in the oil and gas industry in recent years can be characterized by two keywords: Big Data and Shale Boom. The industry has been accumulated substantial amounts of data during the process of unconventional oil and gas production to better optimize the architecture and completion of wells and, furthermore, to predict future production behaviors of wells

to minimize drilling cost. And it is indisputable that the oil and gas production from shale formation was one of the epicenters that overturned the recent world energy market. The United States Energy Information Administration (US EIA) estimated that technically recoverable shale gas resources in the world are 7201 trillion cubic feet, contained in a survey by US EIA which covered 137 shale formations from 41 countries. (See Annual Energy Outlook 2018 https://www.eia.gov/- for a detail.)

Particularly in the US energy market, unconventional oil and gas sources are rapidly gaining in importance. Figure 1 depicts historic and projected productions of natural gas by different sources. On the panel, it is observed that there manifests a surge in shale gas production since the mid-2000s, and it continuously increases through 2050. It is widely acknowledged that the upsurge was due to the widespread utilization, advance, and conjunction of two techniques: horizontal drilling and hydraulic fracturing technologies (Gandossi and et al. 2013). Figure 2 describes a pictorial example of a hydraulically fractured horizontal well. (Refer Fig 1 in Howarth et al. (2011) for a detailed version.) Horizontal drilling is a directional drilling technology such that a well is drilled parallel to the reservoir bedding plane (Giger et al.,

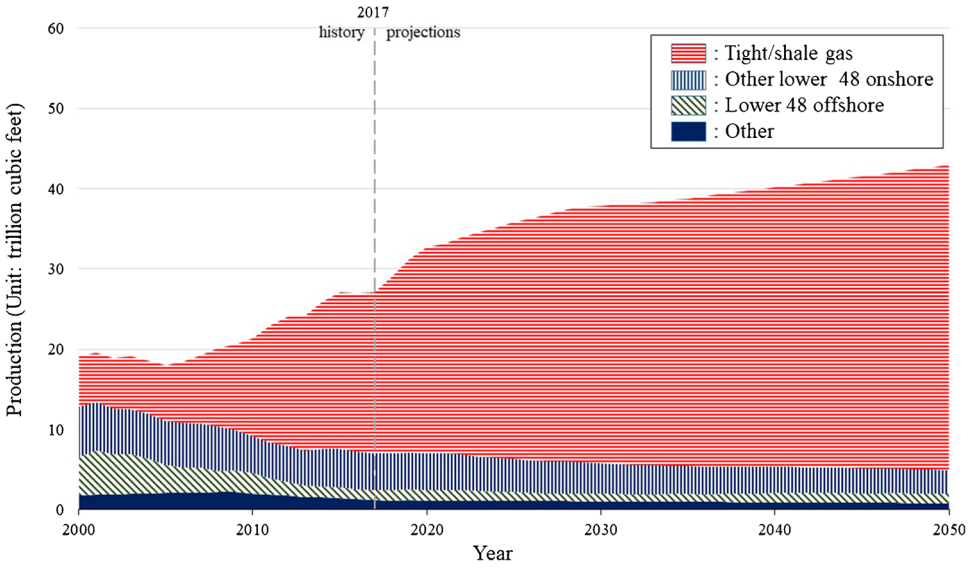

Figure 1: Historic and expected US natural gas production by different sources. Other includes Alaska and coalbed methane. US EIA forecasted that shale oil production would remain as the leading source of US crude oil production from 2017 to 2050

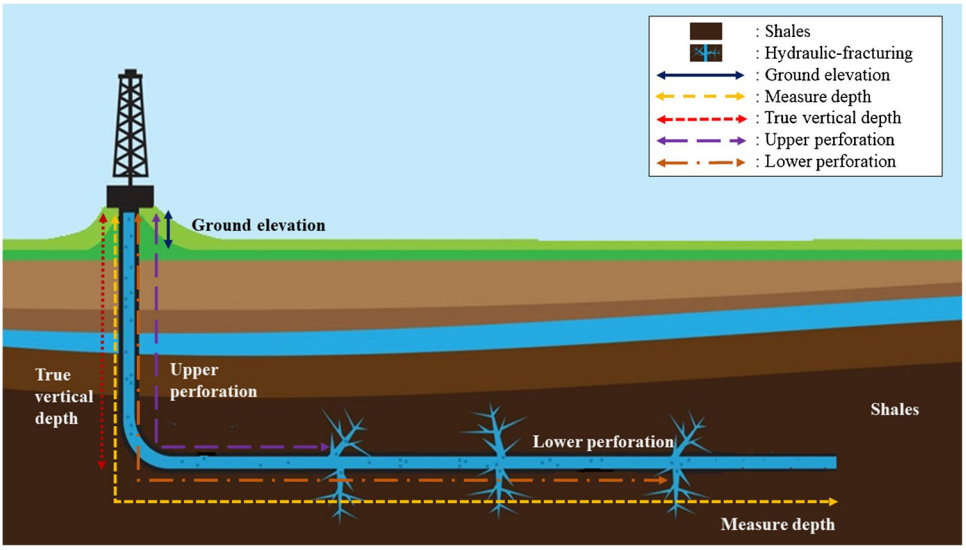

Figure 2: A schematic example of a hydraulically fractured horizontal well

1984). Its main purpose is to increase reservoir contact, eventually improving a well productivity (Joshi, 1991). Well productivity of a horizontal well is known to be often 3 to 5 times greater than that of a vertical well (Al-Haddad et al., 1991; Mukherjee et al., 1991), but also costs 1.5 to 2.5 times more than a vertical well (Joshi and et al. 2003). Hydraulic fracturing, also called fracking, is a well stimulation technique that makes use of water-based fluids to break up nonporous rocks to access oil and gas resources that are locked in the oil-bearing formation so that initial flow and ultimate recovery of oil increase (Hubbert and Willis, 1972; Montgomery et al., 2010).

The US shale boom–a product of technological advances in horizontal drilling and hydraulic fracturing that unlocked new stores of energy–has greatly benefited the growth in the US economy. In October 2011, hydraulically fractured horizontal directional drilling became one of the predominant methods to explore a new US crude oil and natural gas development. In 2016, hydraulically fractured horizontal wells accounted for 69% of all oil and gas wells drilled in the US. As end of 2018, about 6.5 million barrels per day of crude oil were produced directly from shale oil resources which accounts for 59% of total US crude oil produced in 2018 (EIA et al., 2019). The US has been a net energy importer since 1953, but due to a decline in energy imports and an increase in energy exports, the US is projected to be a net energy exporter of crude oil by the early or mid-2020s, thereby achieving The Shale

Revolution, America's path to energy independence. (See www.eia.gov/- for the detailed report from US EIA.)

1.2. A Shale Reservoir Problem The eventual success of petroleum development projects relies on a large degree of well construction costs (Valdes et al., 2013). As for unconventional reservoirs, because of very low permeability, and a flow mechanism very different from that of conventional reservoirs, estimates for the well construction cost often contain high levels of uncertainty, and oil companies need to make heavy investment in the drilling and completion phase of the wells. The overall recent commercial success rate of horizontal wells in the US appears to be 65%, which implies that only 2 out of 3 drilled wells will be commercially successful (Joshi and et al. 2003). For this reason, one of the crucial tasks of petroleum engineers is to quantify the uncertainty associated with the process of oil or gas production to reduce extra initial risk for the projects.

Considerable amounts of data are being generated during the development and operation of shale reservoirs. Here, we explain a shale oil wells dataset from a shale reservoir region, and summarize some commonly agreed characteristics of the dataset. After that, a "*shale reservoir problem*" is suggested that can be useful for the upstream petroleum industry.

Let $\{(\mathsf{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$ denotes a collection of N shale oil wells data from a shale reservoir region such that each of the wells is indexed by i ($i = 1, \cdots, N$). We assume that all the wells are hydraulically fractured and horizontally drilled, and no wells were refractured after initial productions. Technically, production results of an individual well i can be described by 3-tuple $(\mathsf{P}_i, \mathbf{x}_i, \mathbf{s}_i)$ whose components are:

- $\mathsf{P}_i = (\mathsf{P}_{i1}, \cdots, \mathsf{P}_{iT_i})^\top \in \mathbb{R}^{T_i}$: monthly oil production rate time series data over T_i months from its initial production;
- $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^\top \in \mathbb{R}^p$: p number of completion data;
- $\mathbf{s}_i = (s_{i1}, s_{i2})^\top \in \mathbb{R}^2$: spatial location (longitude, latitude) of the well.

In petroleum engineering, well completion means the process of transforming a well ready for the initial production (Bellarby, 2009; Wan, 2011). A collection of informative data involved in the completion procedure is called completion data, represented by the vector $\mathbf{x}_i$. We emphasize that the concept of completion motivates a statistical insight. This is because the time series data $\mathbf{P}_i$ is observed *after* well completion predictors $\mathbf{x}_i$ along with the well location $\mathbf{s}_i$ are determined by petroleum engineers. In other words, for each well i, we can view the relationship between $\mathbf{P}_i$ and $(\mathbf{x}_i, \mathbf{s}_i)$ as an output-and-input relationship: the T_i elements of vector $\mathbf{P}_i = (\mathsf{P}_{i1}, \cdots, \mathsf{P}_{it},$

$\cdots, \mathsf{P}_{iT_i})^\top$ can be regarded as discrete points observed at time points $t = 1, \cdots, T_i$, whereas the completion data $\mathbf{x}_i$ coupled with the spatial location $\mathbf{s}_i$ are served as inputs. By this reason, the collection of the 3-tuples from N wells, $\{(\mathbf{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$, formulates a spatial functional data possessing a nested data structure where both temporal and location information should be simultaneously considered.

The followings are summaries about some agreed characteristics of shale oil wells data $\{(\mathbf{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$. Most of them are investigated by Lewis et al. (1918) through empirical studies:

(i) all the wells are rarely drilled at once (For any two wells indexed by i and j with $i \neq j$, the values P_{i1} and P_{j1} can be measured at different time points. This means that the time index '1' in the subscripts, P_{i1} and P_{j1}, represents the initial production time.);

(ii) production period of each well can be different. (For any two wells indexed by i and j with $i \neq j$, the production periods T_i and T_j can be different.);

(iii) ordinarily a well reaches its maximum production at its initial production time or within a few months after its completion (For a well i, the value $\arg\max_{t=1,\cdots,T_i} \mathsf{P}_{it}$ is either $t = 1$ or some t greater than 1 upper bounded by a small integer.);

(iv) eventually the oil or gas production rate becomes so small over time and it becomes unprofitable (For a well i, it holds $\lim_{t\to\infty} \mathsf{P}_{it} = 0$.);

(v) number of wells N in a reservoir region is allowed to be a large number (For example, N can be as large as 6,000 in Eagle Ford Shale Play.);

(vi) the dimension of the completion data p can be as large as 100;

(vii) identification of significant completion predictors among p completion data $\mathbf{x} = (x_1, \cdots, x_p)^\top \in \mathbb{R}^p$ in terms of a well productivity is one of the paramount tasks in reservoir characterization;

(viii) set of well locations $\{\mathbf{s}_i\}_{i=1}^N$ may form a band where oil or gas reserves exist underground as shown on Fig. 3.

We are now in a position to suggest a "*shale reservoir problem*" based on the aforementioned characteristics of the shale oil wells data $\{(\mathbf{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$. Its objective is to achieve the following three sub-goals:

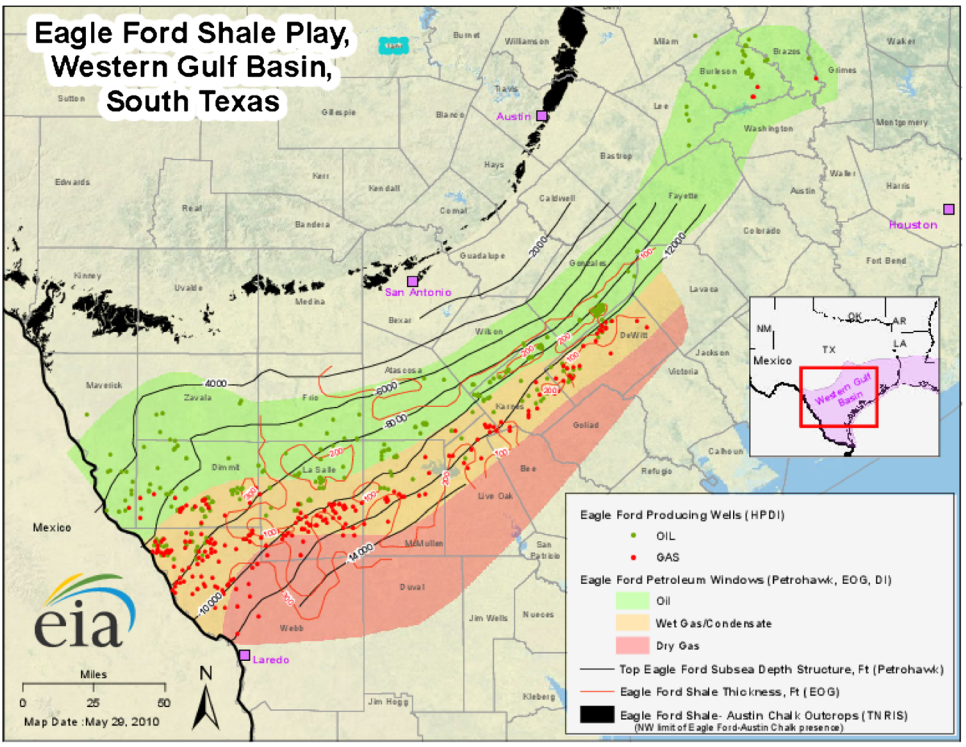

Figure 3: Eagle Ford region with three types of petroleum windows. (Source: US EIA)

(a) uncover a hidden pattern from the time series data $\mathbf{P}_i = (\mathsf{P}_{i1}, \cdots, \mathsf{P}_{iT_i})^\top$ 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 the relationship between the measurements $\mathbf{P}_i$ and some characteristics of the i-th well in terms of productivity;

(b) 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 $\boldsymbol{\vartheta}_i$;

(c) predict the temporal profile of oil production rates of a new well during T^* months, denoted by $\mathbf{P}^* = (\mathsf{P}^*_1, \cdots, \mathsf{P}^*_{T^*})^\top$, 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.

Throughout the paper we use the straight capital letters 'P' and '**P**' to represent 'production', and the asterisk notation '$*$' is superscripted to emphasize that a resultant quantity is associated with 'prediction'.

The suggested shale reservoir problem (a)–(c) is a supervised machine learning problem that combines (a) curve fitting, (b) feature selection, and (c) spatial prediction. An idealistic statistical model should be able to fulfill the sub-goals (a)–(c) integratively.

In the literature of petroleum engineering, the curve fitting analysis dedicated to the purpose with sub-goal (a) is referred to as decline curve analysis (DCA) (Arps and et al. 1945; Ilk et al., 2008; Valkó et al., 2010; Clark, 2011; Duong and et al. 2011) which will be detailed in Section 2. However, to the best of our knowledge, there is no research work aimed at providing an integrative fully Bayesian solution to achieve the three sub-goals (a)–(c) through a single model. Recently, Vyas (2017) & Park et al. (2021) researched statistical modeling to accommodate the three components of shale oil wells data, however, the authors did not model the spatial association through a geostatistical stochastic process, hence, uncertainty quantification associated with the sub-goal (c) is disregarded.

1.3. Research region: Eagle Ford Shale Reservoir of South Texas In this research, we consider 360 hydraulically fractured horizontal oil wells data $\{(\mathbf{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$ ($N = 360$) collected from the Eagle Ford shale reservoir of South Texas. No wells were refractured after their completions. Data has been collected from DrillingInfo online database (info.drillinginfo.com). The Eagle Ford shale reservoir (see Fig. 3) is known as possibly the largest single economic development in the history of Texas and ranked as the largest oil and gas development in the world based on capital invested; visit eagleford-shale.com for a detail.

The followings are detailed descriptions of the components, $\{\mathbf{P}_i\}_{i=1}^N$, $\{\mathbf{x}_i\}_{i=1}^N$, and $\{\mathbf{s}_i\}_{i=1}^N$, from the shale oil wells dataset $\{(\mathbf{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$, $N = 360$.

The time frame of the production time series data of N wells, $\{\mathbf{P}_i = (\mathsf{P}_{i1}, \cdots, \mathsf{P}_{iT_i})^\top\}_{i=1}^{N=360}$, spans from January 2011 through June 2017. Each well i can have different production period T_i. The minimum and maximum production periods of the 360 wells are 42 and 78 months, respectively. (That is, $\min_{i=1,\cdots,360} T_i = 42$ months, and $\max_{i=1,\cdots,360} T_i = 78$ months.) Unit of the production rate P_{it} ($t = 1, \cdots, T_i$) is barrel per month (BBL/Month).

Figure 4 displays the production rate trajectory for a well with API10 number given by 4201334341, and those of other wells are overlaid with gray dashed curves. API10 is a unique identifier assigned to an individual well, hence, there is an one-to-one correspondence between the index i and the

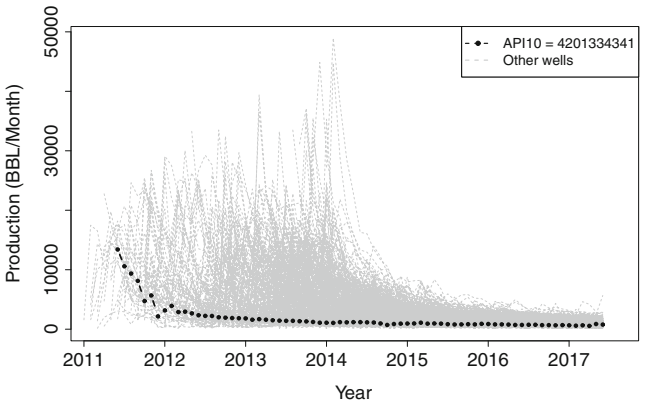

Figure 4: Production rate trajectory for well with API10 = 4201334341. Other wells are transparently overlaid

API10 number. The initial production rate of the well is 13451 BBL/Month in June 2011. This implies that the total amount of oil produced during June is 13451 barrels. As noted from item (iv) in Section 1.2, temporal profiles of production rate eventually show a declining phase.

It is important to note that the declining phase commonly manifested in the oil production rate time series data is due to the exhaustion of expulsive force rather than depletion of oil, because much more oil is left underground than is ordinarily brought to the surface (Lewis et al., 1918). This implies that the advance of the drilling technology, hydraulically fractured horizontal drilling, has been already reflected on the historical records of the monthly rates $\mathbf{P}_i$, and this is the main reason why production 'rate' data (with a unit BBL/Month) are prevailingly used in the DCA.

Well completion data from N wells $\{\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^\top\}_{i=1}^{N=360}$ has $p = 11$ completion predictors. Table 1 lists the 11 completion data, and these fall broadly into two classes: hydraulic fracturing design data or well test data. Pictorial explanation of some hydraulic fracturing design data is contained in Fig. 2. Petroleum engineers tune these predictors and optimize a well design and manufacturing phase to attain the maximum well productivity (Gradl, 2018; Wan, 2011; Deen et al., 2015). (Covariates presented in this article are confined with 11, but actual dimension of completion data can be as large as 100.)

Spatial location data $\{\mathbf{s}_i = (s_{i1}, s_{i2})^\top\}_{i=1}^{N}$ represent (longitude, latitude) coordinates of the N wells, recorded in decimal degrees used in WGS84 coordinate reference system. They are located on the green area in Fig. 3.

Table 1: 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

NOTE: Fig. 2 shows the some hydraulic fracturing design data among well completion data. Unit of abstract is blocks of land units in Texas. It applies to Texas only. (Source: Drillinginfo)

The rest of this paper is organized as follows. Section 2 explains the traditional DCA and how to implement it given temporal profiles $\{\mathbf{P}_i\}_{i=1}^N$. Section 3 introduces a non-spatial extension of the traditional DCA to utilize time series data and completion data $\{(\mathbf{P}_i, \mathbf{x}_i)\}_{i=1}^N$, and four candidate models are suggested. Model choice is employed in Section 4 to select the best model from the four non-spatial models. In Section 5, we further extend the framework of the non-spatial models to a spatial platform, thereby exploiting the entire shale oil wells dataset $\{(\mathbf{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$. This spatial model is applied to the 360 shale oil wells data, and some results are summarized in Section 6, while Section 7 contains some summary and future works.

2 Traditional Decline Curve Analysis

Since the traditional DCA has been introduced by Lewis et al. (1918) in 1918, it has been one of the most popularly utilized tools for petroleum engineers (Mishra and Datta-Gupta, 2017). Provided a collection of the production data $\{(\mathbf{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$ from N well in a shale oil or gas reservoir as explained in Section 1.2, the traditional DCA only utilizes the time series data $\{\mathbf{P}_i\}_{i=1}^N$, while completion and location information, $\{(\mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$, are not involved in the analysis. The objective of the DCA was delineated in Section 1.2 as the sub-goal (a) subsumed under the shale reservoir problem.

Consider an individual well i $(i = 1, \cdots, N)$. A key ingredient of the DCA is a parametric curve $q(t; \boldsymbol{\vartheta}_i)$ to describe the "inherent tendency" of the declining phase from the temporal profile $\mathbf{P}_i = (\mathsf{P}_{i1}, \cdots, \mathsf{P}_{iT_i})^\top$ to evolve over time (as seen from the Fig. 4). Such a curve $q(t; \boldsymbol{\vartheta}_i)$ is referred to as rate decline curve (RDC). Conceptually, any parametric curve that can describe such a declining pattern is a candidate of the RDC. Subject-specific parameter vector $\boldsymbol{\vartheta}_i$ represents an "inherent characteristic" of the i-th well dictating this tendency. Here, a notable point is that petroleum engineers usually elicit and involve certain interpretations on each component of the $\boldsymbol{\vartheta}_i$ in terms of a well productivity.

Most RDCs in the literature are derived from solving certain differential equations that describe a hidden dynamic from production rate trajectory (Arps and et al. 1945; Ilk et al., 2008; Valkó et al., 2010; Clark, 2011; Duong and et al. 2011), whereas some RDCs are developed by generalizing certain probability density functions (pdf) (Weibull and et al. 1951). See (Fetkovich and et al. 1980; Ali et al., 2015; Harris et al., 2014; Miao et al., 2018) for an overview for some RDCs. In this paper, we introduce four widely used RDCs: Weibull, Arps' hyperbolic, stretched exponentiated decline, and Duong models.

2.1. Weibull Model Weibull model is given by

$$\begin{aligned} q(t;M,\lambda,k) &= M \cdot f(t;\lambda,k) \quad M>0,\lambda>0,k>0. \\ \text{such that } f(t;\lambda,k) &= \text{Weibull}(t;b,k) = (k/\lambda)(t/\lambda)^{k-1}\exp(-(t/\lambda)^k) \end{aligned} \tag{2.1}$$

where M, λ, and k are called the carrying capacity parameter, scale parameter, and shape parameter, respectively. Note that Weibull model is simply a generalization of Weibull distribution (Weibull and et al. 1951). In petroleum context, there are *separate* interpretations for each of the curve parameters in terms of a well productivity: M means the maximum cumulative production set by the equation; λ is the value of time at which 63.2% of the resource have been produced; and k shows how the rate of growth changes with time and is usually less than 1 for unconventional reservoirs (Mishra, 2012a).

In what follows, for the purpose of estimation, we use a reparameterized Weibull model obtained via a transformation $b = \lambda^{-k}$ in Eq. 2.1:

$$q(t;M,b,k) = M \cdot f(t;b,k), \quad M>0,b>0,k>0 \tag{2.2}$$

$$\text{such that } f(t;b,k) = \text{Weibull}(t;b,k) = bkt^{k-1}\exp(-bt^k), \tag{2.3}$$

where M, b, and k are called the carrying capacity parameter, scale parameter, and shape parameter, respectively.

2.2. Arps' Hyperbolic Model Arps' hyperbolic model (Arps and et al. 1945) is defined by

$$q(t;q_1,h,D) = q_1 \cdot (1+hDt)^{-1/h}, \quad q_1>0, 0<h<1, D>0, \tag{2.4}$$

where q_1, h, and D are called the initial production rate parameter, hyperbolic decline exponent parameter, and initial decline rate parameter, respectively. For technical interpretations for the curve parameters in terms of well productivity, refer to Arps and et al. (1945).

Under reparameterization $M = q_1/((1-h)D)$, $b = 1/((1-h)D)$, and $k = h/(1-h)$, we reexpress Arps' hyperbolic model (2.4) as a generalization of the generalized Pareto distribution (GPD) (Pickands and et al. 1975) as follows:

$$q(t;M,b,k) = M \cdot f(t;b,k), \quad M>0,b>0,k>0 \tag{2.5}$$

$$\text{such that } f(t;b,k) = \text{GPD}(t;b,k) = (1/b)(1+kt/b)^{-1/k-1}, \tag{2.6}$$

where M, b and k are the carrying capacity parameter, scale parameter and shape parameter, respectively.

2.3. Stretched Exponentiated Decline Model The stretched exponentiated decline model (SEDM) (Valkó et al., 2010) is defined by

$$q(t; q_1, \lambda, k) = q_1 \exp(-(t/\lambda)^k), \quad q_1 > 0, \lambda > 0, k > 0, \tag{2.7}$$

where q_1, λ and k are called the initial production rate parameter, the characteristic time parameter, and exponent parameter, respectively. For a technical interpretation for the curve parameters, refer to Valkó et al. (2010).

Under reparameterization $M = q_1$ and $b = \lambda^{-k}$, SEDM is represented as follow:

$$q(t; M, b, k) = M \cdot f(t; b, k), \quad M > 0, b > 0, k > 0 \tag{2.8}$$

$$\text{such that } f(t; b, k) = \exp(-bt^k), \tag{2.9}$$

where M, b and k are the carrying capacity parameter, scale parameter, and shape parameter, respectively.

2.4. Duong Model Duong model (Duong and et al. 2011) is defined by

$$q(t; q_1, a, m) = q_1 t^{-m} \exp(a(t^{1-m} - 1)/(1-m)), \quad q_1 > 0, a > 0, m > 0, \tag{2.10}$$

where q_1, a, and m are called the initial production rate parameter, intercept constant parameter, and slope parameter, respectively. Duong and et al. (2011) empirically showed that m is always greater than unity for shale reservoirs, and m is less than unity for conventional tight well. As shale oil is our concern, we restrict parameter space of m to be $m > 1$. For a technical interpretation for the curve parameters, refer to Duong and et al. (2011).

Under reparameterization $M = q_1 \exp(a/(m-1))/a$, $b = a/(m-1)$, and $k = m - 1$, Duong model (2.10) is represented as a generalization of Fréchet distribution (Fréchet, 1928) as follows:

$$q(t; M, b, k) = M \cdot f(t; b, k), \quad M > 0, b > 0, k > 0, \tag{2.11}$$

$$\text{such that } f(t; b, k) = \text{Fréchet}(t; b, k) = bkt^{-1-k} \exp(-bt^{-k}), \tag{2.12}$$

where M, b and k are the carrying capacity parameter, scale parameter, and shape parameter, respectively.

The Arps' and Duong models are originally derived by solving ordinary differential equations (Arps and et al. 1945; Duong and et al. 2011), but after certain reparameterization they can be viewed as generalizations of pdfs typically used in extreme value theory (DeHaan and Ferreira, 2007; Lee and Kim, 2019). After the reparameterization, each of the curve parameters of the RDCs, Eqs. 2.2, 2.5, 2.9, and 2.11, share coincided notations, denoted by carrying capacity parameter ($M > 0$), scale parameter ($b > 0$), and shape parameter ($k > 0$), respectively.

2.5. Implementation of Traditional Decline Curve Analysis We illustrate how to implement the traditional DCA based on the Weibull model (2.2), but the scheme is applicable to any choice of RDC. Consider production rate trajectories $\{\mathbf{P}_i\}_{i=1}^N$ from N wells. DCA may assume an additive error model where errors are distributed according to independent and identical Gaussian noises across all the wells ($i = 1, \cdots, N$) and time points ($t = 1, \cdots, T$):

$$y_{it} = \mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2), \quad i = 1, \cdots, N, \; t = 1, \cdots, T_i, \tag{2.13}$$

where $y_{it} = \log \mathsf{P}_{it}$ is log-scaled oil production rate of the i-th well at the t-th time point. The mean part in Eq. 2.13 is the log transformation of the Weibull model (2.2), given by $\mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) = \log q(t; M_i, b_i, k_i) = \theta_{1i} + \theta_{2i} + \theta_{3i} + (e^{\theta_{3i}} - 1) \cdot \log t - e^{\theta_{2i}} \cdot t^{e^{\theta_{3i}}}$, dictated by three real-valued curve parameters $\theta_{1i} = \log M_i$, $\theta_{2i} = \log b_i$, and $\theta_{3i} = \log k_i$.

Let T_i-dimensional vector $\mathbf{y}_i = (y_{i1}, \cdots, y_{iT_i})^\top \in \mathbb{R}^{T_i}$ represents the log-scaled oil production rate trajectory for the i-th well during T_i months from its initial production, and its corresponding mean vector is denoted as $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i}) = (\mu(1; \theta_{1i}, \theta_{2i}, \theta_{3i}), \cdots, \mu(T_i; \theta_{1i}, \theta_{2i}, \theta_{3i}))^\top = (\mu_{i1}, \cdots, \mu_{iT_i})^\top \in \mathbb{R}^{T_i}$. Note that the $\boldsymbol{\mu}_i$ can be viewed as a T_i-dimensional vector-valued function with respect to the curve parameters $(\theta_{1i}, \theta_{2i}, \theta_{3i}) \in \mathbb{R}^3$.

A likelihood function for the i-th well can be re-written in a vector-form:

$$\mathcal{L}(\theta_{1i}, \theta_{2i}, \theta_{3i}, \sigma^2 | \mathbf{y}_i) = \mathcal{N}_{T_i}(\mathbf{y}_i | \boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i}), \sigma^2 \mathbf{I}), \quad i = 1, \cdots, N,$$

where $\mathcal{N}_T(\mathbf{a}, \boldsymbol{\Sigma})$ represents the T-dimensional Gaussian distribution with mean vector $\mathbf{a}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\mathbf{I}$ denotes an identity matrix. Now, define three N-dimensional vectors formulated from each of the curve parameters, $\boldsymbol{\theta}_1 = (\theta_{11}, \cdots, \theta_{1N})^\top$, $\boldsymbol{\theta}_2 = (\theta_{21}, \cdots, \theta_{2N})^\top$, and $\boldsymbol{\theta}_3 = (\theta_{31}, \cdots, \theta_{3N})^\top$ across the N wells. We emphasize again that each vector $\boldsymbol{\theta}_l$ ($l = 1, 2, 3$) has its own meaning in terms of a well productivity that petroleum engineer want to infer about (Mishra and et al. 2012b).

A likelihood function based on the N observations $\mathbf{y}_{1:N} = \{\mathbf{y}_i\}_{i=1}^N$ is

$$\mathcal{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \sigma^2 | \mathbf{y}_{1:N}) = \prod_{i=1}^N \mathcal{N}_{T_i}(\mathbf{y}_i | \boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i}), \sigma^2 \mathbf{I}). \tag{2.14}$$

To implement the traditional DCA, one need to maximize the likelihood (2.14) with respect to the curve parameters $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ given σ fixed, to obtain:

$$\begin{aligned}(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2, \widehat{\boldsymbol{\theta}}_3) &= \underset{(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3)\in\mathbb{R}^N\times\mathbb{R}^N\times\mathbb{R}^N}{\arg\max} \log\ \mathcal{L}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \sigma^2|\mathbf{y}_{1:N}) \\ &= \underset{(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3)\in\mathbb{R}^N\times\mathbb{R}^N\times\mathbb{R}^N}{\arg\min} \sum_{i=1}^{N} \|\mathbf{y}_i - \boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i})\|_2^2 \\ &= \underset{(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3)\in\mathbb{R}^N\times\mathbb{R}^N\times\mathbb{R}^N}{\arg\min} \sum_{i=1}^{N}\sum_{t=1}^{T_i} \Big(y_{it} - \mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i})\Big)^2,\end{aligned}$$

where $\|\cdot\|_2$ is the Euclidean norm.

Following acquisition of point estimates $(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2, \widehat{\boldsymbol{\theta}}_3)$ (which includes $(\widehat{\theta}_{1i}, \widehat{\theta}_{2i}, \widehat{\theta}_{3i})$) through an optimization technique via Newton-Raphson method or Gradient descent method (Boyd et al., 2004), one may plug them into the mean part of the model (2.13) to derive a fitted curve $\mu(t; \widehat{\theta}_{1i}, \widehat{\theta}_{2i}, \widehat{\theta}_{3i})$ $(t = 1, \cdots, T_i)$ for each well i. This function is then viewed as a denoised temporal tendency for the log-scaled oil production rates of the i-th well during T_i months, which can be further extrapolated to see a future production pattern. In conclusion, the traditional DCA is a deterministic curve fitting method based on least squares estimation (Larkey, 1925; Arps and et al. 1945; McNulty et al., 1981).

Recently, it seems that there increasingly arises a need to use probabilistic approaches in carrying out the DCA to evaluate an uncertainty associated with reserves estimates (Cheng et al., 2010). The uncertainty quantification is even more important in exploring production results from unconventional reservoirs. This may be because the extreme low-permeability of unconventional reservoirs, generally in the nano-darcy range, ensures the continuing presence of the pressure-transient effects (Kazemi and et al. 1969) during well production. Reservoir modeling in such systems is an extremely complicated task (Schuetter et al., 2015), and pointwise forecasting for the oil production decline curve can produce unreliable prediction outcomes. As remedies, two approaches are mainly used in the petroleum engineering literature: (Jochen et al., 1996; Cheng et al., 2010) carried out bootstrapping (Tibshirani and Efron, 1993), and Gong et al. (2011, 2014); Zhang et al. (2015) employed Bayesian approaches. One of the limitations of the authors' works is that they only utilize the production trajectories $\{\mathbf{P}_i\}_{i=1}^N$, hence, the full aspects of the shale oil wells data are not harnessed.

3 Non-Spatial Bayesian Hierarchical Model

3.1. Individual-level model Consider shale oil wells data $\{(\mathbf{P}_i, \mathbf{x}_i)\}_{i=1}^N$ without spatial locations. We propose a non-spatial Bayesian hierarchical model that extends the traditional DCA to achieve the sub-goals (a) and (b) of the shale reservoir problem. The data level of the hierarchical model inherits a likelihood function from the traditional DCA (2.13): that is, by letting $y_{it} = \log\ \mathsf{P}_{it}$ with $i = 1, \cdots, N$, $t = 1, \cdots, T_i$, the response profile follows

$$y_{it} = \mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2). \tag{3.1}$$

The mean part in Eq. 3.1 is described by a log-scaled RDC $\mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) = \log\ q(t; M_i, b_i, k_i)$, $\theta_{1i} = \log\ M_i$, $\theta_{2i} = \log\ b_i$, and $\theta_{3i} = \log\ k_i$, which has four options according to what RDC has been chosen from Eqs. 2.2, 2.5, 2.9), and Eq. 2.11:

- Weibull model ($\mathcal{M}_1$):

$$\mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) = \theta_{1i} + \theta_{2i} + \theta_{3i} + (e^{\theta_{3i}} - 1) \cdot \log t - e^{\theta_{2i}} \cdot t^{e^{\theta_{3i}}}. \tag{3.2}$$

- Arps model ($\mathcal{M}_2$):

$$\mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) = \theta_{1i} - \theta_{2i} - (1 + e^{-\theta_{3i}}) \cdot \log(1 + e^{\theta_{3i} - \theta_{2i}} \cdot t). \tag{3.3}$$

- SEDM model ($\mathcal{M}_3$):

$$\mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) = \theta_{1i} - e^{\theta_{2i}} \cdot t^{e^{\theta_{3i}}}. \tag{3.4}$$

- Duong model ($\mathcal{M}_4$):

$$\mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) = \theta_{1i} + \theta_{2i} + \theta_{3i} - (1 + e^{\theta_{3i}}) \cdot \log t - e^{\theta_{2i}} \cdot t^{-e^{\theta_{3i}}}. \tag{3.5}$$

For the purpose of articulation, the four models are assigned model notations $\mathcal{M}_m$ $(m = 1, 2, 3, 4)$ as denoted in Eqs. 3.2–3.5.

3.2. Population Model The curve parameters θ_{li} $(l = 1, 2, 3,\ i = 1, \cdots, N)$ in (3.1) assume the following two roles: first, for each well i, the 3-tuple $(\theta_{1i}, \theta_{2i}, \theta_{3i}) \in \mathbb{R}^3$ dictates the shape of the log-scaled RDC $\mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i})$, $(t = 1, \cdots, T_i)$, to fit the i-th well's temporal response $\mathbf{y}_i = (y_{i1}, \cdots, y_{iT_i})^\top \in \mathbb{R}^{T_i}$ over time; and second, for each l, the collection of the latent variables $\{\theta_{li}\}_{i=1}^N$ embeds some interpretation regarding a productivity across N wells,

thereby motivating a population regression analysis. Roughly speaking, the first and second roles are relevant with the sub-goals (a) and (b), respectively.

For each l $(l = 1, 2, 3)$, we construct a linear regression to explain the between-individual differences in $\{\theta_{li}\}_{i=1}^{N}$ through the p completion predictors $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^\top \in \mathbb{R}^p$ as follows:

$$\theta_{li} = \alpha_l + \mathbf{x}_i^\top \boldsymbol{\beta}_l + \epsilon_{li}, \quad \epsilon_{li} \sim \mathcal{N}(0, \sigma_l^2), \quad l = 1, 2, 3, i = 1, \cdots, N, \tag{3.6}$$

where α_l is the intercept term and $\boldsymbol{\beta}_l = (\beta_{l1}, \cdots, \beta_{lp})^\top$ is the p coefficients corresponding to the p completion predictor $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^\top$. In multilevel analysis, the random error terms in Eqs. 3.1 and 3.6, that is, ε_{it} and ϵ_{li}, are referred to as (remaining) intra-individual and inter-individual variabilities, respectively.

Aligned with the sub-goal (b), we wish to identify significant predictors in this regression setup (3.6), thereby necessitating a sparsity on the coefficients $\boldsymbol{\beta}_l$ $(l = 1, 2, 3)$. To impose a hard sparsity (Wainwright, 2019) on the coefficients which only allows zero posterior means for the noise coefficients, one may use the spike-and-slab type priors (Mitchell and Beauchamp, 1988; Ishwaran et al., 2005) for the coefficients. Essentially, a spike-and-slab type prior is a mixture distribution of a Dirac measure at zero and a continuous distribution (O'Hara et al., 2009). One of the drawbacks of such priors is its computational burden (Van Der Pas et al., 2014) as posterior computation may require exploration of a model space of size 2^p from each of the linear regressions (3.6) (George and McCulloch, 1993).

Instead, we place a soft-sparsity (Wainwright, 2019) on the coefficients $\boldsymbol{\beta}_l$ $(l = 1, 2, 3)$, where the posterior mean corresponding to noises coefficients are allowed to be zeros or very close to zeros. Continuous shrinkage priors can do this (Polson and Scott, 2010): examples include the Bayesian lasso (Park and Casella, 2008), horseshoe prior (Carvalho et al. 2009, 2010), and generalized double Pareto shrinkage prior (Armagan et al., 2013), GLT prior (Lee et al., 2020c), etc, among many others (Bhattacharya et al., 2015; Bae and Mallick, 2004).

In this paper, we use the horseshoe prior (Carvalho et al. 2009, 2010), which is known to be computationally efficient than the spike-and-slab type priors, and useful for addressing the variable selection problem (Van Der Pas et al., 2014). For each l $(l = 1, 2, 3)$, a hierarchical structure underneath the coefficients $\boldsymbol{\beta}_l$ in Eq. 3.6 is then given by

$$\beta_{lj} | \lambda_{lj}, \tau_l, \sigma_l \sim \mathcal{N}(0, \lambda_{lj}^2 \tau_l^2 \sigma_l^2),\ \lambda_{lj} \sim \mathcal{C}^+(0, 1),\ \tau_l \sim \mathcal{C}^+(0, 1),\ l = 1, 2, 3,\ j = 1, \cdots, p, \tag{3.7}$$

where $\mathcal{C}^+(0,1)$ denotes the unit-scaled half-Cauchy distribution. (Density is given by $\pi(x) \propto 1/(1+x^2)$ for $x > 0$.) For each l, the λ_{lj}'s and τ_l's are referred as the local and global scale parameters, respectively. Essentially, the local scale parameters are associated with signal selection from the coefficients, whereas the global one imposes a strong gravity over the entire coefficients (Polson and Scott, 2010).

As for the error terms, we use the unit-scaled half-Cauchy prior for the standard deviation of the likelihood (3.1) and Jeffreys priors for the variances of the latent regressions (3.6). Intercept terms in the regressions (3.6) are given by the flat priors:

$$\sigma \sim \mathcal{C}^+(0,1),\ \alpha_l \sim \pi(\alpha_l) \propto 1,\ \sigma_l^2 \sim \pi(\sigma_l^2) \propto 1/\sigma_l^2, \quad l = 1,2,3. \tag{3.8}$$

Figure 5 displays results of curve fitting for the well with API10 = 4201334341 in the Fig. 4 obtained by the four non-spatial Bayesian hierarchical models. Virtually, the difference of four fitted curves is shown around the first ten months, but then it is indistinguishable as time evolves.

4 Model Choice

The aim of the present section is to choose the best model among four non-spatial Bayesian hierarchical models where only RDCs are different, specified by Weibull model $\mathcal{M}_1$ (3.2), Arps' hyperbolic model $\mathcal{M}_2$ (3.3), SEDM $\mathcal{M}_3$ (3.4), and Duong model $\mathcal{M}_4$ (3.5). To that end, we use three model comparison criteria: deviance information criterion (DIC) (Spiegelhalter et al., 2002; Gelman et al., 2004), widely applicable information criterion (WAIC) (Watanabe, 2010), and posterior predictive loss criterion (PPLC)

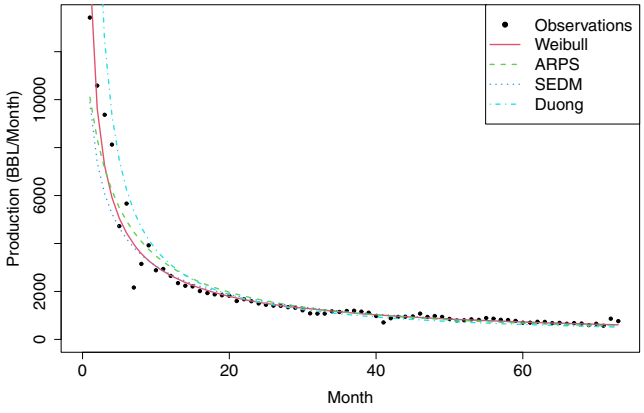

Figure 5: Bayesian decline curve analysis based on the four models

(Gelfand and Ghosh, 1998). In the next subsections, we shall provide some brief summaries of the criteria, and then apply them to the proposed non-spatial models. For a detailed explanation of the criteria, refer to Banerjee et al. (2014) & Gelman et al. (2014).

4.1. Deviance Information Criterion Assume that $\mathbf{y}_{1:N} = \{\mathbf{y}_i\}_{i=1}^N$ is a collection of observations from N subjects such that a likelihood function is denoted by $\mathcal{L}(\boldsymbol{\vartheta}|\mathbf{y}_{1:N})$. Dimension of the $\mathbf{y}_i$ is allowed to be multidimensional. In general, when dealing with a Bayesian hierarchical model, a possible form of a likelihood function $\mathcal{L}(\boldsymbol{\vartheta}|\mathbf{y}_{1:N})$ may depend on which latent variables within the hierarchy are regarded as the parameters of primary interest, represented by $\boldsymbol{\vartheta}$: this specification issue is referred to as *focus issue* (Spiegelhalter et al., 2002). Having determined the parameters of primary interest, deviance is defined by $D(\boldsymbol{\vartheta}) = -2\log\ \mathcal{L}(\boldsymbol{\vartheta}|\mathbf{y}_{1:N})$. The deviance is a goodness-of-fit statistics, and a lower value indicates a better fitting (Celeux et al., 2006). Typically, because a complex model gets rewards in terms of the deviance, we often need to additionally impose a penalty for the model complexity, also called an effective number of parameters, in manufacturing a model comparison criterion. Roughly speaking, a common principle underlying the three criteria, DIC, WAIC and PPLC, is to induce some trade-off effect between the goodness-of-fit and penalty terms.

Gelman et al. (2004) suggested DIC, a generalized version of Akaike information criterion (Akaike, 1998) for a Bayesian hierarchical model, given by

$$\text{DIC} = D(\bar{\boldsymbol{\vartheta}}) + 2 \cdot p_{\text{D}}, \tag{4.1}$$

where the goodness-of-fit term is the value of deviance evaluated at the posterior mean of $\boldsymbol{\vartheta}$, denoted by $\bar{\boldsymbol{\vartheta}} = \mathbb{E}\boldsymbol{\vartheta}|\mathbf{y}_{1:N}]$, and the effective number of parameters in the penalty term is obtained by $p_{\text{D}} = \text{Var}[D(\boldsymbol{\vartheta})|\mathbf{y}_{1:N}]/2 = 2 \cdot \text{Var}[\log\ \mathcal{L}(\boldsymbol{\vartheta}|\mathbf{y}_{1:N})|\mathbf{y}_{1:N}]$. A model with a smaller value for DIC indicates a better predictive performance among considered models.

Going back to our examples, we measure the DIC corresponding to each of the four models, $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$ (3.2)–(3.5), provided that the parameters of primary interest are specified to be $\boldsymbol{\vartheta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \sigma^2)$, that is, the collection of curve parameters and error variance for the N wells. Under this formulation, deviance based on Eq, 3.1 is

$$D(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \sigma^2) = -2\sum_{i=1}^{N} \log\ \mathcal{N}_{T_i}(\mathbf{y}_i|\boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i}), \sigma^2\mathbf{I}),$$

where only mean part $\boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i})$ differs according to what RDC is employed from $\mathcal{M}_m$, $m = 1, 2, 3, 4$, Eqs. 3.2–3.5. To obtain the DIC (4.1), we resort to Markov chain Monte Carlo (MCMC) techniques (Robert and Casella, 2013) to approximate the DIC because the posterior mean $\bar{\boldsymbol{\vartheta}} = \mathbb{E}[\boldsymbol{\vartheta}|\mathbf{y}_{1:N}]$ and the effective number of parameters p_{D} are not expressed in closed-forms.

4.2. Widely Applicable Information Criterion (Watanabe, 2010) introduced WAIC which is regarded as a fully Bayesian version of the DIC (4.1) in the sense that a goodness-of-fit term exploits the entire posterior distribution. Note that the goodness-of-fit term of the DIC (4.1) is acquired by plugging-in the posterior mean $\bar{\boldsymbol{\vartheta}}$ to the deviance $D(\boldsymbol{\vartheta})$, which lacks a fully Bayesian sense. It is known that WAIC is asymptotically equals to Bayesian cross-validation (Vehtari and Gelman, 2014), and also applicable to singular models.

WAIC is defined by

$$\mathrm{WAIC} = -2 \cdot \mathrm{LPPD} + 2 \cdot p_{\mathrm{W}}, \tag{4.2}$$

where the goodness-of-fit term is called the log posterior predictive density (LPPD), $\mathrm{LPPD} = \sum_{i=1}^{N} \log\, \mathbb{E}[\mathcal{L}(\boldsymbol{\vartheta}|\mathbf{y}_i)|\mathbf{y}_{1:N}]$, and the effective number of parameter in the penalty term is defined by $p_{\mathrm{W}} = \sum_{i=1}^{N} \mathrm{Var}[\log\, \mathcal{L}(\boldsymbol{\vartheta}|\mathbf{y}_i)|\mathbf{y}_{1:N}]$. Gelman et al. (2014) pointed out that (i) although without the factor 2 in the second term, p_{W} looks similar to p_{D} in Eq. 4.1 and (ii) p_{W} is computationally stable than p_{D} because p_{W} computes the variance separately for each data point and then summation is carried out.

Given posterior samples $\{(\boldsymbol{\vartheta})^{(s)}\}_{s=1}^{S} \sim \pi(\boldsymbol{\vartheta}|\mathbf{y}_{1:N})$, the LPPD and p_W terms are approximated by

$$\widehat{\mathrm{LPPD}} = \sum_{i=1}^{N} \log\, \left(\frac{1}{S} \sum_{s=1}^{S} \mathcal{L}((\boldsymbol{\vartheta})^{(s)}|\mathbf{y}_i) \right), \tag{4.3}$$

$$\widehat{p_{\mathrm{W}}} = \sum_{i=1}^{N} \left\{ \frac{1}{S-1} \sum_{s=1}^{S} \left(\log\, \mathcal{L}((\boldsymbol{\vartheta})^{(s)}|\mathbf{y}_i) - \frac{1}{S} \sum_{s=1}^{S} \log\, \mathcal{L}((\boldsymbol{\vartheta})^{(s)}|\mathbf{y}_i) \right)^2 \right\}. \tag{4.4}$$

Returning to our examples, given that the parameter of primary interest is set by $\boldsymbol{\vartheta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \sigma^2)$ (as we did to get the DIC), we can approximate the value of WAIC corresponding to each of the four models as follows. First, replace $\mathcal{L}(\boldsymbol{\vartheta}|\mathbf{y}_i)$ in Eqs. 4.3 and 4.4 with the likelihood $\mathcal{L}(\theta_{1i}, \theta_{2i}, \theta_{3i}, \sigma^2|\mathbf{y}_i) = \mathcal{N}_{T_i}(\mathbf{y}_i|\boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i}), \sigma^2\mathbf{I})$ (3.1), $i = 1, \cdots, N$, where only $\boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i})$ differs based on the four options $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$ (3.2)–(3.5), and second, approximate $\widehat{\mathrm{LPPD}}$ and $\widehat{p_{\mathrm{W}}}$ by using a MCMC method, and finally, obtain an approximation of WAIC (4.2) corresponding to each model.

4.3. Posterior Predictive Loss Criterion (Gelfand and Ghosh, 1998) introduced PPLC as an alternative to DIC (4.1) or WAIC (4.2), which is particularly useful to compare nonnested hierarchical models. A notable feature of PPLC different from DIC and WAIC is its use of replicated observations, denoted by $\mathbf{y}_{i,rep}$ $(i = 1, \cdots, N)$, corresponding to the actual observations $\mathbf{y}_{i,obs}$. A replicate $\mathbf{y}_{i,rep}$ is drawn from the posterior predictive density

$$f(\mathbf{y}_{i,rep}|\mathbf{y}_{1:N}) = \int \mathcal{L}(\boldsymbol{\vartheta}|\mathbf{y}_{i,rep})\pi(\boldsymbol{\vartheta}|\mathbf{y}_{1:N})d\boldsymbol{\vartheta}, \quad i = 1, \cdots, N, \tag{4.5}$$

where $\mathcal{L}(\cdot)$ is a likelihood function for the i-th component and $\pi(\boldsymbol{\vartheta}|\mathbf{y}_{1:N})$ is posterior distribution. The idea of using such replicates $\{\mathbf{y}_{i,rep}\}_{i=1}^N$ for a criticism of the model in light of the observed data $\{\mathbf{y}_{i,obs}\}_{i=1}^N$ is also purported by Box (1980).

A general rule of the PPLC is principled on a balanced loss function (Zellner, 1994). Given any loss function $l(\cdot)$ and a positive real number k, a balanced loss function is defined by

$$l(\mathbf{y}_{i,rep}, \mathbf{a}_i; \mathbf{y}_{1:N}) = l(\mathbf{y}_{i,rep}, \mathbf{a}_i) + k \cdot l(\mathbf{y}_{i,obs}, \mathbf{a}_i), \quad k > 0,\ i = 1, \cdots, N, \tag{4.6}$$

where $\mathbf{a}_i$ is a non-stochastic action vector, k is a weight, and $\mathbf{y}_{i,rep}$ is a replicate for $\mathbf{y}_{i,obs}$. Conceptually, the role of action vector $\mathbf{a}_i$ is to accommodate both $\mathbf{y}_{i,obs}$, and what we predict for $\mathbf{y}_{i,rep}$. Note that the loss function on the left-hand side of Eq. 4.6 penalizes actions $\mathbf{a}_i$ both for departure from the corresponding observed value (fit) as well as for departure from what we expect the replicate to be (smoothness) (Banerjee et al., 2014). A generic version of PPLC is defined by $D_k = \sum_{i=1}^N \min_{\mathbf{a}_i} \mathbb{E}[l(\mathbf{y}_{i,rep}, \mathbf{a}_i; \mathbf{y}_{1:N})|\mathbf{y}_{1:N}]$, where the expectation $\mathbb{E}[\cdot|\mathbf{y}_{1:N}]$ is taken with respect to the predictive density $f(\mathbf{y}_{i,rep}|\mathbf{y}_{1:N})$ (4.5) for some specified $k \geq 0$. Note that the resultant D_k does not depend on the action vector $\mathbf{a}_i$ and replicates $\{\mathbf{y}_{i,rep}\}_{i=1}^N$ as they are marginalized out by the minimum and expectation functions, respectively, but is dependent on $k > 0$.

Having chosen the quadratic loss $l(\mathbf{y}, \mathbf{a}) = \|\mathbf{y} - \mathbf{a}\|_2^2$ in Eq. 4.6, the generic PPLC D_k is simplified into

$$D_k = \frac{k}{k+1} G + P, \quad k \geq 0, \tag{4.7}$$

where $G = \sum_{i=1}^N \|\boldsymbol{\nu}_i - \mathbf{y}_{i,obs}\|_2^2$ and $P = \sum_{i=1}^N \boldsymbol{\varsigma}_i^2$ represent the goodness-of-fit and penalty terms, respectively, with $\boldsymbol{\nu}_i = \mathbb{E}[\mathbf{y}_{i,rep}|\mathbf{y}_{1:N}]$ and $\boldsymbol{\varsigma}_i^2 = \mathbb{E}[\|\mathbf{y}_{i,rep} - \boldsymbol{\nu}_i\|_2^2|\mathbf{y}_{1:N}]$, $i = 1, \cdots, N$. Eventually, a model with a smaller

value for the D_k (4.7) is preferable. It is known that ordering of models is insensitive to the particular choice of k (Banerjee et al., 2014).

To apply the PPLC (4.7) to our examples, let $\Omega = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \sigma, \{\alpha_l, \boldsymbol{\beta}_l, \sigma_l^2, \boldsymbol{\lambda}_l, \tau_l\}_{l=1}^3\}$ denotes the collection of all the latent variables for each of the models. Then the posterior predictive distribution of $\mathbf{y}_{i,rep}$ is

$$f(\mathbf{y}_{i,rep}|\mathbf{y}_{1:N}) = \int \mathcal{N}_{T_i}(\mathbf{y}_{i,rep}|\boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i}), \sigma^2\mathbf{I})\pi(\Omega|\mathbf{y}_{1:N})d\Omega, \quad i = 1, \cdots, N.$$

To approximate D_k (4.7) for each model, first, choose a number k, saying $k = 1$, and second, approximate $\boldsymbol{\nu}_i$ and $\boldsymbol{\varsigma}_i^2$ through replicates $\mathbf{y}_{i,rep}$ drawn from the predictive density $f(\mathbf{y}_{i,rep}|\mathbf{y}_{1:N})$ (4.5) for each $i = 1, \cdots, N$, and finally, complete the G and P to get an approximated value for the D_k (4.7).

4.4. Results of Model Comparison via Three Criteria Table 2 summarizes the results of model choice criteria to compare the four non-spatial Bayesian hierarchical models, $\mathcal{M}_m$ $(m = 1, 2, 3, 4)$. The smaller numbers are indications for better predictive performance for the three creteria. As for PPLC, we used $k = 1$ for the weight. Based on the results, the Weibull model $\mathcal{M}_1$ (3.2) is selected as the best model, whereas the Duong model $\mathcal{M}_4$ (3.5) is chosen as the worst model across all the three criteria.

5 Spatial Bayesian Hierarchical Model

5.1. Spatial Extension of the Non-Spatial Model Consider a collection of the full production results $\{(\mathbf{P}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$ from N wells in a shale reservoir. An essential idea underlying the extension of the non-spatial Bayesian hierarchical model (that used $\{(\mathbf{P}_i, \mathbf{x}_i)\}_{i=1}^N$) to a spatial one is the incorporation of a spatial random effect involving the location set $\{\mathbf{s}_i\}_{i=1}^N$ to the linear regression (3.6), eventually providing a kriging (Cressie, 1990; Handcock and Stein, 1993) through the latent level. Gaussian process (Abrahamsen, 1997) is used to describe a spatial association between wells. For a technical detail of Gaussian process, refer to Rasmussen (2004) or Chapter 6 of Bishop (2006).

Table 2: DIC, WAIC, and PPLC for the four non-spatial models

	DIC	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

A hierarchy of the proposed spatial Bayesian hierarchical model can be represented by the following two stages:

Stage 1: individual-level model.

$$y_{it} = \mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2). \tag{5.1}$$

Stage 2: population model.

$$\theta_{li} = \theta_l(\mathbf{s}_i) = \alpha_l + \mathbf{x}_i^\top \boldsymbol{\beta}_l + \epsilon_l(\mathbf{s}_i) + \eta_l(\mathbf{s}_i), \quad \epsilon_l(\cdot) \sim \mathcal{GWN}(\sigma_l^2), \tag{5.2}$$

$$\eta_l(\cdot) \sim \mathcal{GP}(0, \mathcal{K}_{\gamma_l}(\cdot,\cdot)), \quad \mathcal{K}_{\gamma_l}(\mathbf{s}_i, \mathbf{s}_j) = \gamma_l^2 \exp(-e^{\rho_l} \|\mathbf{s}_i - \mathbf{s}_j\|_2^2), \tag{5.3}$$

$$\beta_{lj} | \lambda_{lj}, \tau_l, \sigma_l \sim \mathcal{N}(0, \lambda_{lj}^2 \tau_l^2 \sigma_l^2), \tag{5.4}$$

$$\sigma, \gamma_l, \lambda_{lj}, \tau_l \sim \mathcal{C}^+(0, 1), \tag{5.5}$$

$$\alpha_l \sim \pi(\alpha_l) \propto 1, \quad \sigma_l^2 \sim \pi(\sigma_l^2) \propto 1/\sigma_l^2. \tag{5.6}$$

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. As same with the non-spatial models, the first stage (5.1) depicts the well-specific model using $y_{it} = \log \mathsf{P}_{it}$, $\mu(t; \theta_{1i}, \theta_{2i}, \theta_{3i}) = \log\ q(t; M_i, b_i, k_i)$, $\theta_{1i} = \log\ M_i$, $\theta_{2i} = \log\ b_i$, and $\theta_{3i} = \log\ k_i$.

Some ingredients used on the second stage (5.2)–(5.6) are as follows. Notation $\mathcal{GP}(0, \mathcal{K}(\cdot,\cdot))$ stands for a Gaussian process with zero-mean function and covariance function $\mathcal{K}(\cdot,\cdot)$. Notation $\epsilon_l(\cdot) \sim \mathcal{GWN}(\sigma_l^2)$ denotes the Gaussian white noise (also called the nugget-effect in geo-statistics (Banerjee et al., 2014)), defined by $\epsilon_l(\cdot) \sim \mathcal{GP}(0, \sigma_l^2 \mathcal{I}(\cdot,\cdot)))$ where $\mathcal{I}(\cdot,\cdot)$ is the indicator function. The spatial random effect in Eq. 5.3 is described by an isotropic Gaussian process $\eta_l(\cdot) \sim \mathcal{GP}(0, \mathcal{K}_{\gamma_l}(\cdot,\cdot))$ based on a Gaussian kernel $\mathcal{K}_{\gamma_l}(\cdot,\cdot)$, where the γ_l is assumed to follow the unit-scaled half-Cauchy distribution. The range parameters ρ_l $(l = 1, 2, 3)$ are hyper-parameters whose possible optimal values will be chosen via a cross-validation later.

Letting the ρ_l ($l = 1, 2, 3$) to be zeros degrades the model (5.1)–(5.6) from a spatial model to the non-spatial model introduced in Section 3. See Appendix for a posterior computation for the proposed model.

The platform of the spatial model (5.1)–(5.6) is called nonlinear mixed effect model, or hierarchical nonlinear model. This framework has been widely used in biological, agricultural, environmental, epidemiological, and medical applications to make inference on features underlying profiles of continuous, repeated measurements from a sample of individuals from a population of interest (Davidian and Giltinan, 1995; Lee et al., 2020b). For example, in

a population pharmacokinetic model (Sheiner and Ludden, 1992), the response profile y_{it} in Eq. 5.1 is plasma concentration measured from the i-th patient at the t-th time point (a unit of time is usually an hour), and the corresponding mean function $\mu(t;\boldsymbol{\vartheta}_i)$ is some compartment model, involving subject-specific parameter $\boldsymbol{\vartheta}_i$, which helps explain the relationship between these measurements and characteristics of the patient (Rosner and Müller, 1997).

5.2. Spatial Prediction via Latent Kriging A prominent feature of the proposed spatial model (5.1)–(5.6) distinctive from most of the existing nonlinear mixed effect models that have been applied to various applications is the latent kriging to achieve the sub-goal (c) of the shale reservoir problem. The present subsection delineates the procedure about how to spatially forecast a (log-scaled) production rate trajectory of a new well using the latent kriging.

Consider shale oil wells data $\{(\mathbf{y}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$ from N wells as explained in Section 1.2 where the oil production rate trajectories are log-transformed (that is, $y_{it} = \log\ \mathsf{P}_{it}$). Having observed the training data $\{(\mathbf{y}_i, \mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$, our objective is now to predict the temporal profile of production rates during T^* months, that is, $\mathbf{y}^* = (y_1^*, \cdots, y_{T^*}^*)^\top$, of a new well at a new location $\mathbf{s}^* = (s_1^*, s_2^*)^\top$ where p completion predictors are specified by $\mathbf{x}^* = (x_1^*, \cdots, x_p^*)^\top$.

To that end, we need to derive the posterior predictive density, $f(\mathbf{y}^*|\mathbf{y}_{1:N}) = f(\mathbf{y}^*|\{(\mathbf{y}_i, \mathbf{x}_i\ , \mathbf{s}_i)\}_{i=1}^N, \mathbf{s}^*, \mathbf{x}^*)$, which embeds the uncertainty over the latent variables of the model by taking all likely values, having seen the training wells data into account. We concretize the derivation procedure step by step as follows.

Let $(\theta_1^*, \theta_2^*, \theta_3^*)^\top \in \mathbb{R}^3$ denotes the curve parameters for the new well. By the likelihood assumption (5.1), predicted response is distributed according to

$$\mathbf{y}^*|\theta_1^*, \theta_2^*, \theta_3^*, \sigma^2 \sim N_{T^*}(\boldsymbol{\mu}^*(\theta_1^*, \theta_2^*, \theta_3^*), \sigma^2\mathbf{I}), \tag{5.7}$$

where the function $\boldsymbol{\mu}^*(\theta_1^*, \theta_2^*, \theta_3^*) = (\mu(1;\theta_1^*, \theta_2^*, \theta_3^*), \cdots, \mu(T^*;\theta_1^*, \theta_2^*, \theta_3^*))^\top$ is a T^*-dimensional vector valued function with respect to $(\theta_1^*, \theta_2^*, \theta_3^*)$. By using the linear regression assumption (5.2) indexed with $l = 1, 2, 3$, we can express the joint density of curve parameters of the new and training wells

$$\begin{bmatrix}\theta_l^* \\ \boldsymbol{\theta}_l\end{bmatrix} \Bigg| \alpha_l, \boldsymbol{\beta}_l, \sigma_l^2, \gamma_l \sim N_{N+1}\left(\begin{bmatrix}\alpha_l + (\mathbf{x}^*)^\top\boldsymbol{\beta}_l \\ \mathbf{1}\alpha_l + \mathbf{X}\boldsymbol{\beta}_l\end{bmatrix}, \begin{bmatrix}\sigma_l^2 + \gamma_l^2 & \gamma_l^2\mathbf{b}^\top \\ \gamma_l^2\mathbf{b} & \sigma_l^2\mathbf{I} + \gamma_l^2\mathbf{B}(\rho_l)\end{bmatrix}\right), \tag{5.8}$$

where $\mathbf{b}$ is a N-dimensional vector whose i-th entry is given by $\exp(-e^{\rho_l}\|\mathbf{s}_i - \mathbf{s}^*\|_2^2)$ $(i = 1, \cdots, N)$, and $\mathbf{B}(\rho_l)$ is a N-by-N matrix whose ij-th entry is

given by $\exp(-e^{\rho_l}\|\mathbf{s}_i - \mathbf{s}_j\|_2^2)$ $(i, j = 1, \cdots, N)$. By carrying out the conditional property of the Gaussian distribution to each of the joint density (5.8) indexed with $l = 1, 2, 3$, we can express the conditional distribution of the curve parameter of the new well given those from the training wells

$$\theta_l^*|\boldsymbol{\theta}_l, \alpha_l, \boldsymbol{\beta}_l, \sigma_l^2, \gamma_l \sim \mathcal{N}(\mu_l^*(\boldsymbol{\theta}_l, \alpha_l, \boldsymbol{\beta}_l, \sigma_l^2, \gamma_l), (\sigma_l^2)^*(\sigma_l^2, \gamma_l)),$$

where the mean and variance functions are given by

$$\begin{aligned}
\mu_l^*(\boldsymbol{\theta_l}, \alpha_l, \boldsymbol{\beta}_l, \sigma_l^2, \gamma_l) &= \alpha_l + (\mathbf{x}^*)^\top \boldsymbol{\beta}_l + \gamma_l^2 \mathbf{b}^\top [\sigma_l^2 \mathbf{I} + \gamma_l^2 \mathbf{B}(\rho_l)]^{-1} (\boldsymbol{\theta}_l - \mathbf{1}\alpha_l - \mathbf{X}\boldsymbol{\beta}_l), \\
(\sigma_l^2)^*(\sigma_l^2, \gamma_l) &= \sigma_l^2 + \gamma_l^2 - (\gamma_l^2)^2 \mathbf{b}^\top [\sigma_l^2 \mathbf{I} + \gamma_l^2 \mathbf{B}(\rho_l)]^{-1} \mathbf{b},
\end{aligned}$$

that are also referred to as kriging mean and variance, respectively, in geostatistical analysis; for a detail, refer to page 41 of Gelfand et al. (2010).

To avoid notational clutter, denote the collection of all the latent variables involved in the prediction as $\Psi = (\theta_1^*, \boldsymbol{\theta}_1, \theta_2^*, \boldsymbol{\theta}_2, \theta_3^*, \boldsymbol{\theta}_3, \sigma, \{\alpha_l, \boldsymbol{\beta}_l, \sigma_l^2, \boldsymbol{\lambda}_l, \tau_l, \gamma_l\}_{l=1}^3)$, and let Ω denotes the collection of all the latent variables used in the posterior inference, that is, $\Omega = \Psi - \{\theta_1^*, \theta_2^*, \theta_3^*\}$. Now, we can express the predictive density $f(\mathbf{y}^*|\mathbf{y}_{1:N})$ by an integral form where all the latent variables in the Ψ are marginalized out:

$$f(\mathbf{y}^*|\mathbf{y}_{1:N}) = \int p(\mathbf{y}^*, \Psi|\mathbf{y}_{1:N}) d\Psi = \int p(\mathbf{y}^*|\Psi, \mathbf{y}_{1:N}) \cdot p(\Psi|\mathbf{y}_{1:N}) d\Psi. \quad (5.9)$$

The first component of integrand in Eq. 5.9 coincides with Eq. 5.7 because of the conditional independence assumption induced from the likelihood (5.1), hence, we have $p(\mathbf{y}^*|\Psi, \mathbf{y}_{1:N}) = p(\mathbf{y}^*|\theta_1^*, \theta_2^*, \theta_3^*, \sigma) = N_{T^*}(\mathbf{y}^*|\boldsymbol{\mu}^*(\theta_1^*, \theta_2^*, \theta_3^*), \sigma^2\mathbf{I})$. The second component of integrand in Eq. 5.9 can be simplified to

$$\begin{aligned}
p(\Psi|\mathbf{y}_{1:N}) &= p(\Omega, \theta_1^*, \theta_2^*, \theta_3^*|\mathbf{y}_{1:N}) = p(\theta_1^*, \theta_2^*, \theta_3^*|\Omega, \mathbf{y}_{1:N}) \pi(\Omega|\mathbf{y}_{1:N}) \\
&= \left\{ \prod_{l=1}^3 p(\theta_l^*|\boldsymbol{\theta}_l, \alpha_l, \boldsymbol{\beta}_l, \sigma_l^2, \gamma_l) \right\} \pi(\Omega|\mathbf{y}_{1:N}) \\
&= \left\{ \prod_{l=1}^3 \mathcal{N}(\theta_l^*|\mu_l^*(\boldsymbol{\theta}_l, \alpha_l, \boldsymbol{\beta}_l, \sigma_l^2, \gamma_l), (\sigma_l^2)^*(\sigma_l^2, \gamma_l)) \right\} \pi(\Omega|\mathbf{y}_{1:N}).
\end{aligned}$$

To summarize, the posterior predictive density (5.9) can be rewritten as

$$f(\mathbf{y}^*|\mathbf{y}_{1:N}) = \int N_{T^*}(\mathbf{y}^*|\boldsymbol{\mu}^*(\theta_1^*,\theta_2^*,\theta_3^*),\sigma^2\mathbf{I}) \cdot\left\{\prod_{l=1}^{3}\mathcal{N}(\theta_l^*|\mu_l^*(\boldsymbol{\theta_l},\alpha_l,\boldsymbol{\beta}_l,\sigma_l^2,\gamma_l),(\sigma_l^2)^*(\sigma_l^2,\gamma_l))\right\}\pi(\Omega|\mathbf{y}_{1:N})d\Psi. \tag{5.10}$$

Note that the second component of the integrand in Eq. 5.10 produces the latent kriging separately for each of the curve parameters indexed by $l = 1, 2, 3$, and then the latent kriging predictors dictate the mean part of the future production rate trajectory of a new well.

Because the density $f(\mathbf{y}^*|\mathbf{y}_{1:N})$ (5.10) is not expressed in a closed-form, we use a Monte Carlo simulation to approximate the density. Given realized S number of samples $\{(\mathbf{y}^*)^{(s)}\}_{s=1}^S \sim f(\mathbf{y}^*|\mathbf{y}_{1:N})$ (5.10), we recommend to use the posterior predictive median $\widehat{\mathbf{y}}_{i,\text{median}}^*$ to represent a predicted trajectory rather than the posterior predictive mean because the former produces more robust outcomes in current application. A spatial prediction corresponding to the original scale of the production rates $\mathbf{P}^* = (\mathsf{P}_1^*,\cdots,\mathsf{P}_{T^*}^*)^\top$ can be obtained by transforming the realized samples $\{(\mathbf{y}^*)^{(s)}\}_{s=1}^S$ through $\mathsf{P} = \exp(y)$, thereby, leading to $\{(\mathbf{P}^*)^{(s)}\}_{s=1}^S = \{((\mathsf{P}_1^*)^{(s)},\cdots,(\mathsf{P}_{T^*}^*)^{(s)})\}_{s=1}^S \sim f(\mathbf{P}^*|\mathbf{y}_{1:N})$.

As an extended application, we can also predict the estimated ultimate recovery (EUR). In petroleum engineering, the EUR is defined as an approximated quantity of oil from a well which is potentially recoverable by the end of its producing life (Currie et al., 2010). In the oil and gas industry, uncertainty quantification of EUR is of the utmost important procedure and a starting point in the decision-making process for future drilling projects. Also, the oil and gas companies should comply with financial regulations about EUR outlined by the U.S. Securities and Exchange Commission: see www.sec.gov/- for the regulations. Let EUR^* represents a predicted EUR of a new well, defined by a summation of monthly oil production rates over 30-year period (360 months) at a new location $\mathbf{s}^*$ given p completion predictor $\mathbf{x}^*$. (That is, $\text{EUR}^* = \sum_{t=1}^{T^*}\mathsf{P}_t^*$ with $T^* = 360$.) The EUR^* can be also approximated by using a Monte Carlo simulation based on the realized samples $\{(\mathbf{P}^*)^{(s)}\}_{s=1}^S$.

6 Results

Recall that the Weibull model $\mathcal{M}_1$ (3.2) was chosen as the best model among the four non-spatial hierarchical models ($\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$ (3.2)–(3.5)) in Section 4.4. In what follows, we describe results based on

the Weibull model, and for the purpose of articulation the spatial Bayesian hierarchical model (5.1)–(5.6) where the RDC is chosen by the Weibull model is referred to as *spatial Weibull model*. Its non-spatial version is simply obtained by setting the γ_l, $l = 1, 2, 3$, in Eq. 5.3 to be zeros. Some relevant R codes to implement the spatial prediction based on the spatial Weibull model are available on https://github.com/yain22/SWM.

6.1. Spatial Prediction Versus Non-Spatial Prediction The objective of this subsection is two-fold: first, to compare predictive performances of the spatial and non-spatial Weibull models; and second, to choose some reasonable range parameters, ρ_l ($l = 1, 2, 3$), of the spatial Weibull model. Particularly, we can check whether the latent kriging helps improve the performance or not through the first objective. Recall that the total number of shale oil wells is $N = 360$. To evaluate the performances at different scenarios, we set the number of test wells, denoted by N_{T}, to be each of $36, 43, 54, 72, 90$, and 180, respectively: then, the remaining wells are automatically used as training wells. (For example, for the first setting, the number of training wells shall be $360 - N_{\mathrm{T}} = 360 - 36 = 324$.) To simplify the experiments, we use identical range parameters involved in the spatial random effects, that is, $\rho_1 = \rho_2 = \rho_3$, and fix them with -10, -4, -1, 0, and 1.

For evaluation criteria, we calculate mean squared error (MSE) evaluated at a specified number of test wells $N_{\mathrm{T}} \in \{36, 43, 54, 72, 90, 180\}$:

$$\mathrm{MSE} = \frac{1}{N_{\mathrm{T}}} \sum_{i=1}^{N_{\mathrm{T}}} \|\widehat{\mathbf{y}}^*_{i,\mathrm{median}} - \mathbf{y}_i\|_2^2,$$

where $\widehat{\mathbf{y}}^*_{i,\mathrm{median}} = (\widehat{y}^*_{i1}, \cdots, \widehat{y}^*_{iT_i})^\top \in \mathbb{R}^{T_i}$ is the posterior predictive median for the test data $\mathbf{y}_i = (y_{i1}, \cdots, y_{iT_i})$ obtained from the posterior predictive density $f(\mathbf{y}^*|\mathbf{y}_{1:N})$ (5.10) as illustrated in Section 5.2. A smaller value of the MSE indicates a better predictive accuracy.

Figure 6 displays MSEs of the non-spatial and spatial Weibull models across different number of test wells N_{T}. Based on the results, we conclude that the spatial Weibull model *universally* outperforms its non-spatial counterpart regardless of the number of test wells. Additionally, the gap between the MSEs of the spatial and non-spatial Weibull models tends to increase as the number of test wells increases. Particularly, when ρ_l ($l = 1, 2, 3$) are fixed with -4, the MSEs attain the smallest values across the number of the test wells. Henceforth, we use these values, $\rho_1 = \rho_2 = \rho_3 = -4$, as our default choice for the range parameters.

6.2. Bayesian Decline Curve Analysis A Bayesian version of DCA, or shortly Bayesian DCA, based on the spatial Weibull model can describe

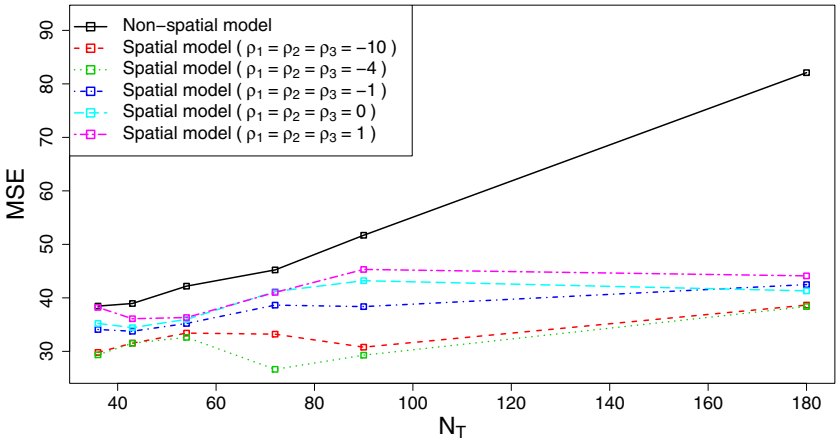

Figure 6: MSEs of non-spatial and spatial Weibull models across different number of test wells. The range parameters of the spatial model are identically set by -10, -4, -1, 0, and 1. MSE attains the smallest values when the range parameters are -4 across different number of test wells

a hidden pattern of production profile at an individual well and reservoir levels. Bayesian DCA utilizes posterior samples realized from the full joint posterior distribution. Suppose that $\{(M_i)^{(s)}, (b_i)^{(s)}, (k_i)^{(s)}\}_{s=1}^{S}$ is S number of posterior samples from posterior density $\pi(M_i, b_i, k_i|\mathbf{y}_{1:N})$ corresponding to the i-th well. Then Bayesian DCA at the individual well is obtained by

$$\begin{aligned}\widehat{q}(t; M_i, b_i, k_i) &= \frac{1}{S}\sum_{s=1}^{S} q(t; (M_i)^{(s)}, (b_i)^{(s)}, (k_i)^{(s)}) \\ &\approx \mathbb{E}[q(t; M_i, b_i, k_i)|\mathbf{y}_{1:N}], \quad t = 1, \cdots, T_i,\end{aligned}$$

Similarly, Bayesian DCA at the reservoir level can be obtained by using posterior samples of the intercept terms α_l $(l = 1, 2, 3)$ in the regressions (5.2). Note that the Bayesian DCA is different from the traditional DCA explained in Section 2.5 in the sense that the parameters of interests are integrated out rather than relying on a plug-in method.

Panels in Fig. 7 displays the estimation results of the Bayesian DCA for the well with API10 = 4201334341 (left panel) and Eagle Ford Shale reservoir (right panel). To train the spatial Weibull model, we used the total $N = 360$ wells.

6.3. Identification of significant completion predictors Figure 8 shows the result of posterior inferences for the coefficients in spatial linear regression (5.2) obtained by training the total 360 wells to the spatial Weibull

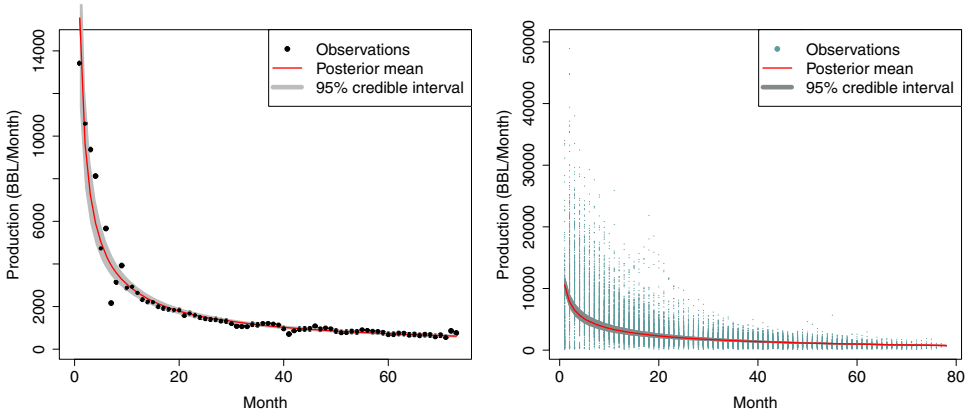

Figure 7: Estimation results of Bayesian decline curve analysis at an individual well (left) and reservoir levels (right). Shaded regions on both panels represent the pointwise posterior 95% credible intervals

model. The blue vertical bars in each panel represent the posterior 95% credible intervals corresponding to the elements of the $\boldsymbol{\beta}_l = (\beta_{l1}, \cdots, \beta_{lj}, \cdots, \beta_{lp})^\top \in \mathbb{R}^p$, $(p = 11;\ l = 1\,(\text{left}), 2\,(\text{middle}), 3\,(\text{right}))$, where the covariate index j follows the Table 1.

The followings are some summaries of the results. (Technical interpretation is omitted and left to petroleum engineers.) First, as for the log-scaled carrying capacity, $\theta_1 = \log\ M$, true vertical depth and the first test oil volume indexed by $j = 4$ and 9 are selected as important well completion

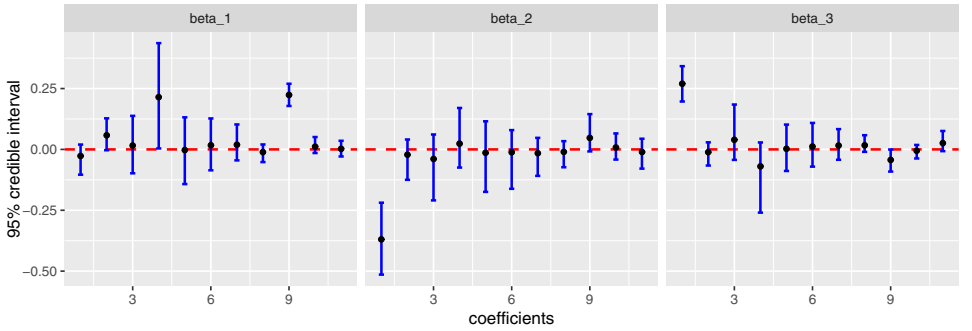

Figure 8: Posterior 95% credible intervals for the coefficients: $\boldsymbol{\beta}_1$ (left), $\boldsymbol{\beta}_2$ (middle), and $\boldsymbol{\beta}_3$ (right). The symbol • represents the posterior mean. The integer numbers on the x-axis of the plots correspond the index j of the 11 completion predictors in Table 1

predictors. Second, commonly from the log-scaled scale and shape parameters, $\theta_2 = \log\ b$ and $\theta_3 = \log\ k$, the first test flowing tubing pressure has been found to be the most important completion predictor indexed with $j = 1$.

6.4. Spatial Predictions at Two New Locations We use 324 wells as a training dataset and predict the oil production decline curve at two test locations. The locations of the training and two test wells are displayed on the map in Fig. 9. API10 numbers of the two test wells are: well-1 (API10 = 4249332630) and well-2 (API10 = 4231135594). Note that the well-1 is positioned near a certain group of training wells, while the well-2 is more isolatedly located, relatively far from other neighboring wells.

Panels in Fig. 10 display the prediction results. As a metric, we report root-mean-square deviation (RMSD) at individual well level, measured via $(\|\widehat{\mathbf{y}}^*_{i,\text{median}} - \mathbf{y}_i\|_2^2/T_i)^{1/2} = \{(1/T_i)\cdot\sum_{t=1}^{T_i}(\widehat{y}^*_{it,\text{median}} - y_{it})^2\}^{1/2}$: a lower value for the RMSD indicates a better predictive accuracy among considered test wells. RMSDs of the two test wells based on the spatial Weibull model are 0.178 (well-1) and 0.305 (well-2). From the panels, we can also observe that pointwise posterior predictive 95% credible interval of the production rates of the well-1 is relatively narrower than that of the well-2 during the production period. To see the benefit of the latent kriging (5.3), we also measured RMSDs based on the non-spatial Weibull model, which resulted in 0.180 (well-1) and 0.348 (well-2), higher than those based on the spatial one, respectively.

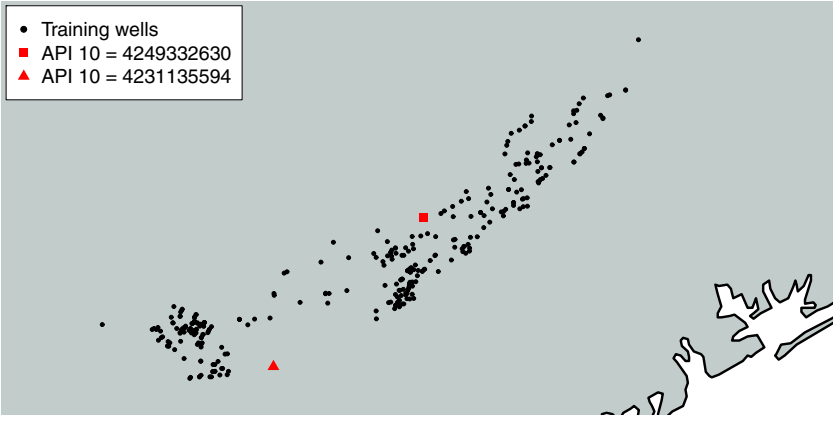

Figure 9: Locations of 324 training wells and two test wells. The symbols ■ and ▲ represent the locations of well-1 (API10=4249332630) and well-2 (API10=4231135594), respectively

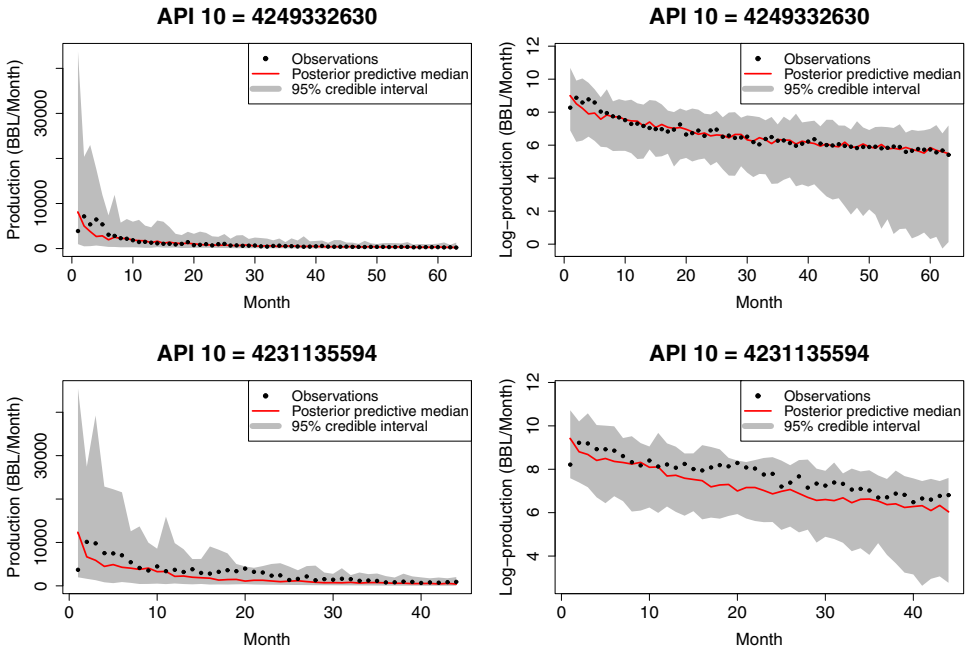

Figure 10: Spatial predictions for two test wells: well-1 (top) and well-2 (bottom). The left and right panels are displayed in the original and log-transformed scales for the oil production rates, respectively

Table 3 summarizes prediction results for the EUR for the two test wells: posterior mean, posterior median, and posterior predictive 95% credible interval of EUR*. Because either of the wells did not produce oil for 30 years, there is no reference value to validate a predictive accuracy. Posterior predictive 95% credible interval of EUR* of the well-1 is narrower than that of the well-2.

Based on the prediction outcomes of the two test wells and groupwise comparison in Section 6.1, we conclude that the latent kriging helps improve the predictive accuracy (the RMSD and MSE) and quality (the length of 95% credible interval). This means that the spatial model (5.1)–(5.6) takes

Table 3: 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)

NOTE: Unit of the measurements is Barrel over 30 years

an advantage of utilization of the geological proximity information in the prediction via the latent kriging. More concretely, spatial prediction at the location of the well-1 borrows some useful information from neighboring wells, while the isolated position of the well-2 renders its prediction sub-optimal. Having no use for the location information forfeits the benefit.

7 Discussion

In this paper, we explored shale oil production results from a shale reservoir and described some characteristics of the obtainable dataset, and suggested a "shale reservoir problem" that can be useful for the upstream petroleum industry. Drawing attention to a limitation of the traditional DCA that only utilizes the temporal profile from the production results, we eventually proposed a new platform of a spatial model that harnesses the full aspects of the shale oil wells data with a fully Bayesian approach. One of the notable results is the superiority of the spatial Weibull model over its non-spatial counterpart, which elucidates that reasonable exploitation of geological information can bring an advantage to better forecasting the future behavior of oil or gas production before actual drilling takes place. This can save a large degree of well construction costs.

We regard that the followings are salient contributions of the paper: first, we clarified the shale oil wells dataset and articulated a recent request from the petroleum industry in the Era of Big Data from the perspective of statisticians so as to call for more interdisciplinary researches across statistics and petroleum engineering communities; and second, we extended the platform of the traditional DCA to that of non-linear mixed effect models endowed with the latent kriging, perhaps a new trial in the interdisciplinary works.

Some future directions of research are as follows. First, provided the presence of a strong serial correlation in the data level (5.1), one may include autoregressive parameters in the error assumption or introduce nonhomogeneous Ornstein-Uhlenbeck process (Zhang et al., 1998) to model such a correlation. Second, the notion of the seemingly unrelated regressions (Fiebig, 2001) can be used instead of the separate regressions (5.2) to associate interaction between the individual regressions. Third, one may seek a fully Bayesian approach to estimate the range parameters in Eq. 5.3. Fourth, various correlation functions, for example, the Matérn correlation function or an anisotropic covariance function (Banerjee et al., 2014), can be used to improve the predictive performance using the latent kriging. Fifth, to enhance the computational speed, variational methods (Blei et al., 2017; Bishop, 2006) can be used to approximate the posterior distribution; this

new optimization technique will be particularly useful when analyzing an enormously large number of shale oil wells in a reservoir region. Lastly, it will be interesting to accommodate some time-varying covariates such as water or gas production rate time series data to adjust oil production decline curve in the curve-fitting. These time-varying covariates may be used in data level (5.1) endowed with center-adjusted inference or prediction (Li et al., 2010).

Funding. The research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award number R01CA194391, NSF TRIPODS grant no. NSF CCF-1934904, and by NSF grant no. IIS-1741173.

Compliance with Ethical Standards.

We declare that we have no conflict of interest.

The source of data is from Drillinginfo. The data is not publically available and there can be a cost associated with it charged by the company. Users can purchase the data from the website or can get some similar data and will be able to use our codes.

R codes to implement the spatial prediction based on the spatial Weibull model are publically available on https://github.com/yain22/SWM.

References

ABRAHAMSEN, P. (1997). *A review of gaussian random fields and correlation functions.* Norsk Regnesentral/Norwegian Computing Center Oslo.

AKAIKE, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199–213): Springer.

AL-HADDAD, S. M., CRAFTON, J. W. and ET AL. (1991). Productivity of horizontal wells. In *Low Permeability Reservoirs Symposium: Society of Petroleum Engineers.*

ALI, T. A., SHENG, J. J. and ET AL. (2015). Production decline models: A comparison study. In *SPE Eastern Regional Meeting*: Society of Petroleum Engineers.

ANDRIEU, C., DEFREITAS, N., DOUCET, A. and JORDAN, M. I. (2003). An introduction to mcmc for machine learning. *Machine Learning* **50**, 1-2, 5–43.

ARMAGAN, A., DUNSON, D. B. and LEE, J. (2013). Generalized double pareto shrinkage. *Statistica Sinica* **23**, 1, 119.

ARPS, J. J. and ET AL. (1945). Analysis of decline curves. *Transactions of the AIME* **160**, 01, 228–247.

BAE, K. and MALLICK, B. K. (2004). Gene selection using a two-level hierarchical bayesian model. *Bioinformatics* **20**, 18, 3423–3430.

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2014). *Hierarchical modeling and analysis for spatial data.* CRC press.

BELLARBY, J. (2009). *Well completion design.* Elsevier.

BHATTACHARYA, A., PATI, D., PILLAI, N. S. and DUNSON, D. B. (2015). Dirichlet–laplace priors for optimal shrinkage. *Journal of the American Statistical Association* **110**, 512, 1479–1490.

BISHOP, C. M. (2006). *Pattern recognition and machine learning.* Springer.

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association* **112**, 518, 859–877.

BOX, G. E. P. (1980). Sampling and bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A (General)*, 383–430.

BOYD, S., BOYD, S. P. and VANDENBERGHE, L. (2004). *Convex optimization*. Cambridge university press.

CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2009). Handling sparsity via the horseshoe. In *Artificial Intelligence and Statistics* (pp. 73–80).

CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97**, 2, 465–480.

CASELLA, G. and GEORGE, E. I. (1992). Explaining the gibbs sampler. *The American Statistician* **46**, 3, 167–174.

CELEUX, G., FORBES, F., ROBERT, C. P., TITTERINGTON, D. M. and ET AL. (2006). Deviance information criteria for missing data models. *Bayesian Analysis* **1**, 4, 651–673.

CHENG, Y., WANG, Y., MCVAY, D., LEE, W. J. and ET AL. (2010). Practical application of a probabilistic approach to estimate reserves using production decline data. *SPE Economics & Management* **2**, 01, 19–31.

CHIB, S. and GREENBERG, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician* **49**, 4, 327–335.

CLARK, A. J. (2011). Decline curve analysis in unconventional resource plays using logistic growth models. Ph.D. Thesis.

CRESSIE, N. (1990). The origins of kriging. *Mathematical Geology* **22**, 3, 239–252.

CURRIE, S. M., ILK, D., BLASINGAME, T. and ET AL. (2010). Continuous estimation of ultimate recovery. In *SPE Unconventional Gas Conference*: Society of Petroleum Engineers.

DAVIDIAN, M. and GILTINAN, D. M. (1995). *Nonlinear models for repeated measurement data, 62*. CRC press.

DEHAAN, L. and FERREIRA, A. (2007). *Extreme value theory: an introduction*. Springer Science & Business Media.

DEEN, T., DAAL, J., TUCKER, J. and ET AL. (2015). Maximizing well deliverability in the eagle ford shale through flowback operations. In *SPE Annual Technical Conference and Exhibition*: Society of Petroleum Engineers.

DUONG, A. N. and ET AL. (2011). Rate-decline analysis for fracture-dominated shale reservoirs. *SPE Reservoir Evaluation & Engineering* **14**, 03, 377–387.

EIA, U. S., OIL, U. S. C. and RESERVES, N. G. P. (2019). Year-end 2018 (december 13, 2019).

FETKOVICH, M. J. and ET AL. (1980). Decline curve analysis using type curves. *Journal of Petroleum Technology* **32**, 06, 1–065.

FIEBIG, D. G. (2001). Seemingly unrelated regression. *A companion to theoretical econometrics*, 101–121.

FRÉCHET, M. (1928). Sur la loi de probabilité de l'écart maximum. In *Annales de la societe Polonaise de Mathematique*: [sn].

GANDOSSI, L. and ET AL. (2013). An overview of hydraulic fracturing and other formation stimulation technologies for shale gas production. Eur. Commisison Jt. Res. Cent. Tech. Reports 26347.

GELFAND, A. E. and GHOSH, S. K. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika* **85**, 1, 1–11.

GELFAND, A. E., DIGGLE, P., GUTTORP, P. and FUENTES, M. (2010). *Handbook of spatial statistics*. CRC press.

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2004). *Bayesian data analysis*. Chapman and Hall/CRC.

GELMAN, A., HWANG, J. and VEHTARI, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing* **24**, 6, 997–1016.

GEORGE, E. I. and MCCULLOCH, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association* **88**, 423, 881–889.

GIGER, F. M., REISS, L. H., JOURDAN, A. P. and ET AL. (1984). The reservoir engineering aspects of horizontal drilling. In *SPE Annual Technical Conference and Exhibition*: Society of Petroleum Engineers.

GONG, X., GONZALEZ, R. A., MCVAY, D., HART, J. D. and ET AL. (2011). Bayesian probabilistic decline curve analysis quantifies shale gas reserves uncertainty. In *Canadian Unconventional Resources Conference*: Society of Petroleum Engineers.

GONG, X., GONZALEZ, R., MCVAY, D. A., HART, J. D. and ET AL. (2014). Bayesian probabilistic decline-curve analysis reliably quantifies uncertainty in shale-well-production forecasts. *SPE Journal* **19**, 06, 1–047.

GRADL, C. (2018). Review of recent unconventional completion innovations and their applicability to egs wells.

HANDCOCK, M. S. and STEIN, M. L. (1993). A bayesian analysis of kriging. *Technometrics* **35**, 4, 403–410.

HARRIS, S., LEE, W. J. and ET AL. (2014). A study of decline curve analysis in the elm coulee field. In *SPE Unconventional Resources Conference*: Society of Petroleum Engineers.

HOFF, P. D. (2009). *A first course in bayesian statistical methods, 580*. Springer.

HOWARTH, R. W., INGRAFFEA, A. and ENGELDER, T. (2011). Natural gas: Should fracking stop?. *Nature* **477**, 7364, 271.

HUBBERT, M. K. and WILLIS, D. G. (1972). Mechanics of hydraulic fracturing.

ILK, D., RUSHING, J. A., PEREGO, A. D., BLASINGAME, T. A. and ET AL. (2008). Exponential vs. hyperbolic decline in tight gas sands: understanding the origin and implications for reserve estimates using arps' decline curves. In *SPE annual technical conference and exhibition*: Society of Petroleum Engineers.

ISHWARAN, H., RAO, J. S. and ET AL. (2005). Spike and slab variable selection: frequentist and bayesian strategies. *The Annals of Statistics* **33**, 2, 730–773.

JOCHEN, V. A., SPIVEY, J. P. and ET AL. (1996). Probabilistic reserves estimation using decline curve analysis with the bootstrap method. In *SPE Annual Technical Conference and Exhibition*: Society of Petroleum Engineers.

JOSHI, S. D. (1991). *Horizontal well technology*. PennWell Books.

JOSHI, S. D. and ET AL. (2003). Cost/benefits of horizontal wells. In *SPE Western Regional/AAPG Pacific Section Joint Meeting*: Society of Petroleum Engineers.

KAZEMI, H. and ET AL. (1969). Pressure transient analysis of naturally fractured reservoirs with uniform fracture distribution. *Society of Petroleum Engineers Journal* **9**, 04, 451–462.

LARKEY, C. S. (1925). Mathematical determination of production decline curves. *Trans AIME* **71**, 1322, 69–21.

LEE, S. and KIM, J. H. T. (2019). Exponentiated generalized pareto distribution: Properties and applications towards extreme value theory. *Communications in Statistics-Theory and Methods* **48**, 8, 2014–2038.

LEE, S. Y. (2020a). A note on Gibbs sampler and coordinate ascent variational inference. arXiv:2008.01006.

LEE, S. Y., LEI, B. and MALLICK, B. (2020b). Estimation of covid-19 spread curves integrating global data and borrowing information. *PLOS ONE* **15**, 7, 1–17.

lee, s. y., pati, d. and mallick, b. k. (2020c). Tail-adaptive Bayesian shrinkage. arXiv:2007.02192.

lewis, j. o., beal, c. h. and et al. (1918). Some new methods for estimating the future production of oil wells. *Transactions of the AIME* **59**, 01, 492–525.

li, y., lin, x. and müller, p. (2010). Bayesian inference in semiparametric mixed models for longitudinal data. *Biometrics* **66**, 1, 70–78.

makalic, e. and schmidt, d. f. (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters* **23**, 1, 179–182.

mcnulty, r. r., knapp, r. m. and et al. (1981). Statistical decline curve analysis. In *SPE Annual Technical Conference and Exhibition*: Society of Petroleum Engineers.

miao, y., li, x., lee, j., zhao, c., zhou, y., li, h., chang, y., lin, w., xiao, z., wu, n. and et al. (2018). Comparison of various rate-decline analysis models for horizontal wells with multiple fractures in shale gas reservoirs. In *SPE Trinidad and Tobago Section Energy Resources Conference*: Society of Petroleum Engineers.

mishra, s. (2012a). A new approach to reserves estimation in shale gas reservoirs using multiple decline curve analysis models. *Society of Petroleum Engineers*.

mishra, s. and et al. (2012b). A new approach to reserves estimation in shale gas reservoirs using multiple decline curve analysis models. In *SPE Eastern Regional Meeting*: Society of Petroleum Engineers.

mishra, s. and datta-gupta, a. (2017). *Applied statistical modeling and data analytics: A practical guide for the petroleum geosciences*. Elsevier.

mitchell, t. j. and beauchamp, j. j. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association* **83**, 404, 1023–1032.

montgomery, c. t., smith, m. b. and et al. (2010). Hydraulic fracturing: history of an enduring technology. *Journal of Petroleum Technology* **62**, 12, 26–40.

mukherjee, h., economides, m. j. and et al. (1991). A parametric comparison of horizontal and vertical well performance. *SPE Formation Evaluation* **6**, 02, 209–216.

o'hara, r. b., sillanpää, m.j. and et al. (2009). A review of bayesian variable selection methods: what, how and which. *Bayesian Analysis* **4**, 1, 85–117.

park, t. and casella, g. (2008). The bayesian lasso. *Journal of the American Statistical Association* **103**, 482, 681–686.

park, j., iino, a., datta-gupta, a., bi, j. and sankaran, s. (2021). Novel hybrid fast marching method-based simulation workflow for rapid history matching and completion design optimization of hydraulically fractured shale wells. *Journal of Petroleum Science and Engineering* **196**, 107718.

pickands, iii., j. and et al. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics* **3**, 1, 119–131.

polson, n. g. and scott, j. g. (2010). Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian Atatistics* **9**, 501-538, 105.

rasmussen, c. e. (2004). Gaussian processes in machine learning. In *Advanced lectures on machine learning* (pp. 63–71): Springer.

robert, c. and casella, g. (2013). *Monte carlo statistical methods*. Springer Science & Business Media.

rosner, g. l. and müller, p (1997). Bayesian population pharmacokinetic and pharmacodynamic analyses using mixture models. *Journal of Pharmacokinetics and Biopharmaceutics* **25**, 2, 209–233.

SCHUETTER, J., MISHRA*, S, ZHONG, M. and LAFOLLETTE, R. (2015). Data analytics for production optimization in unconventional reservoirs. In *Unconventional Resources Technology Conference* (pp. 249–269). San Antonio: Society of Exploration Geophysicists, American Association of Petroleum.

SHEINER, L. B. and LUDDEN, T. M. (1992). Population pharmacokinetics/dynamics. *Annual Review of Pharmacology and Toxicology* **32**, 1, 185–209.

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 4, 583–639.

TIBSHIRANI, R. J. and EFRON, B. (1993). An introduction to the bootstrap. *Monographs on Statistics and Applied Probability* **57**, 1–436.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 1, 267–288.

VALDES, A., MCVAY, D. A., NOYNAERT, S. F. and ET AL. (2013). Uncertainty quantification improves well construction cost estimation in unconventional reservoirs. In *SPE Unconventional Resources Conference Canada*: Society of Petroleum Engineers.

VALKÓ, P.P., LEE, W. J. and ET AL. (2010). A better way to forecast production from unconventional gas wells. In *SPE Annual technical conference and exhibition*: Society of Petroleum Engineers.

VAN DER PAS, S. L., KLEIJN, B. J. K., VAN DER VAART, A. W. and ET AL. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics* **8**, 2, 2585–2618.

VEHTARI, A. and GELMAN, A. (2014). Waic and cross-validation in stan, Aalto University, Helsinki.

VYAS, A. (2017). Application of machine learning and genetics algorithm in reservoir characterization. Ph.D. Thesis, Texas A & M University.

WAINWRIGHT, M. J. (2019). High-dimensional statistics: A non-asymptotic viewpoint, 48. Cambridge University Press.

WAN, R. (2011). *Advanced well completion engineering*. Gulf professional publishing.

WATANABE, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**, 3571–3594.

WEIBULL, W. and ET AL. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics* **18**, 3, 293–297.

ZELLNER, A. (1994). Bayesian and non-bayesian estimation using balanced loss functions. In Statistical decision theory and related topics V (pp. 377–390): Springer.

ZHANG, D., LIN, X., RAZ, J. and SOWERS, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association* **93**, 442, 710–719.

ZHANG, H., COCCO, M., RIETZ, D., CAGLE, A., LEE, J. and ET AL. (2015). An empirical extended exponential decline curve for shale reservoirs. In: SPE Annual Technical Conference and Exhibition: Society of Petroleum Engineers.

Appendix : posterior computation

A.1 A vector-form expression of the spatial model

The present section aims to provide a full description of a posterior computation for the spatial Bayesian hierarchical model (5.1)–(5.6) through a MCMC sampling technique (Robert and Casella, 2013). To start with, we shall integrate out the Gaussian process $\eta_l(\cdot)$ in each of the spatial linear regressions (5.2) indexed with $l = 1, 2$ and 3 to alleviate the computational burden associated with sampling from multivariate Gaussian distribution involving the spatial random effects. After the marginalization, a hierarchical structure of the proposed model can be expressed in a vector-form stage-wisely.

Stage 1: individual-level model.

$$\mathbf{y}_i|\theta_{1i},\theta_{2i},\theta_{3i},\sigma \sim \mathcal{N}_{T_i}(\boldsymbol{\mu}_i(\theta_{1i},\theta_{2i},\theta_{3i}),\sigma^2\mathbf{I}), \qquad (i=1,\cdots,N). \quad \text{(A.1)}$$

Stage 2: population model.

$$\boldsymbol{\theta}_l|\alpha_l,\boldsymbol{\beta}_l,\sigma_l^2,\gamma_l \sim \mathcal{N}_N(\mathbf{1}\alpha_l+\mathbf{X}\boldsymbol{\beta}_l,\sigma_l^2\mathbf{I}+\gamma_l^2\mathbf{B}(\rho_l)), \qquad (l=1,2,3), \text{(A.2)}$$

$$\boldsymbol{\beta}_l|\tau_l,\boldsymbol{\lambda}_l,\sigma_l^2 \sim \mathcal{N}_p(\mathbf{0},\sigma_l^2\tau_l^2\boldsymbol{\Lambda}_l), \qquad (l=1,2,3), \text{(A.3)}$$

$$\sigma,\gamma_l,\lambda_{lj},\tau_l \sim \mathcal{C}^+(0,1), \qquad (l=1,2,3,\ j=1,\cdots,p), \text{(A.4)}$$

$$\alpha_l \sim \pi(\alpha_l)\propto 1, \quad \sigma_l^2 \sim \pi(\sigma_l^2)\propto 1/\sigma_l^2, \qquad (l=1,2,3). \quad \text{(A.5)}$$

On the first stage, for each well i $(i = 1,\cdots,N)$, the vector notations $\mathbf{y}_i$ and $\boldsymbol{\mu}_i(\theta_{1i},\theta_{2i},\theta_{3i})$ represent T_i-dimensional vectors recording the observed log-scaled oil production rate time series data and its corresponding mean part from the likelihood (A.1), respectively, over the production period T_i months:

$$\begin{aligned}\mathbf{y}_i &= (y_{i1},\cdots,y_{it},\cdots,y_{iT_i})^\top\\ \boldsymbol{\mu}_i(\theta_{1i},\theta_{2i},\theta_{3i}) &= (\mu(1;\theta_{1i},\theta_{2i},\theta_{3i}),\ldots,\mu(t;\theta_{1i},\theta_{2i},\theta_{3i})\ldots,\mu(T_i;\theta_{1i},\theta_{2i},\theta_{3i}))^\top,\end{aligned}$$

where $\mu(t;\theta_1,\theta_2,\theta_3)$ is a log-scaled RDC whose options are Weibull model $\mathcal{M}_1$ (3.2), Arps' hyperbolic model $\mathcal{M}_2$ (3.3), SEDM $\mathcal{M}_3$ (3.4), and Duong model $\mathcal{M}_4$ (3.5).

On the second stage, the vector notation $\boldsymbol{\theta}_l = (\theta_{l1}, \cdots, \theta_{lN})^\top$ $(l = 1, 2, 3)$ represents the N-dimensional vector for the curve parameters participated in a chosen log-scaled RDC across the N wells. The matrix notation $\mathbf{X}$ is N-by-p design matrix whose i-th row vector is given by the p completion predictors $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^\top \in \mathbb{R}^p$. Before implementing a sampler, it is recommended that each of the column vectors of the design matrix $\mathbf{X}$ is standardized (Tibshirani, 1996; Armagan et al., 2013): that is, each column vector has been centered, and then columnwisely scaled so that each column vector has mean zero and unit standard deviation. The matrix notation $\mathbf{I}$ stands for an identity matrix. The matrix notation $\mathbf{B}(\rho_l)$ represents N-by-N correlation matrix whose ij-entry is $[\mathbf{B}(\rho_l)]_{i,j} = \exp(-e^{\rho_l}\|\mathbf{s}_i - \mathbf{s}_j\|_2^2)$, $(i, j = 1, \cdots, N)$, induced by the Gaussian correlation function in Eq. 5.3. The matrix $\boldsymbol{\Lambda}_l$ is a p-by-p diagonal matrix formulated by the local-scale parameters of the horseshoe prior, $\boldsymbol{\Lambda}_l = \mathrm{diag}(\lambda_{l1}^2, \cdots, \lambda_{lp}^2)$ $(l = 1, 2, 3)$.

Directed asymmetric graphical model representation (Bishop, 2006) of the model (A.1)–(A.5) is depicted in Fig. 11. Following the grammar of the graphical model, the circled variables indicate stochastic variables, while observed ones are additionally colored in grey. Hyper-parameters is not circled, implying non-stochasticity. The arrows indicate the conditional dependency between the variables.

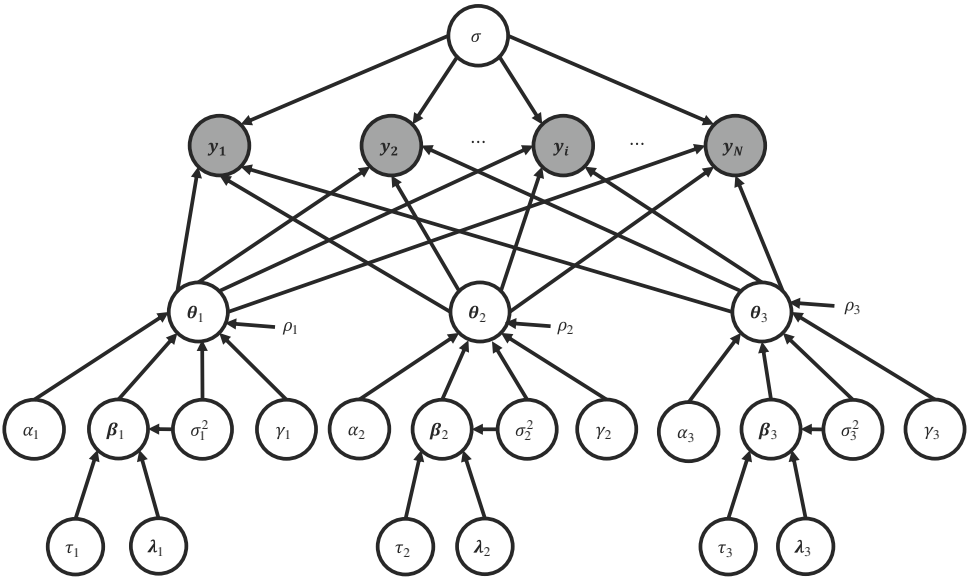

Figure 11: The proposed spatial model as a graphical model

A.2 Gibbs sampling algorithm

Under the formulation of the spatial Bayesian hierarchical model (A.1)–(A.5), the eventual goal of the present section is to sample from its full joint posterior density whose proportional part is given by

$$\left\{\prod_{i=1}^{N} \mathcal{N}_{T_i}(\mathbf{y}_i|\boldsymbol{\mu}_i(\theta_{1i},\theta_{2i},\theta_{3i}),\sigma^2\mathbf{I})\right\}\left\{\prod_{i=1}^{3}\mathcal{N}_N(\boldsymbol{\theta}_l|\mathbf{1}\alpha_l+\mathbf{X}\boldsymbol{\beta}_l,\sigma_l^2\mathbf{I}+\gamma_l^2\mathbf{B}(\rho_l))\right.$$
$$\left.\cdot\mathcal{N}_p(\boldsymbol{\beta}_l|\mathbf{0},\sigma_l^2\tau_l^2\boldsymbol{\Lambda}_l)\cdot\pi(\tau_l)\cdot\pi(\boldsymbol{\lambda}_l)\cdot\pi(\sigma_l^2)\cdot\pi(\gamma_l)\right\}\cdot\pi(\sigma).$$

To sample from the full joint posterior density, we use a Gibbs sampler (Casella and George, 1992; Lee, 2020a) to exploit conditional independence relationships induced by the hierarchy of the model. Before constructing a relevant Gibbs sampler, we shall make use of a parameter expansion about a half-Cauchy random variable (Makalic and Schmidt, 2015). Generally, it is known that a half-Cauchy random variable X can be decomposed as follow: $X\sim\mathcal{C}^+(0,1)\Leftrightarrow X^2|Y\sim\mathcal{IG}(1/2,1/Y)$ and $Y\sim\mathcal{IG}(1/2,1)$. Here, the notation $\mathcal{IG}(a,b)$ stands for the inverse-gamma distribution with the shape parameter a and the scale parameter b, and the notation $\Leftrightarrow$ denotes a distributional equivalence between the random variables on the left and right hand sides.

Half-Cauchy distributed latent variables in the spatial model are listed in Eq. A.4, and they can be equivalently expressed as:

$$\sigma \sim \mathcal{C}^+(0,1)\Leftrightarrow\sigma^2|\phi\sim\mathcal{IG}(1/2,1/\phi)\text{ and }\phi\sim\mathcal{IG}(1/2,1), \quad (A.6)$$
$$\gamma_l \sim \mathcal{C}^+(0,1)\Leftrightarrow\gamma_l^2|\omega_l\sim\mathcal{IG}(1/2,1/\omega_l)\text{ and }\omega_l\sim\mathcal{IG}(1/2,1), \quad (A.7)$$
$$\lambda_{lj} \sim \mathcal{C}^+(0,1)\Leftrightarrow\lambda_{lj}^2|\nu_{lj}\sim\mathcal{IG}(1/2,1/\nu_{lj})\text{ and }\nu_{lj}\sim\mathcal{IG}(1/2,1), \quad (A.8)$$
$$\tau_l \sim \mathcal{C}^+(0,1)\Leftrightarrow\tau_l^2|\xi_l\sim\mathcal{IG}(1/2,1/\xi_l)\text{ and }\xi_l\sim\mathcal{IG}(1/2,1), \quad (A.9)$$

independently for each $l=1,2,3$ and $j=1,\cdots,p$.

We are now in a position to propose a Gibbs sampler. The following algorithm describes a straightforward Gibbs sampler that approximates the full joint posterior distribution obtained from the hierarchical formulation (A.1)–(A.5) where in half-Cauchy distributed latent variables (that is, the latent variables in Eq. A.4) are replaced by Eqs. A.6–A.9.

> ***Step 1.*** Sample ν_{lj}, $l=1,2,3$, $j=1,\cdots,p$, independently from their full conditional distributions
>
> $$\pi(\nu_{lj}|-)\sim\mathcal{IG}(1,1+1/\lambda_{lj}^2).$$

Step 2. Sample ξ_l, $l = 1, 2, 3$, independently from their full conditional distributions

$$\pi(\xi_l|-) \sim \mathcal{IG}(1, 1 + 1/\tau_l^2).$$

Step 3. Sample λ_{lj}^2, $l = 1, 2, 3$, $j = 1, \cdots, p$, independently from their full conditional distributions

$$\pi(\lambda_{lj}^2|-) \sim \mathcal{IG}(1, 1/\nu_{lj} + \beta_{lj}^2/2\sigma_l^2\tau_l^2).$$

Step 4. Sample τ_l^2, $l = 1, 2, 3$, independently from their full conditional distributions

$$\pi(\tau_l^2|-) \sim \mathcal{IG}((p+1)/2, 1/\xi_l + \boldsymbol{\beta}_l^\top \boldsymbol{\Lambda}_l^{-1}\boldsymbol{\beta}_l/2\sigma_l^2),$$

where $\boldsymbol{\Lambda}_l = \text{diag}(\lambda_{l1}^2, \cdots, \lambda_{lp}^2) \in \mathbb{R}^{p\times p}$.

Step 5. Sample σ_l^2, $l = 1, 2, 3$, independently from their full conditional distributions. Proportional parts of the distributions are given by

$$\begin{aligned}\pi(\sigma_l^2|-) \quad &\propto \quad (\sigma_l^2)^{-\frac{N+p}{2}-1} \cdot \exp\left\{-\frac{1}{\sigma_l^2}\left(\frac{\widetilde{\boldsymbol{\theta}}_l{}^\top[\mathbf{I} + (\gamma_l^2/\sigma_l^2)\mathbf{B}(\rho_l)]^{-1}\widetilde{\boldsymbol{\theta}}_l + \boldsymbol{\beta}_l^\top \boldsymbol{\Lambda}_{*l}^{-1}\boldsymbol{\beta}_l}{2}\right)\right\} \\ &\cdot \quad |\mathbf{I} + (\gamma_l^2/\sigma_l^2)\mathbf{B}(\rho_l)|^{-\frac{1}{2}} \end{aligned} \tag{A.10}$$

where $\widetilde{\boldsymbol{\theta}}_l = \boldsymbol{\theta}_l - \mathbf{1}\alpha_l - \mathbf{X}\boldsymbol{\beta}_l \in \mathbb{R}^N$. The notation $|\mathbf{A}|$ indicates the determinant of the matrix $\mathbf{A}$. The density $\pi(\sigma_l^2|-)$ is not a known distribution, and we use a Metropolis-Hastings algorithm (Chib and Greenberg, 1995) within this Gibbs sampler to sample from the density. A proposal density for the algorithm is suggested in the next subsection.

Step 6. Sample $\boldsymbol{\beta}_l$, $l = 1, 2, 3$, independently from their full conditional distributions

$$\pi(\boldsymbol{\beta}_l|-) \sim \mathcal{N}_p(\boldsymbol{\Sigma}_{\boldsymbol{\beta}_l}\mathbf{X}^\top[\sigma_l^2\mathbf{I} + \gamma_l^2\mathbf{B}(\rho_1)]^{-1}(\boldsymbol{\theta}_l - \mathbf{1}\alpha_l), \boldsymbol{\Sigma}_{\boldsymbol{\beta}_l}),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_l} = \{\mathbf{X}^\top[\sigma_l^2\mathbf{I} + \gamma_l^2\mathbf{B}(\rho_1)]^{-1}\mathbf{X} + (1/\sigma_l^2)\boldsymbol{\Lambda}_{*l}^{-1}\}^{-1} \in \mathbb{R}^{p\times p}$ with $[\mathbf{B}(\rho_l)]_{i,j} = \exp(-e^{\rho_l}\|\mathbf{s}_i - \mathbf{s}_j\|_2^2)$ $(i, j = 1, \cdots, N)$, and $\boldsymbol{\Lambda}_{*l} = \tau^2\boldsymbol{\Lambda}_l \in \mathbb{R}^{p\times p}$. The notation $\|\cdot\|_2$ indicates the l_2-norm.

Step 7. Sample α_l, $l = 1, 2, 3$, independently from their full conditional distributions

$$\pi(\alpha_l|-) \sim \mathcal{N}_1\left(\frac{\mathbf{1}^\top[\sigma_l^2\mathbf{I}+\gamma_l^2\mathbf{B}(\rho_l)]^{-1}(\boldsymbol{\theta}_l-\mathbf{X}\boldsymbol{\beta}_l)}{\mathbf{1}^\top[\sigma_l^2\mathbf{I}+\gamma_l^2\mathbf{B}(\rho_l)]^{-1}\mathbf{1}}, \frac{1}{\mathbf{1}^\top[\sigma_l^2\mathbf{I}+\gamma_l^2\mathbf{B}(\rho_l)]^{-1}\mathbf{1}}\right),$$

where the notation $\mathbf{1}$ stands for a vector whose entries are ones.

Step 8. Sample $\boldsymbol{\theta}_1$ from its full conditional distribution

$$\pi(\boldsymbol{\theta}_1|-) \sim \mathcal{N}_N(\boldsymbol{\Sigma}_{\boldsymbol{\theta}_1}\{(1/\sigma^2)\mathbf{r} + [\sigma_l^2\mathbf{I} + \gamma_l^2\mathbf{B}(\rho_1)]^{-1}(\mathbf{1}\alpha_1 + \mathbf{X}\boldsymbol{\beta})\}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}_1}),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}_1} = \{(1/\sigma^2)\mathbf{T} + [\sigma_l^2\mathbf{I} + \gamma_l^2\mathbf{B}(\rho_1)]^{-1}\}^{-1} \in \mathbb{R}^{N\times N}$.
Here, the $\mathbf{r}$ is a N-dimensional vector-valued function of $\{(\theta_{2i}, \theta_{3i})\}_{i=1}^N$ across the N wells, given by

$$\mathbf{r} = (\mathbf{1}^\top\{\mathbf{y}_1 - \mathbf{h}(\theta_{21}, \theta_{31})\}, \ldots, \mathbf{1}^\top\{\mathbf{y}_N - \mathbf{h}(\theta_{2N}, \theta_{3N})\})^\top,$$

whereas the $\mathbf{h}(\theta_{2i}, \theta_{3i})$ is a T_i-dimensional vector-valued function of $(\theta_{2i}, \theta_{3i})$ for the i-th well $(i = 1, \cdots, N)$

$$\begin{aligned}\mathbf{h}(\theta_{2i}, \theta_{3i}) &= \boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i}) - \mathbf{1}\theta_{1i} \\ &= (\mu(1; \theta_{1i}, \theta_{2i}, \theta_{3i}) - \theta_{1i}, \ldots, \mu(T_i; \theta_{1i}, \theta_{2i}, \theta_{3i}) - \theta_{1i})^\top.\end{aligned}$$

Step 9. Sample θ_{2i} and θ_{3i}, $i = 1, \cdots, N$, independently from their full conditional distributions. Proportional parts of the distributions are

$$\begin{aligned}\pi(\theta_{2i}|-) &\propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}_i - \boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i})\|_2^2 - \frac{1}{2(\sigma_2^2+\gamma_2^2)}(\theta_{2i} - \alpha_2 - \mathbf{x}_i^\top\boldsymbol{\beta}_2)^2\right), \\ \pi(\theta_{3i}|-) &\propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}_i - \boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i})\|_2^2 - \frac{1}{2(\sigma_3^2+\gamma_3^2)}(\theta_{3i} - \alpha_3 - \mathbf{x}_i^\top\boldsymbol{\beta}_3)^2\right).\end{aligned}$$

Note that the above densities are not be expressed as known distributions because two parameters, θ_{2i} and θ_{3i}, participate to the function $\boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i})$ in a nonlinear way. We use the Metropolis algorithm (Andrieu et al., 2003) with Gaussian proposal densities within this Gibbs sampler algorithm.

Step 10. Sample ϕ from its full conditional distribution

$$\pi(\phi|-) \sim \mathcal{IG}(1, 1 + 1/\sigma^2).$$

Step 11. Sample σ^2 from its full conditional distribution

$$\pi(\sigma^2|-) \sim \mathcal{IG}\left(\frac{\text{trace}(\mathbf{T})+1}{2}, \frac{1}{\phi} + \frac{1}{2}\sum_{i=1}^{N} \|\mathbf{y}_i - \boldsymbol{\mu}_i(\theta_{1i}, \theta_{2i}, \theta_{3i})\|_2^2\right),$$

where $\mathbf{T} = \text{diag}(T_1, \cdots, T_N) \in \mathbb{R}^{N\times N}$.

Step 12. Sample γ_l^2, $l = 1, 2, 3$, independently from their full conditional distributions. Proportional parts of the distributions are given by

$$\begin{aligned}\pi(\gamma_l^2|-) \quad &\propto \quad (\gamma_l^2)^{-\frac{N+1}{2}-1} \cdot \exp\left\{-\frac{1}{\gamma_l^2}\left(\frac{\widetilde{\boldsymbol{\theta}}_l^{\top}[(\sigma_l^2/\gamma_l^2)\mathbf{I} + \mathbf{B}(\rho_l)]^{-1}\widetilde{\boldsymbol{\theta}}_l}{2} + \frac{1}{\omega_l}\right)\right\} \\ &\cdot \quad |(\sigma_l^2/\gamma_l^2)\mathbf{I} + \mathbf{B}(\rho_l)|^{-\frac{1}{2}}.\end{aligned}$$

The density $\pi(\gamma_l^2|-)$ is not a known distribution and we use a Metropolis-Hastings algorithm where its proposal density is suggested in the next subsection.

Step 13. Sample ω_l, $l = 1, 2, 3$, independently from their full conditional distributions

$$\pi(\omega_l|-) \sim \mathcal{IG}(1, 1 + 1/\gamma_l^2).$$

R code for the Gibbs sampler is available on https://github.com/yain22/SWM.

A.3 Proposal densities for the Step 5 and 12

In the present subsection, we suggest a proposal density that can be used to implement the Metropolis-Hastings algorithm for the ***Step 5***. For a fixed l ($l = 1, 2, 3$), suppose that the signal-to-noise ratio, $c_l = \gamma_l^2/\sigma_l^2$, is constant in the ***Step 5*** (A.10): then, the full conditional posterior of σ_l^2 can be expressed in a closed form distribution

$$\pi(\sigma_l^2|-) \sim \mathcal{IG}\left(\frac{N+p}{2}, \frac{\widetilde{\boldsymbol{\theta}}_l^{\top}[\mathbf{I} + c_l\mathbf{B}(\rho_l)]^{-1}\widetilde{\boldsymbol{\theta}}_l + \boldsymbol{\beta}_l^{\top}\boldsymbol{\Lambda}_{*l}^{-1}\boldsymbol{\beta}_l}{2}\right),$$

because the determinant term in Eq. A.10 is constant with respect to σ_l^2. In the Metropolis-Hastings algorithm, we use this inverse-gamma distribution

as a proposal density. That being said, given previous sample $(\sigma_l^2)^{(s)}$, a candidate for the next sample $(\sigma_l^2)^{(s+1)}$ is drawn from

$$J(\sigma_l^2|(\sigma_l^2)^{(s)}) = \mathcal{IG}\left(\sigma_l^2 \middle| \frac{N+p}{2}, \frac{\widetilde{\boldsymbol{\theta}}_l^\top [\mathbf{I} + (c_l)^{(s)}\mathbf{B}(\rho_l)]^{-1}\widetilde{\boldsymbol{\theta}}_l + \boldsymbol{\beta}_l^\top \boldsymbol{\Lambda}_{*l}^{-1}\boldsymbol{\beta}_l}{2}\right), \quad c_l^{(s)} = \gamma_l^2/(\sigma_l^2)^{(s)},$$

and then evaluated whether to accept it or not based on the Metropolis-Hastings rejection criteria: see page 184 of Hoff (2009). The suggested proposal density is a canonical byproduct because it becomes the full conditional posterior distribution of σ_l^2 when the spatial random effects are removed (that is, $\gamma_l^2 = 0$) from the regression (5.2).

As similar to how we derived a proposal density for the ***Step 5***, we obtain a proposal density for the Metropolis-Hastings algorithm for the ***Step 12*** to sample from the full conditional distribution of γ_l^2 $(l = 1, 2, 3)$:

$$J(\gamma_l^2|(\gamma_l^2)^{(s)}) = \mathcal{IG}\left(\gamma_l^2 \middle| \frac{N+1}{2}, \frac{\widetilde{\boldsymbol{\theta}}_l^\top [(1/c_l^{(s)})\mathbf{I} + \mathbf{B}(\rho_l)]^{-1}\widetilde{\boldsymbol{\theta}}_l}{2} + \frac{1}{\omega_l}\right), \quad c_l^{(s)} = (\gamma_l^2)^{(s)}/\sigma_l^2.$$

Se Yoon Lee
Bani K. Mallick
Department of Statistics, Texas A &
M University, 3143 TAMU, College
Station, TX, 77843, USA
E-mail: seyoonlee@stat.tamu.edu
bmallick@stat.tamu.edu

Paper received: 20 February 2020; accepted 10 December 2020.