SPATIO-TEMPORAL ESTIMATES OF LONG-TERM OIL DRILLING LOCATIONS

Prepared by:

EunSu Lee, Ph.D.
Nimish Dharmadhikari, Ph.D. Candidate
Alan Dybing, Ph.D.

Upper Great Plains Transportation Institute
North Dakota State University, Fargo

UGPTI Department Publication No. 275

October 2014

Acknowledgments

The authors would like to acknowledge the sponsorship of this project by North Dakota Department of Transportation. The results and opinions in this paper are those of the authors and do not necessarily reflect the policy of the sponsors.

Disclaimer

The contents presented in this report are the sole responsibility of the Upper Great Plains Transportation Institute and the authors.

ABSTRACT

Estimating drilling locations plays an important role in forecasting truck trips derived by hydraulic fracturing oil development for long-range transportation planning. Predicting drilling locations among more than 7,000 oil drilling lands shows a random pattern with uncertainty. Twenty-year multi-period forecasting architecture is proposed in this study using maximum likelihood estimation to fit in the logistic regression. Probability of drilling on each leased space for drilling was predicted in order to forecast truck movements with respect to frequency and paths for a 20-year period. The maximum likelihood estimates of the parameters of a logistic model provide future locations for drilling considering well aging, the number of wells on a leased land space, closeness to the current wells, and oil density in an oil development zone. While, in the short term, the drilling locations were concentrated in the middle of the space, the probability of the drilling was marginally distributed throughout the region in the long run.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Drilling shale oil wells generates a large amount of truck traffic. Heavy loads required by drilling activities damage local roads, resulting in the need for several hundred million dollars for repair and maintenance. To support logistical activities for efficient energy development, a proactive approach is required for allocating investments for paving road and timely maintenance. Forecasting load impact on a road network is essential for estimating pavement and repair costs to support energy logistics [1].

Long-range transportation planning requires reasonably accurate information for transportation agencies and legislators to make appropriate decisions and justify strategic long-term budget allocations. Such planning requires an understanding of the equipment movement and traffic sources required for large-scale and comprehensive efforts over widespread oil production zones.

This paper focuses on the prediction of oil drilling locations in order to predict rig movements between current drilling locations and the next drill locations for the next 20-year planning period where producing wells need to be fractured in order to increase oil production. Oil-related traffic models involve historical and current oil wells across 17 oil counties in North Dakota.

In 2010, more than 100 million barrels of crude oil was produced in North Dakota, which was about a 42% increase from 2009. This level of production represents 309,679 barrels oil per day, making the state the fourth largest oil producing state in the United States. By April 2012, North Dakota had become the second largest oil producing state in the United States, surpassing Alaska, and producing almost one million barrels per day. However, there is only one crude oil refinery in the state, located near Mandan, with a capacity of 58,000 barrels per day. The rest of the state's oil production is transported through five major pipe transloading facilities and 18 BNSF and Canadian National railroads crude-by-rail facilities to oil refineries in Texas, Oklahoma, and Louisiana. In addition to crude oil, 13 natural gas processing plants handle 80 billion cubic feet of the total 114 bcf of natural gas produced as by-products of oil drilling in 2011. Drilling rig count is a prime barometer for measuring oil and gas activity. The count averaged 200 rigs per day in North Dakota in 2011, breaking the record of 126 rigs per day set in 2010 and 182 in 2011.

Thus, estimating drilling locations plays an important role in estimating truck trips for long-term planning. Oil production-related traffic is generated by directional shipments: inbound shipments of sand, fresh water, pipes, gravel, and supplies and outbound shipments of salt-water as by-products and crude oil. Truck traffic generated between origins and destinations is based on commodities that must be shipped. Horizontal drilling using hydraulic fracturing generates about 2,024 inbound and outbound truck movements [2]. Approximately half of the trips are loaded. Based upon the projected number of rigs and drilling activities, the number of oil wells can be estimated. Drilling activities vary by the duration of the drilling process as well as by the locations in each county or smaller unit. In general, it takes six weeks to two months to drill one oil well [1].

Predicting drilling locations among more than 7,000 oil lease units shows a random pattern with uncertainty. A spacing unit is the land which covers an area at least as large as can be "efficiently and economically be drained by one well" [3]. The spacing units within an oil field have uniform sizes and shapes in general; nonetheless, the sizes and shapes can differ within in the same oil field. Most of the spacing units for horizontal wells are either 640 acres or 1,280 acres [4]. Instead of drilling one well in a new spacing unit, some wells are drilled horizontally, crossing into neighboring spacing units. The challenges of the choice-based prediction stem from the unknown history of the rigs' movements and their pattern and randomness under uncertainty of oil price and rig activity. Because of the severity of traffic and the uncertainty of the drilling and production locations, estimating well locations to estimate traffic generation plays an important role in long-range transportation planning (LRTP).

The paper regards this prediction problem of predicting oil spacing units as choice-based prediction using maximum likelihood method. Twenty-year multi-period forecasting architecture is proposed for use in this study. State transportation agencies and legislators require long-term traffic estimations and cost analysis in order to make appropriate decisions and to justify strategic, long-term budget allocations. Thus, this study estimates drilling probability on each spacing unit using the maximum likelihood estimation in order to forecast truck movements with respect to frequency and paths for a 20-year period.

2. LITERATURE REVIEW

Geographic information systems (GIS) are widely used for spatial pattern analysis to predict unpredictable movements of humans, animals, diseases, and other biological organisms. The analyses integrate several data sources, including geographical locations, tracks of historical movements, and behavioral patterns [5]. Similarly, pattern recognition analysis is commonly used to recognize the patterns referred to as a group of statistical techniques [6]. These techniques can be found in disease, crime analysis, and biology studies.

To estimate randomly distributed locations, several spatial estimations were introduced in criminal analysis and disease and epidemic models to handle dynamic movement under uncertainty. Volz and Meyers [7] present a new model, “neighbor exchange” (NE). This model is a dynamic addition to the static contact network model. “The model assumes that at any given time, an individual will be in contact with an individual-specific number of neighbors with whom disease transmission is possible. Each contact is temporary, lasting a variable amount of time before coming to end, at which point the neighbor is replaced by a different individual.” This model expresses the dynamic characteristics of the epidemic propagation thus capturing the susceptible–infected–recovered (SIR) dynamics in a population.

Bertuzzo et al. [8] conducted research on the propagation of cholera through the hydrological network connections. They test different lattice models as networks to calculate the front speed of an epidemic. They suggest their approach can be better for the solution of the heterogeneous networks. Danon et al. [9] used a degree of distribution to present a risk assessment model of infection in a social network. They used surveys to generate better results from their model. They provide survey results and their implications about different models, which use different assumptions to model disease propagation in a social network.

Brown, Dalton, and Hoyle [10] used spatial choice analysis to develop an empirical prediction model for future suicide bombings. They study two models: the spatial preference model and the logistics regression model, and their fusion with each other. They compare these models with the naïve model, which shows significant improvement in the forecasting.

Bernasco [11] analyzed criminal locations at a detailed spatial resolution. The author proposes that the offenders make their decisions on the detailed level, so the study should use detailed resolution, too, rather than on a larger level like census tracts. Discrete choice and spatial choice models are studied taking into account smaller resolution. It is noted that small spatial units depend strongly on their environment and models that take spatial interdependence into account are needed.

Another way of spatial modeling is spatial choice modeling. Hunt, Boots, and Kanaroglou [12] compared spatial choice modeling. They provide general choice models, and then the study attempts to add a spatial component in choice models. Later, they present newer choice models which can be useful for geographers: the generalized extreme value model, the open-form choice model, and the choice set model. They propose that these newer models are evolved enough to take into account spatial component while keeping random utility. Bhat and Sener [13] use a spatial logit model to accommodate spatial correlation across observation units. They propose a copula-based, closed-form binary logit choice model. Their method highlights the power of closed-form techniques for accommodating spatial effects.

Zhu and Timmermans [14] extended the heuristic rules, conjunctive, disjunctive and lexicographic rule, and introduce heterogeneity. They suggest that instead of conventional discrete choice models using principles of bounded rationality, it is better to model pedestrian behavior as it is a complex environmental process.

Alamá-Sabater, Artal-Tur, and Navarro-Azorín [15] introduced the neighborhood effect in spatial conditional logit framework. They investigated the drivers of the location choices of industrial firms in cases of inter-territorial spillovers. They confirmed the influence of the spatial factors on the decision analysis with other major factors. They developed their model based on the standard random utility maximization (RUM) framework.

For pattern recognition, discriminant analysis (DA) would determine where each well belongs with the classification of groups and variables [6]. The DA finds the best linear combinations of the geographical and behavioral variables for the oil wells. Forecasting production from oil wells is different from the traditional prediction of oil production and population growth in terms of geographical dynamics and flow. For example, oil and gas production rate and growth per region and zone can be estimated based on the production plan and oil richness; however, the estimation is appropriate for fixed regions and zones, which are known. Similarly, the number of oil wells can be estimated by using linear projection in response to trends. Mitra et al. [1] estimated the future locations of drilling rigs from lease data. They assumed that the drilling activity would begin at the final year of the lease right before the lease expired. They also assigned three to five additional wells to the mineral land lease. Due to lack of information available for the estimation, the study uses distances fron the nearest oil wells.

In summary, the literature review indicates that drilling locations can be estimated by using a variety of methods, including a spatial choice model with the combination of logit regression. The required spatial information can be collected, and the outputs for the decision-making process can be visualized by using GIS. To our knowledge, based on extensive literature review, no application has been published for predicting horizontal drilling locations for a long-term period.

3. METHODOLOGY

3.1 Datasets

An oil spacing unit represents the public land lease for drilling and producing oil available from the North Dakota Oil and Gas Division. The division provides all oil activity related GIS files through ArcIMS Viewer [16]. Historical oil wells and existing rig locations are available from the information server; however, the historical drilling activities have not been published to the public, which make clustering analysis difficult.

The previous active oil wells can be tracked by spud year in the given oil wells from the North Dakota Department of Mineral Resources [16]. Since 2005, more than 20 rigs actively drilled horizontal wells, which increased from the previous year by 78.6%. At the same time, the increased number of wells produced more than 50 million barrels per year and steadily increased. Thus, the years from 2005 to 2011 were considered in the historical period to estimate the oil production and historical rig locations in time series. The prediction model proposed herein does not account for the oil wells located out of the spacing units. This study uses the forecasting data from the North Dakota Department of Labor for the number of oil wells for a 20-year time window through oil zones.

3.2 Modules of the Model

The drilling process takes six weeks to two months, which means that rig equipment can drill up to 10 wells in a year. A search for the next drilling sites was based on the probability of drilling for each unit (Figure 1). The potential drilling locations were sequenced by the probability of drilling a spacing unit, and then, most likely, wells were selected by the numbers in accordance with oil zones and oil counties. The selected spacing units are drilled, thereby generating truck trips to ship sand, gravel, fresh water, pipes, etc. Different commodities need different types of trucks. When the drilling phase is finished, the oil wells are equipped with well overhead to produce crude oil, thereby generating truck trips to transport crude oil and salt water and for maintenance [1]. The trip generation process is explained in Figure 3.1. The searching process is repeated after updating the drilling information of the spacing units over the 20-year time window.

The maximum likelihood model estimates the parameters of the temporal and spatial factors for placing drilling locations. External factors, such as oil in the market and operator's characteristics, are also critical for the study; however, the study used the annual forecasts of wells and production per county obtained from the Oil and Gas Division of North Dakota. This study assumes that the agency forecasts the number of oil wells considering the oil price trend and operators' activities throughout the state. Thus, this study relies on their estimates and just focuses on predicting vertical wellbores on a spacing unit. The maximum likelihood estimation explains the choice of drilling locations because the number of drilling sites in a county is limited based on the drilling forecast data for each county. All the spacing units' predicted values and limits lie between 0 and 1.

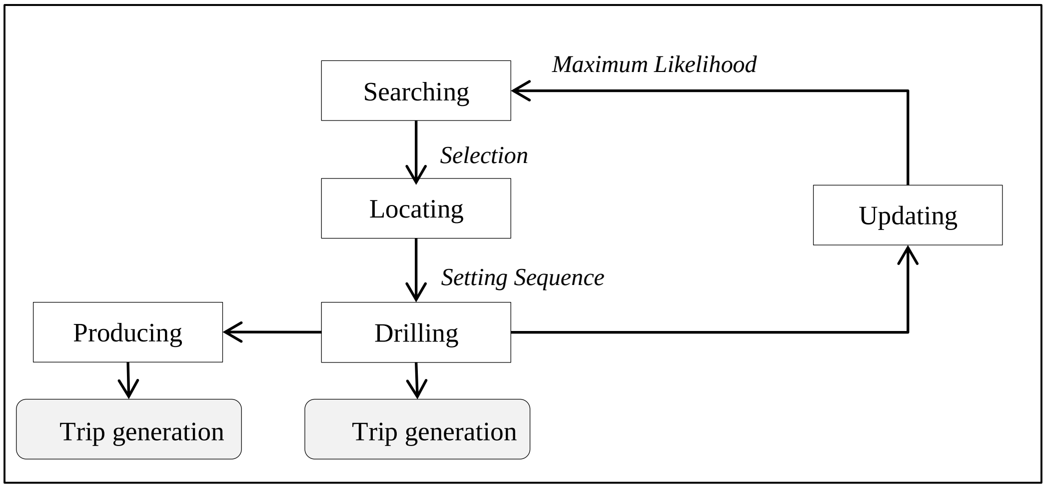

Figure 3.1 Five sub-systems for forecasting of oil spacing units for trip generation (truck trips).

Maximum likelihood estimation is used for selecting drilling location to fit logistic regression model in this study. The response probability π, regression coefficient *β,* and explanatory variable *x* are elements of the logistic regression model in the form of

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1\chi_1 + \cdots + \beta_n\chi_n \qquad (1)$$

The odds of the event (*y* = drilling) occurring follows $odd = \frac{\pi}{1-\pi} = \frac{P(y = drilling)}{P(y = nodrilling)}$

where y is a binary response [17]. Thus, the prediction equation for the probability of drilling, *E*(y), in a spacing unit *i* is

$$E(y_i) = \frac{e_i^{\pi}}{1+e_i^{\pi}} \qquad (2)$$

, that consists of the maximum likelihood estimates of $\beta_0, \beta_1, \ldots, \beta_{16}$ from the model:

$$\begin{aligned}\pi_i &= \beta_0 + \beta_1 YEAR5_i + \beta_2 YEAR4_i + \beta_3 YEAR3_i + \beta_4 YEAR2_i + \beta_5 YEAR1_i + \beta_6 AQUA_i + \\ &\beta_7 \ln WELLFRQ5_i + \beta_8 \ln WELLFRQ10_i + \beta_9 \ln WELLFRQ15_i + \beta_{10} \ln AREAMILE_i + \\ &\beta_{11} \ln PIPENEAR + +\beta_{12} \ln WDEPNEAR + \beta_{13} \ln NEARWELL_i + \beta_{14} \ln NEARTRUST_i + \\ &\beta_{15} \ln ZONEDENS_i + \beta_{16} \ln ASSESUNIT_i + \varepsilon_i\end{aligned} \qquad (3)$$

In the model, the β_0 is the intercept and the *β*s, all s={1,…,16}, present the estimated coefficients of each independent variable for the maximum likelihood ratio. *WELLFRQ5* presents the well density by measuring the number of wells producing within five miles and *WELLFRQ10* and *WELLFRQ15* within 10 miles and 15 miles, respectively. The oil wells along a river and lake or in the middle of the water locations are presented by the binary variable of *AQUA*. The next sections explain the other variables in detail.

3.3 Dynamic Estimation of Temporal-Spatial Factors

This study considers a variety of temporal and spatial factors in order to estimate likely locations for drilling oil wells.

3.3.1 U.S. Geological Survey's Oil Assessment Information

This study spatially joins the U.S. Geological Survey (USGS)'s Williston Basin oil assessment data to the spacing units in order to measure the attractiveness of each spacing unit for drilling. The assessments includes Elm Coulee-Billings Nose (z = 1), Missouri-Little Knife (z = 2), Central Basin-Poplar Dome (z = 3), Northwest Expulsion Threshold (z = 4), and Eastern Expulsion Threshold (z = 5) assessment units (AUs) in the Bakken formation [17] [18]. Each factor is converted into a value between 0 and 1. The undiscovered oil is a technically recoverable resource, so it was matched to each assessment unit out of 3.64 billion barrels from the Bakken formation.

$$ASSESSUNIT_i = Barrel_z \times \frac{Barrel_{i \in z}}{\sum_{i=1}^{n} Barrel_{i \in z}}, \forall z = \{1,2,3,4,5\} \tag{4}$$

where i is a spacing unit, and z represents an assessment unit (AU) in the Bakken formation. The number of barrels is estimated based on the area of the spacing unit in each zone.

3.3.2 Distance from Pipe Loading Sites

We measure the Euclidian distance from an oil spacing unit to the closest pipe transloading facility (5). By drilling an oil well near the pipe transloading facility, oil producers can reduce transportation costs and required resources and also lessen the impact on public roads. In that sense, drillers would prefer to lease the public trust mineral land and private oil fields near the pipe transloading sites in cooperation with other factors. The distance was estimated using an as-the-crow-flies route between a site proposed (x_i, y_i) and the closest pipe transloading facility (x_q, y_q).

$$PIPENEAR_i = \sqrt{(x_i - x_q)^2 + (y_i - y_q)^2} \tag{5}$$

3.3.3 Distance from Fresh Water Depots

The distance between a spacing unit and the closest water depot is measured by Euclidean straight line between two coordinates for a rig (x_i, y_i) and for the nearest source of water (x_w,y_w) (6). Drilling activities will take place as close as possible to existing water depots to minimize logistics cost and increase productivity. Approximately 400 truck movements are needed to ship fresh water in order to drill a horizontal well. In consideration of this, the oil spacing units are more likely to be located near the water depots.

$$WDEPNEAR_i = \sqrt{(x_i - x_w)^2 + (y_i - y_w)^2} \tag{6}$$

3.3.4 Area Density

To avoid the risk of having a dry well, a rig would be placed in a productive area, which is proven. Oil density with a threshold distance can be used to estimate the maximum likelihood of drilling. Each spacing unit's $ZONEDENS_i$ counts the number of wells in the oil drilling history within $d_0 = 5$, 10, or 15 miles, respectively, because the study uses homogeneous buffer sizes for each spacing unit (7).

$$ZONEDENS_i = F^{j}_{d_0=\{5,10,15\}} \tag{7}$$

3.3.5 Well Age

The variable of a well's age verifies if drilling occurred in the past five years from the baseline scenario (i.e., 1, 2, 3, 4, or 5). For example, the estimation of the drilling location on a spacing unit in 2010 sees the drilling history of 2009 for $k = 1$ (8). In other words, this model estimates dynamic causal effects. In this distributed lag model, the maximum of the k value is five years in the state. The independent variable of aging can be weighted by special coefficient of β in the model.

$$YEAR(K)_i = \begin{cases} 1, \text{if a spacing unit } i \text{ has a horizontal well drilled in the year} \textit{(yyyy-k)} \\ 0, \text{otherwise} \end{cases} \tag{8}$$

3.3.6 Distance from the Closest Well

By placing a rig at the closest spacing unit, the workover rigs and all other resources can be used with shorter movements. Furthermore, the new drilling can minimize the risk of having a dry oil well. The distance is measured by Euclidean straight line between two coordinates (x_i, y_i) and ($x_{neighbor}$, $y_{neighbor}$).

$$NEARWELL_i = \sqrt{(x_i - x_{neighbor})^2 + (y_i - y_{neighbor})^2} \tag{9}$$

However, some newly deployed equipment will not start from the current rig location points. Instead, they will find neighboring spacing units, which have potential. For example, the spacing unit (2, b) is drilled (see Figure 3.2-a), and then the rig chooses a cell (2, b) and starts to drill the zone (see Figure 3.2-b). Cells (2, a) and (2, b) are now excluded for the next drilling locations (c), so cell (1, b) was chosen (see Figure 3.2-c) following by 1c (see Figure 3.2-d).

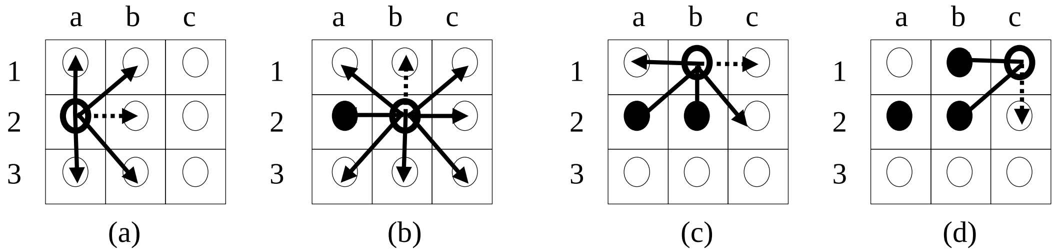

Figure 3.2 Searching neighbor units for drilling location. Among the alternate movement (solid lines), the dashed line is selected for the next drilling location in accordance with the probability of drilling.

3.4 Zonal Plan as a Constraint

Additional new wells based on oil zones (i.e., seven oil zones in 17 counties) were used to predict the total oil wells in Phase I and Phase II for planning. This study relies on the information given for predicting the oil activities for a 20-year planning horizon. In addition to predicting drilling behavior, the plan acts as a restriction or guideline for future drilling activities in the regions.

3.5 Selection Process in GIS

The study clipped the oil activity-related layers for the 17 oil counties (Figure 3) and then exported drilled oil and gas wells for the last five years based on the spud date (2006-2011) from the oil wells dataset. The oil spacing units are grouped into 2010 and 2011 by start date.

In addition to public land leases, this study includes oil exploration and leasing activities on privately owned land. Since data are not available as to the lease negotiation and timing on private lands, it's assumed that leasing activities on public and private lands within a geographic area occur simultaneously. That is, once a company acquires a public land lease, it will negotiate leases with private landowners in contiguous areas. This model assumes that if there is a public land lease acquisition and actual drilling activity occurs that private land within three miles of the public lease will also have similar drilling characteristics, as long as it has not already been drilled.

3.6 Spatial and Temporal Autocorrelation

A group of wells from 2005 to 2012 was used to check out the spatial and temporal autocorrelation. The research tested the autocorrelation using Moran's I index. Inverse distance was used for the conceptualization of spatial relationships using Euclidean distance of 30Km. Moran's index was 0.509, which indicates fairly distributed and clustered at the same time. In general, the positive Moran's I index express the clustered pattern of the oil wells throughout the periods with variance of 0.000322 and expected index of -0.000273. Across the oil fields in the Bakken formation, the spatial clustering is frequently shown because the oil drilling activities are concentrated within the Williston thermal maturity in that productive wells are coincident with the area [19]. In addition, gathering modes, such as rail terminals and pipelines, are one of the critical factors of production. During this period, the wells were randomly distributed by p-value of less than 0.0000001 and z-score of 28.376223.

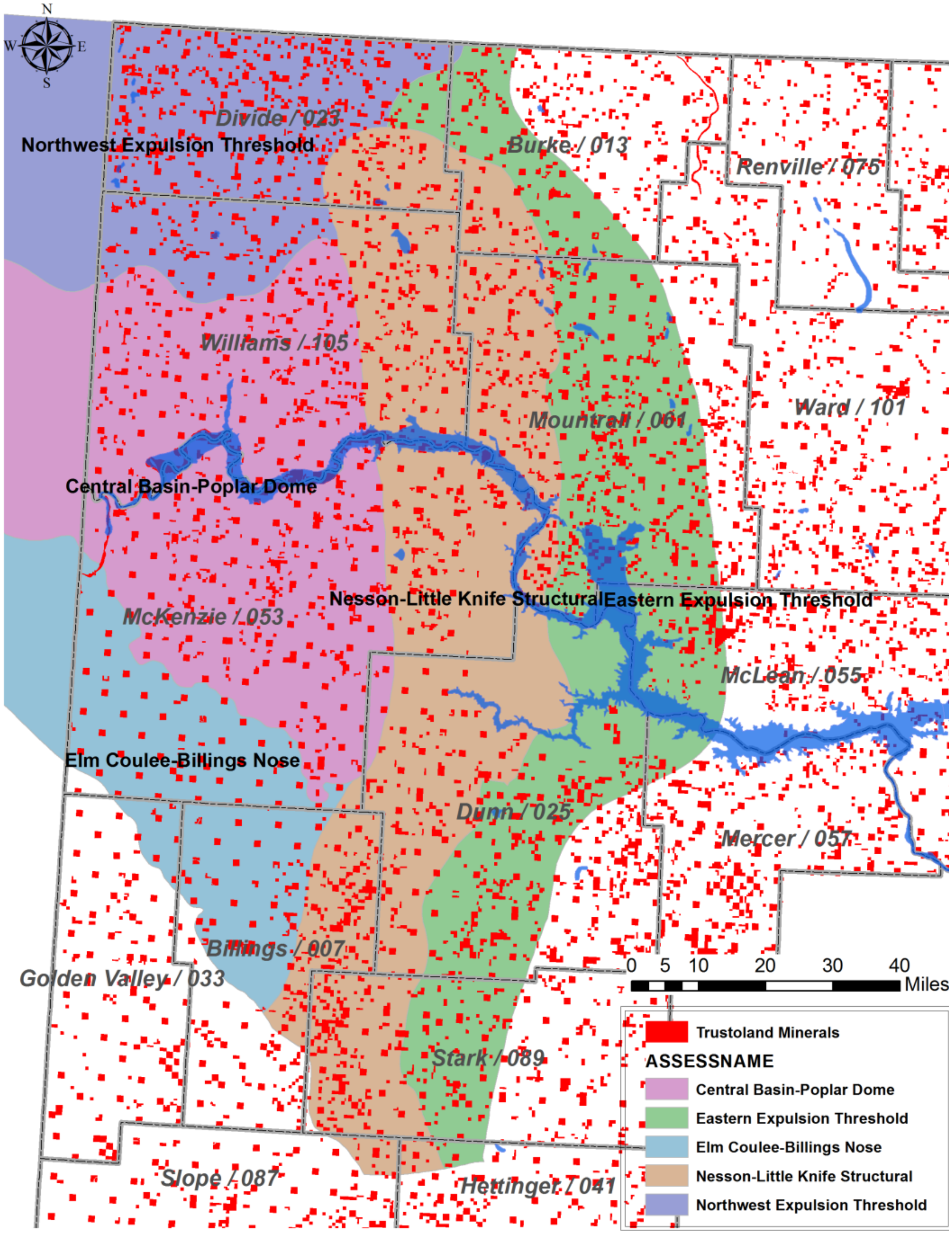

Figure 3.3 Regional public trust land mineral and U.S. Geologic Survey oil assessments.

4. EMPIRICAL RESULTS

4.1 Model Validation

The output of the model estimate is displayed in the map in Figure 4.1. We estimated the 589 drilling locations in 2011. The estimated drilling locations are compared with the 693 oil wells spudded in 2011 and the 200 sample rigs during 2011. The yearly drilling locations are not traced, so the exact locations of drilling are not known to the public. The distribution of the spacing units and oil wells are reasonably distributed through the overall oil counties (Figure 4). The visualized distribution of the estimated drilling sites and rig locations indicates the differences are shown in the counties of Golden Valley, Stark, and Billings. The south of Williams County also shows the gaps from the estimation.

The average distance from drilling locations to the sample rig locations was 4,137.79 meters. At the 95% confident level, the sample average of distance between an estimated drilling site and an existing rig location will be within between 3673.56 meters and 4602.03 meters. The standard error was 236 and standard deviation 5,736.54 meters. Among 589 drilling locations, about 5% of them are farther than 10 kilometers away from an existing rig location, and about 10% of them are off more than 8115.79 meters. Because the size of a spacing unit is one mile by two miles, the average distance is acceptable for the model by 2.57 miles (Table 4.1).

Table 4.1 Summary statistics of the distance between observed and estimated drilling sites.

Measure	Values
Mean	4137.79 meters
Mean (Lower 95%)	3673.56 meters
Mean (Upper 95%)	4602.03 meters
Mode	0 meter
Standard Deviation	5736.54 meters
Minimum	0 meter
Maximum	85497.11 meters
Count	589 spacing units
90th Percentile	8115.79 meters
95th Percentile	10398.13 meters
Coefficient of Variation	138.64 %

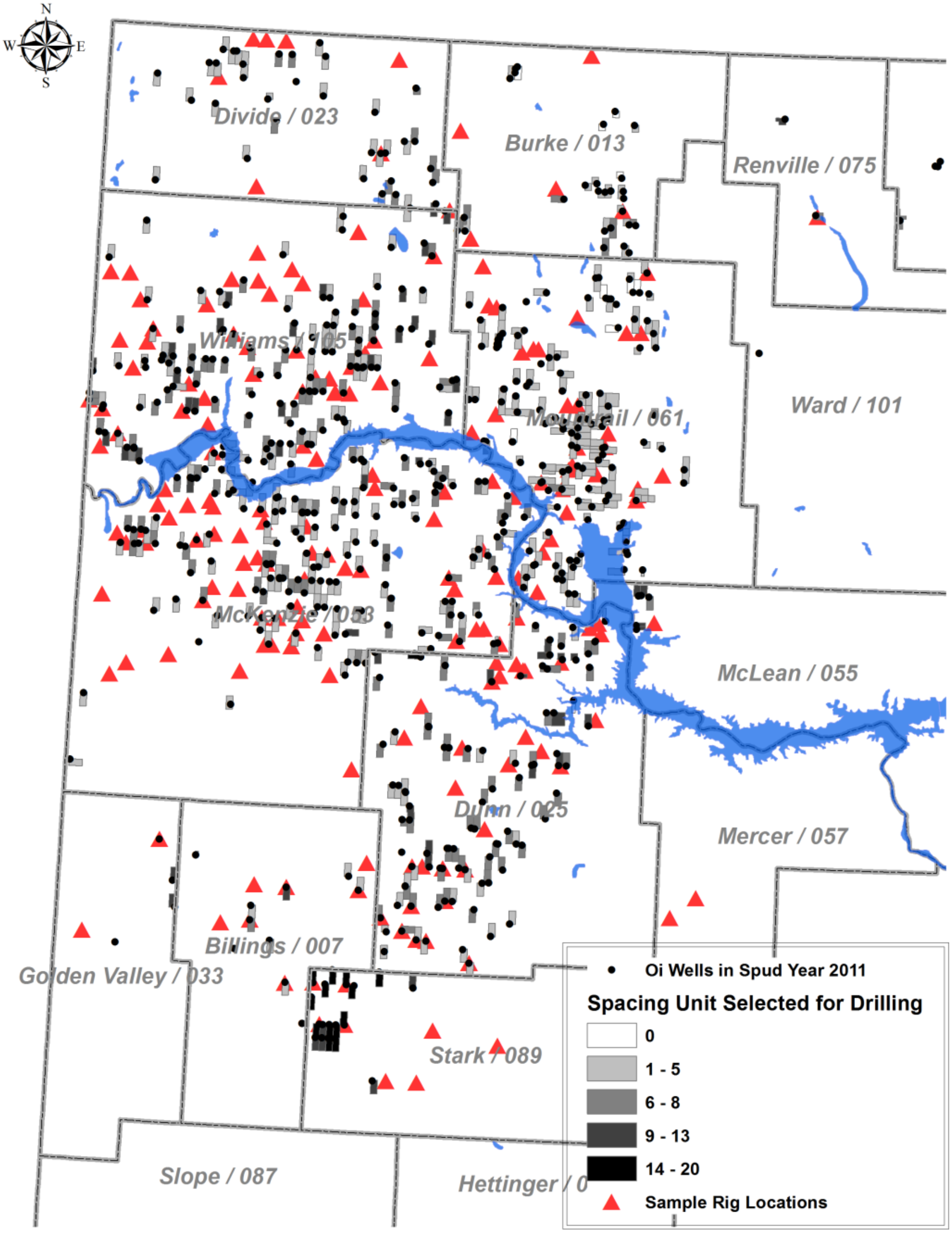

Figure 4.1 Oil wells drilled in 2011 (base year) and spacing units selected to drill in 2011. Sample rigs of all drilling activities through 2011.

4.2 Model Comparison and Interpretation

The overall model as a whole fits significantly given a p-value of 0.001. The Wald test indicates that the model is statistically significant. Table 4.2 shows the coefficients (labeled Estimates), their standard errors (error), the Wald Chi-Square statistic, and associated p-values. The coefficients of the geospatial variables, WELLFREQ5, AREAMILE, and ZONEDENS are statistically significant. For the temporal variables, YEAR2 has a significant impact with a 90% confidence level in the full model. The logit regression coefficients give the change in the *log* odds of the outcome for one unit increase in the full prediction model.

For every one-unit change in AREAMILE, the log odd of drilling increases by 1.8464. Similarly, for every one-unit increase in ZONEDENS, the log odds of being drilled to produce oil increase by 0.5151. The coefficients for the categories of historical drilling year have a slightly different interpretation. However, only significant spatial factors and temporal factors are considered for the nested model to reduce the size of the model and for ease of implementing it in GIS software. The portion of the output labeled Model Fit Statistics describes and tests the overall fit of the model. The -2 Log L (1408.390) can be used in comparisons of nested models; thereby indicating two models (full model and nested model) are not significantly different by both 5% and 10% level (i.e., -2(1248.488-1249.478)=1.98<12.02(90%)). Thus, this study implements the nested model in GIS for simplicity.

In the nested model (Table 4.2), the number of wells within 5 miles (WELLFREQ5) and 15 miles (WELLFREQ15) are statistically significant as well as AREAMILE and ZONEDENS. Among the temporal factors, YEAR2 is statistically significant; however, this study includes all the other temporal variables in the nested model.

4.3 Output and Visualization

In 2005, many wells from the middle of the region gained high weights because the regions have a concentration of shale oils (Figure 4.2). The southeast corner also gains high probability because relatively fewer oil spacing units are shown in the map. In 2020, the likelihood ratio of the drilling locations moves westward because many of rigs have already drilled in the middle of the region, and there are many newer oil wells to the west. In 2025, the probability of drilling spacing units is widely dispersed while the zones located in the middle of the region have high chances to be drilled. In 2030, the probability of drilling spacing units is almost evenly distributed across the region because most of the zones have been developed in the 20 years since 2012.

Table 4.2 Comparison of Full Model and Nested Model of selection of drilling locations.

Independent Variables	Names	Full Model			Nested Model		
		Estimates	Wald Chi-sq	Pr>Chisq	Estimates	Wald Chi-sq	Pr>Chisq
Constant		5.6552	0.0695	0.7921	-9.2857	73.3581	0.0001
YEAR5	Oil drilled 5 year back	-12.9064	0.0006	0.9803	-	-	-
YEAR4	Oil drilled 4 year back	-0.2035	0.1274	0.7211	-0.4736	0.6961	0.4041
YEAR3	Oil drilled 3 year back	-0.9420	2.7207	0.0991^{+}	-0.6961	2.4346	0.1187
YEAR2	Oil drilled 2 year back	-1.1233	9.2485	0.0024*	-0.8349	9.0205	0.0027*
YEAR1	Oil drilled 1 year back	-0.1790	0.5604	0.4541	-0.2362	1.1282	0.2881
AQUA	Water and river	-1.3061	3.1817	0.0745^{+}	-0.8791	2.2032	0.1377
ASSESSUNIT	Relative Oil Recoverable in Williston Basin	1.8464	0.5939	0.4409	-	-	-
OilMEAN	Million Barrels	-1.5482	0.3683	0.5439	-	-	-
WDEPNEAR	Distance from the nearest water depot	-0.1219	1.2905	0.2560	-	-	-
PIPENEAR	Distance from the nearest pipe transloading sites	0.2494	2.5291	0.1118	-	-	-
WELLFREQ5	Number of oil wells within 5 miles	0.4670	4.2297	0.0397*	0.3240	5.9016	0.0151**
WELLFRQ10	Number of oil wells within 10 miles	-0.4782	0.9808	0.3220	-	-	-
WELLFRQ15	Number of oil wells within 15 miles	0.8151	2.8662	0.0905^{+}	0.6582	8.1520	0.0043*
NEARTRUST	Distance from the nearest trust land				-	-	-
NEARWELL	Distance from the nearest active oil well	-0.0890	0.1599	0.6892	-	-	-
AREAMILE	Space of the spacing unit	0.5244	3.8584	0.0495**	0.6316	6.9912	0.0082*
ZONEDENS	Density of the oil zone	0.5151	4.5806	0.0323**	0.5367	8.3890	0.0038*
Global Testing	Likelihood Ratio		81.3151	<0.0001		94.0081	<0.0001
	Score		80.2154	<0.0001		80.0007	<0.0001
	Wald		71.4877	<0.0001		80.4878	<0.0001
Model Fit	-2LogL	1248.488			1249.478		

Note: dependent variable: drilling = 1 if a spacing unit is selected, drilling = 0 if it is not selected; 6670 observations used out of 6711. Individual coefficients are statistically significant at the *1%, **5%, and $^{+}$10% level.

5. CONCLUSION

This paper proposed a geospatial analysis for estimating drilling locations for a 20-year long-term period by investigating the probability of drilling for each spacing unit. The maximum likelihood estimation of the parameters of a logistic regression provides information regarding locations for drilling with regard to well aging, drilling limitations on a spacing unit, closeness to the current wells, and oil density in an oil development zone. The estimation also includes the U.S. Geological Survey's oil assessment.

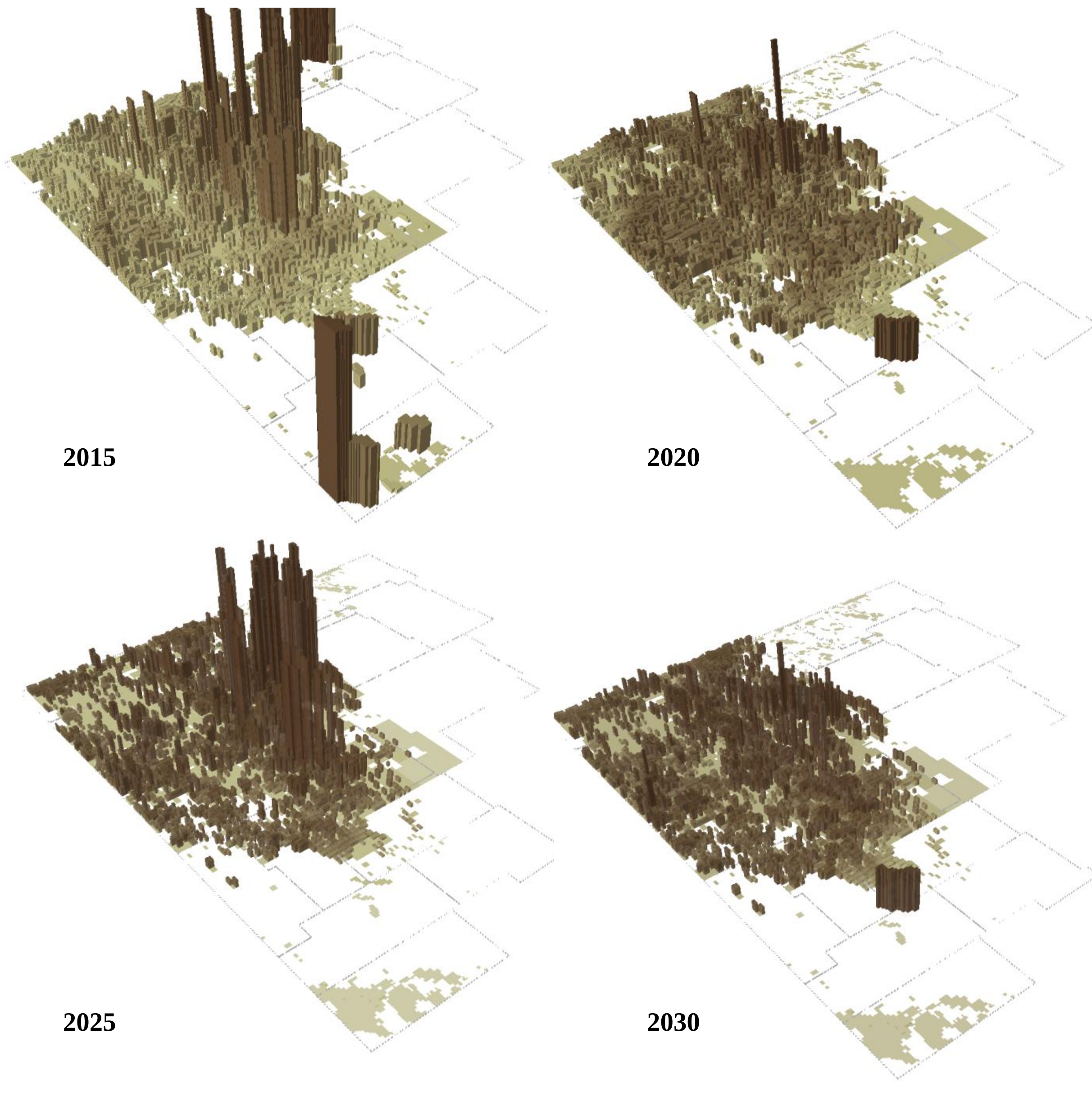

Figure 5.1 Probability of drilling on spacing units using for the next 20 years.

Nevertheless, the estimation process of the drilling locations heavily relies on the future oil forecasting information from state agencies and the U.S. Geological Survey's assessment of the Bakken formation. Therefore, the output of this study should be more in line with the potential oil development strategy for long-range transportation planning. Small variations in the process will result in huge variations at the end of the 20-year long-term planning period. Nonetheless, this model provides information for long-range transportation planning related to shale oil development that provides heavy impact on roads in the United States.

In a future study, other methods, such as bioinformatics, clustering, and simulation, could be implemented and compared with the proposed model in order to improve quality of the output. The proposed model in this study can be utilized for the regions where development begins. When the vertical and horizontal drilling activities are mixed, two different models could be developed to represent each group. The model provides a low level of transferability due to the variation of geospatial patterns, although the procedure can be adopted.

REFERENCES

[1] S. Mitra, D. Tolliver, and A. Dybing, “Analyzing Investments Needed to Support Oil and Gas Production and Distribution,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. No. 2307, pp. 1-8, 2012.

[2] Upper Great Plains Transportation Institute, “Additional Road Investment Needed to Support Oil and Gas Production and Distribution in North Dakota,” September 2010.

[3] Washington State Legislature, “Development Units - Size and Shape,” 1994.

[4] North Dakota Petroleum Council, “North Dakota Oil & Gas Industry: Facts and Figures,” North Dakota Petroleum Council, Bismarck, 2012.

[5] T. Lacoma, “Spatial Forecasting Pattern Analysis,” *eHow: Money,* August 2011.

[6] P. Flom, “Pattern Recognition Analysis,” 2012. [Online]. Available: http://www.ehow.com/facts_6847751_pattern-recognition-analysis.html. [Accessed 16 February 2012].

[7] E. Volz and L.A. Meyers, “Susceptible–infected–recovered epidemics in dynamic contact networks,” *Proceedings of the Royal Society,* p. 2925–2933, 2007.

[8] E. Bertuzzo, R. Casagrandi, M. Gatto, I. Rodriguez-Iturbe, and A. Rinaldo, “On spatially explicit models of cholera epidemics,” *Journal of the Royal Society Interface,* pp. 321-333, 2010.

[9] L. Danon, T.A. House, J.M. Read, and M.J. Keeling, “Social encounter networks: collective properties and disease transmission,” *Journal of the Royal Society Interface,* p. 2826–2833, 2012.

[10] D. Brown, J. Dalton, and H. Hoyle, “Spatial Forecast Methods for Terrorist Events in Urban Environments,” *Intelligence and Security Informatics,* p. 426–435, 2004.

[11] W. Bernasco, “Modeling Micro-Level Crime Location Choice: Application of the Discrete Choice Framework to Crime at Places,” *Journal of Quantitative Criminology*, vol. 26, no. 1, pp. 113-138, 2010.

[12] L. M. Hunt, B. Boots, and P.S. Kanaroglou, “Spatial choice modelling: new opportunities to incorporate space into substitution patterns,” *Progress in Human Geography,* vol. 28, no. 6, pp. 746-766, 2004.

[13] C.R. Bhat and I.N. Sener, “A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units,” *Journal of Geographical Systems,* vol. 11, no. 3, pp. 243-272, 2009.

[14] W. Zhu and H. Timmermans, “Modeling pedestrian shopping behavior using principles of bounded rationality: model comparison and validation,” *Journal of Geographical Systems,* vol. 13, no. 2, pp. 101-126, 2011.

[15] L. Alamá-Sabater, A. Artal-Tur, and J.M. Navarro-Azorín, “Industrial location, spatial discrete choice models and the need to account for neighbourhood effects,” *The Annals of Regional Science,* vol. 47, no. 2, pp. 393-418, 2011.

[16] North Dakota Department of Mineral Resources, “Oil and Gas: ArcIMS Viewer,” 22 Feburary 2012. [Online]. Available: https://www.dmr.nd.gov/OaGIMS/viewer.htm. [Accessed 5 Feburary 2012].

[17] U.S. Geological Survey, “Assessment of Undiscovered Oil and Gas Resources of the Williston Basin Province of North Dakota, Montana, and South Dakota, 2008,” USGS Central Energy Team, 2008.

[18] U.S. Geological Survey, “Energy Resources Program: Williston Basin,” 2008.

[19] G.K. Osadetz, R.L. Snowdon, and W.P. Brooks, “Relationships Amongst Oil Quality, Thermal Maturity and Post-Accumulation Alteration in Canadian Williston Basin (Southeastern Saskatchewan and Southwestern Manitoba),” in *Sixth International Williston Basin Symposium*, Saskatchewan, 1991.

[20] W. Mendenhall and T. Sincich, A Second Course in Statistics: Regression Analysis, Upper Saddle River, New Jersey: Pearson Education Inc., 2003, pp. 449-456.