Unnatural landscapes in ecology: generating the spatial distribution of brine spills

Henriette I. Jager1,∗, Rebecca A. Efroymson1, Kerry L. Sublette2 and Tom L. Ashwood1

1Oak Ridge National Laboratory, Oak Ridge, TN 37831-6036, U.S.A.
2University of Tulsa, Tulsa, OK 74104, U.S.A.

SUMMARY

Quantitative tools are needed to evaluate the ecological effects of increasing petroleum production. In this article, we describe two stochastic models for simulating the spatial distribution of brine spills on a landscape. One model uses general assumptions about the spatial arrangement of spills and their sizes; the second model distributes spills by siting rectangular well complexes and conditioning spill probabilities on the configuration of pipes. We present maps of landscapes with spills produced by the two methods and compare the ability of the models to reproduce a specified spill area. A strength of the models presented here is their ability to extrapolate from the existing landscape to simulate landscapes with a higher (or lower) density of oil wells. Published in 2005 by John Wiley & Sons, Ltd.

KEY WORDS: brine spill; process water; petroleum; stochastic spill model; risk assessment; gamma distribution; Poisson distribution

1. INTRODUCTION

Human activities and their impacts on landscapes and the biological populations they support are a primary concern in applied landscape ecology. Despite this, most research focuses on natural processes and patterns rather than those associated with human activities. Stochastic models are often used to describe natural landscape patterns: for example, Markov models describe both the effects of climate change on vegetation (Shugart and Smith, 1996) and patterns of landuse change (Soares-Filho et al., 2001), and Weibull models have been used to describe spread of forest fire (Li and Apps, 1996). As we demonstrate here, similar models can be used to describe human-created patterns.

Human processes are harder to conceptualize than natural processes using statistical models because we are so intimately familiar with the details. It is difficult for us to imagine that alternative realities might have resulted from the same processes. For example, a particular road might have been here and not there. Human-produced patterns in the landscape strike us as unique, rather than as realizations of a random process that might have produced a different result. Models of human impacts

∗Correspondence to: H. I. Jager, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6036, U.S.A.
†E-mail: jagerhi@ornl.gov
‡This article is US Government work and is in the public domain in the USA.
often reflect this mindset by adopting a map of the actual landscape as a given, rather than conceptualizing disturbance as a stochastic process, simulating potential landscapes, and thereby obtaining more general results. Perhaps another reason that stochastic models are not used more often is that the most common stochastic models produce landscapes with more irregular, fragmented and isotropic patterns than those produced by human activities. Stochastic models, including point processes and spatial statistical models, are rarely designed to produce regular geometric shapes, such as rectangular pastures and linear roads.

Stochastic models that describe human-created patterns offer two advantages that are important in a risk assessment context. First, a stochastic landscape model can be used to address general questions about the influence of human activities. For example, ‘How would ecological communities respond to doubling oil production?’ In contrast, ecological risk predictions based on just the existing landscape apply only to other landscapes with human influences in precisely the same place and following precisely the spatial arrangement as those observed. Second, the use of a stochastic model permits the effects of spatial uncertainty (sensu Jager and King, 2004) in brine scar locations to be incorporated in estimates of ecological risk.

1.1. Disturbance by spills

This article focuses on one type of anthropogenic disturbance: brine spills at petroleum production sites. Petroleum exploration and production can occur in valuable ecosystems such as grasslands that provide habitat to numerous species. Brine spills are common at these sites. For example, the state of Oklahoma reported over 900 brine spills per year (1993–2002; Oklahoma Corporation Commission). ‘Produced water’ or brine is a by-product of petroleum production and can have salt concentrations exceeding 100,000 mg/kg (API, 1997). Because much more saltwater is produced with oil than gas, releases of crude oil and saltwater are more closely allied to oil production than to gas production (Fisher and Sublette, submitted).

Oilfield brine spills represent a much longer-term disturbance than hydrocarbon (oil) spills. Unlike oil, brine has an immediate lethal impact on established vegetation because its high salinity makes it difficult for plant roots to absorb water by osmosis. Recovery of vegetation after a brine spill is much slower than recovery after oil spills because soil conditions suitable for plant life are not restored by biological (microbial decomposition) or by physical processes (leaching, volatilization). Vegetation recovery is particularly slow in clay soils because leaching is retarded (Colgan et al., 2002), and because the high salinity produces a sodic soil with poor structure and low hydraulic conductivity. As sodium displaces calcium from aggregated clays, clay particles disperse in the soil water until they plug off smaller pores in the soil. In the long-term, soil that is unable to support plant life will succumb to erosion, creating an even larger scar. The resultant ‘brine scar’ remains a nearly permanent mark on the landscape.

1.2. Study objectives

The purpose of this study is to develop and parameterize stochastic models for the spatial distribution of brine spills. Levins (1966) asserted that models cannot be simultaneously realistic, general and precise. In this article, we propose two spill-generator models: one more general and the other more realistic. The more general Poisson–gamma model generates a random point pattern of spills with different areal extent. The more realistic, well-complex model sites rectangular well complexes at random, and predicts the distribution of brine spills associated with each complex. These spill-generator models can be used to compare landscapes with increased or decreased petroleum
production. In this article, we describe the two models and demonstrate how input parameters for the models can be estimated for a site-specific application in the Tallgrass Prairie Preserve (TPP) in Osage County, Oklahoma.

2. METHODS

2.1. Empirical spill data

Both spill generator models require parameter estimates for the distribution of brine spill sizes. We constructed an empirical size distribution for brine scars in the TPP. These data were derived from a map of brine spill boundaries (Figure 1) for barren areas outlined visually from Digital Orthophoto Quarter Quadrangles (Bryan Tapp, University of Tulsa, personal communication).

Small spills are much more common than larger spills. We found this to be true for spills in the TPP, and Fisher and Sublette (submitted) found the same pattern among quantified spills reported in Oklahoma between 1993 and 2002. The median reported volume of spills was 6359 L (40 barrels or bbl), but large spills > 79 493 L (500 bbl) did occur each year. The average volume of brine spills in Oklahoma ranged from 14 150 to 25 120 L (89–158 bbl) during this period.

This situation can be described by strongly left-skewed positive distributions, such as the gamma. We therefore fitted a 2-parameter gamma distribution to the observed spill areas in the TPP. We used both a frequentist and Bayesian approach to identify reasonable values of the scale and shape parameters. The univariate procedure in SAS® provided maximum likelihood (ML) parameter estimates, along with a Kolmogorov-Smirnov test for goodness of fit.

Figure 1. Map of brine spills at the Tallgrass Prairie Preserve, Oklahoma, digitized from Digital Orthophoto Quarter Quadrangles by Elizabeth Word, and provided by Bryan Tapp of The University of Tulsa.
The Bayesian approach finds the most likely distribution of parameter values given the observed data. We used Markov chain Monte Carlo to solve for posterior distributions of the two parameters. We adopted an exponential prior distribution with mean 0.6826 for shape parameter, \( \alpha \). We used a gamma prior distribution with shape parameter \( \beta \) (assumes CV = 1) and scale parameter \( \beta \) = 0.001 for the inverse of the scale parameter, \( \beta \). Convergence of estimates was fast, as shown by comparing estimates with five different initial conditions (chains). We used a burn-in of 1000 iterations, and report parameter means and credibility intervals (2.5%ile and 97.5%ile) from 9000 later iterations of 1 chain.

2.2. Poisson-gamma model

The Poisson–gamma model uses a two-step process to generate brine spills on the landscape. First, the spatial distribution of spill centers is generated according to a Poisson distribution in 2D space. Second, the area of each spill is drawn from a gamma distribution.

2.2.1. Distribution of spill centers. The first step is to locate the centers for a specified number of spills, \( N \), on the landscape using a homogeneous 2D Poisson distribution. To visualize this, think of placing each spill by throwing a virtual dart at the gridded landscape.

2.2.2. Distribution of spill areas. The second step is to determine the size of each spill and to distribute this area around each spill center. We ensure that the realized total spill area and number of spills matches the requested values in the following way. If the spill areas are distributed according to a gamma distribution, then the proportions of spill area comprised by the \( N \) spills follow a Dirichlet distribution. We simulate the proportions by randomly selecting the gamma variates, dividing by their sum, and multiplying by the specified total spill area. The gamma parameters can be estimated from empirical data as described in Section 2.1.

2.2.3. Diffusion of spill area from center. We use a random walk algorithm to simulate the spread of spill area outward from the center of each spill. This method is appropriate because diffusion is the aggregated result of random motion of many individual particles (Haefner, 1996). In our case, we envision these particles as ‘random walkers’ carrying brine from one cell to another. The algorithm assumes that random walkers ignore cells that are already filled; they continue from a ‘full’ cell without counting the move as a step. We adjust the ‘viscosity’ by changing the maximum number of steps allowed to each random walker. When landscape boundaries are encountered, random walkers are reflected to the center of the spill. The algorithm stops sending out random walkers: (i) when the spill area is completely distributed to surrounding cells, or (ii) when the total area of the spill reaches the specified total spill area.

If one is interested in representing spills in a particular landscape, and the topography of the region is such that one would expect a directional down-slope bias in the shapes of spills, then it is possible to simulate a biased random walk driven by spatially explicit slope data (e.g. Grunbaum, 2000). Here, we present a conceptual model that assigns a higher probability to walking in the downslope direction. Whereas simple random diffusion assumes that movement in each direction is equally likely, biased diffusion permits a bias in direction (i.e. a non-uniform distribution). Morales et al. (2002) simulated a biased random walk by using a wrapped Cauchy distribution with two parameters, mean direction, \( \mu \), and mean cosine of angular distribution, \( S \). The wrapped Cauchy distribution is symmetric and unimodal, with its mode at \( \mu \). As \( S \to 0 \), the distribution converges to the uniform distribution and as \( S \to 1 \) the distribution tends to a point distribution concentrated in direction \( \mu \) (Fisher, 1993).
Our maps have eight discrete directions instead of one continuous angle. It makes sense to set mean direction $\mu$ to the angle between the current cell and the neighboring cell with the lowest elevation, with a random choice between or among ties. We would like the distribution of probabilities among the eight directions to become more concentrated in one direction as the slope, $\tan \beta = \Delta \text{elevation}/\Delta \text{distance}$, becomes steeper. As a first approximation, we set $S = \frac{\theta}{2}$, which reaches one when the slope is a vertical $90^\circ$ drop. To discretize the Cauchy distribution, we integrate over the $\frac{\pi}{4}$ interval surrounding the average angle of each neighboring quadrant, where $0^\circ$ is north. The resulting discrete probabilities, $P_{r_i}$, of diffusing toward each of eight adjacent octants are obtained by integrating the density function from plus to minus $\frac{\pi}{8}$ surrounding the mean angle for the quadrant, $\frac{\pi}{8}(2i - 1)$, or subtracting endpoint values of the cumulative distribution function, $F$ (Equation 1):

$$F(\theta) = \frac{1}{2\pi} \cos^{-1}\left[\frac{(1 - S^2) \cos(\theta - \mu) - 2S}{1 - S^2 - 2S \cos(\theta - \mu)}\right], \quad 0 \leq S \leq 1, 0 \leq \theta \leq 2\pi$$

$$P_{r_i} = F\left(\frac{\pi i}{4}\right) - F\left(\frac{\pi(i - 1)}{4}\right), \quad i = 1, 2, \ldots, 8 \tag{1}$$

2.3. Well-complex model

The well-complex model is more realistic than the Poisson–gamma model because spill probabilities are conditioned on the locations of wells and the pipes that connect them. We developed a conceptual model during a visit to the Tallgrass Prairie Preserve, which lies within a large and mature oil production region of Oklahoma. Small, independent producers currently operate approximately 100 wells in the TPP. Wells are arranged in rectangular grids spaced about 270 m apart (Figure 2). A typical rectangular complex might contain four rows with five wells per row.

Each active well produces 160–320 L (1–2 bbl) of oil daily, and ten times that much brine. Gathering lines carry an oil–brine mixture from a pump jack at each individual well to a tank battery. Gathering lines often run directly to the battery of tanks, but lines from newer wells are sometimes joined to those from older ones to minimize pipe. These pipes leak with the highest frequency (3 spills y$^{-1}$) for three reasons: (i) they represent the greatest total length of pipe in the system; (ii) they are vulnerable because they lie on the surface; and (iii) they carry a corrosive fluid.

At the tank battery, oil is separated from brine. Reverse gathering lines return brine to non-producing wells, where the brine is re-injected. Brine-only spills typically occur at the tank battery or between the tank battery and the brine injection well. Larger collection lines carry the oil in a straight path from each tank battery to a consolidating tank, where the oil is collected by truck.

In the well-complex model, we focus on spills along gathering lines running from individual wells to each tank battery. These spills are assumed to contain brine. Collection-line spills are not represented in these models because they are less common than gathering-line spills and they do not carry brine.

The well-complex model has three components for simulating the spatial distribution of spills in a 2D landscape partitioned into rectangular cells. First, we use a statistical model to locate well complexes. Second, we randomly select distances along the pipeline at which spills occur within each complex, and identify the cells at which these distances are reached. Third, we assign the area of each spill and distribute spill area to the central cell and its neighbors.

2.3.1. Well-complex geometry. After specifying the number and dimensions of well complexes, we locate the first well complex by siting the northwest corner in a 30 by 30-m cell selected according to uniform distribution from the grid of cells that span the landscape. The algorithm then checks whether the complex fits entirely within the dimensions of the landscape. If not, the process is repeated until a suitable location is found. Once the first complex is sited, subsequent complexes are sited with the additional constraint that no overlap with previously sited complexes is allowed.

To simulate the gathering lines, we locate a tank battery in the cell north and west of the northwest corner of each complex. For simplicity, pipes are assumed to run directly from each well in the complex to the tank. The model determines the length of pipe running through each cell, which is used in estimating the likelihood of a spill at that location.

2.3.2. Number of spills. We assume that the likelihood of encountering a spill along any segment of pipe of a specified length is constant, so that the likelihood of a spill within a cell increases with the length of pipe running through it. This is a simplification, as we know that local conditions can influence the probability of spill (see Discussion). Equation (2) describes the distance along the pipes before encountering the next spill, where parameter $\lambda$ is the average number of spills per meter of pipe, and $x$ is the linear distance of pipe (m). We specify the expected number of spills per complex, which is used to obtain $\lambda$, but the sequences of random distances selected according to the gamma distribution may result in fewer spills or more spills:

$$\text{Prob} \{ D \leq x \} = 1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda x} (\lambda x)^k}{k!}, \quad x > 0$$  \hspace{1cm} (2)
2.3.3. Area of brine spills. Two input parameters provided to the well-complex model are the number of well complexes and the total area of spill for the entire landscape. We assume that the specified total spill area is apportioned equally among the complexes. The areas of brine spills within each complex are assumed to follow a gamma distribution (see Section 2.1), and the same model is used to simulate diffusion as in the Poisson–gamma model.

2.4. Simulations

We demonstrate the two spill-generating models by generating landscapes that differ in the number of spills and the percentage of area on which spills have occurred. Note that a site-specific application would use the parameters obtained from empirical spill data. We partitioned a 268.2099-km² landscape centered on the TPP into 900-m² grid cells organized into 617 rows and 483 columns. We allowed up to 50 steps during the random walk process. For the Poisson–gamma model, we set the $CV = 3.5$, and compared landscapes with 25 and 50 spills. For the well-complex model, we fixed the number of spills per complex to five, and created landscapes with 5 or 10 well complexes. For both models, we compared landscapes generated with different percentages of total spill area (10%, 20%, 30%, 40%, and 50%). We present maps to illustrate the effects of varying these parameters.

3. RESULTS

3.1. Empirical spill data

Empirical spill areas give an indication of the shape of frequency distribution typical of brine spills in the TPP. The empirical distribution for these 122 spills in the TPP is highly left-skewed (Figure 3), with mean area, $A = 911.8$ m² ($CV = 1.21$). The maximum likelihood (ML) estimate of shape, $\alpha = 0.6827$, and scale, $\beta = 1335.7$, were very close to moment estimators, shape parameter $\alpha = CV^{-2}$ and the scale parameter $\beta = A/\alpha$. Using the mean of the posterior distribution gave lower parameter estimates than ML. For shape, mean $\alpha = 0.5521$, and 95% of the posterior distribution fell between 0.4437 and 0.6741. For scale, mean $\beta = 1253.4$, and 95% of the posterior distribution fell between 938.1 and 1777.1. The Kolmogorov–Smirnov D-statistic, the maximum deviation between the empirical and ML-fitted cumulative distribution, was 0.1037, which corresponds to $p > 0.25$. This simply means that the quantity of data was not sufficient to detect deviation of the empirical distribution from the gamma distribution with ML parameter estimates.

Figure 3. Comparison of empirical and fitted frequency distributions of brine spill areas in the Tallgrass Prairie Preserve, Oklahoma

Figure 4. Example landscapes generated by the Poisson–gamma model. Each landscape has 25 spills with areas varying according to a gamma distribution. The percentage of spill area increases from right to left, and two replicate maps are shown top and bottom.

Figure 5. Landscapes generated using the Poisson-gamma model with 20% of area covered by brine spills. The top row shows maps with 25 spills and the bottom row has 50 spills. Landscapes that allow random walkers to travel farther (250 vs. 50 steps) show more diffuse brine scars. Landscapes with a higher C.V. show an uneven distribution of spill areas.
3.2. Poisson–gamma model

The Poisson–gamma method for generating spills produced landscapes with the specified proportion of area covered by brine and the specified number of spills (Figure 4). The series of landscapes shown in Figure 5 demonstrate the effects of varying parameters from the values (center column) described in Section 2.4. Landscapes on the left have more-diffuse spills because each simulated random walker was allowed 250 steps instead of 50 (fewer walkers traveled farther). Landscapes on the right have less variation in spill sizes because the specified coefficient of variation was lower (0.5 instead of 3.5). The number of spills is 25 in the top row of landscapes and 50 in the bottom row (Figure 5).

The Poisson–gamma model provided good control of the proportion of spill area for both 25 and 50 spills (Figure 6a). Total area did not vary much among replicate maps. In a few cases, maps with a lower percentage of spill area than the target were produced (symbols below the 1:1 line in Figure 6a).

Figure 6. The distribution of realized spill percentages in 30 replicate simulations and target percent spill areas is compared for (a) the Poisson–gamma with 25 and 50 spills and (b) the well-complex model with 10 complexes. The box-whisker diagram encloses the 25–75th percentile in a box, with the 5th and 95th percentiles shown as dots or triangles, the median as a solid line, and the mean as a dashed line.

3.3. Well-complex model

The well-complex method for generating spills produced landscapes with a range of specified spill characteristics (Figure 7). These maps illustrate the ability to control the degree of fragmentation in the landscape by varying the number of well complexes.

The proportion of spill area on the landscapes tended to be lower than that specified due to overlap of spills among well complexes (Figure 6b). This deviation increased with increasing proportions of the landscape covered by spills. We fitted a linear regression model to the simulated deviation shown in Figure 6b: deviation = 0.508 (target percent) – 4.375, $R^2 = 0.92$. This deviation could be added to the user-specified percentage in the simulation model to eliminate bias.

4. DISCUSSION

We found that the well-complex model permits less control over the proportion of spill area than the Poisson–gamma model, but it has the advantage of using parameters that are easier to estimate in case studies of oil fields. When a large percentage of the landscape is impacted, simulated spills that originate in different complexes may overlap. As a result, the total realized spill area simulated by the well-complex model tends to underestimate the total specified for the landscape. However, if the percentage of area covered by brine spills at the site of interest is lower than 10%, this bias may be insignificant.

One can choose between the two spill-generator models based on the type of parameter information available and the need to control spill areas. If realistic information on well spacing for wells arranged...
on a grid is available, and if the percentage of spill area is < 10%, then the well-complex model can be implemented. Otherwise, the Poisson–gamma model is simpler to use and gives more general results. Calibration can be accomplished by fitting a gamma model to the observed distribution of spill areas, either for a single oil field or over a larger spatial area. Note that other spill size distributions could also be implemented.

The spill-generator models presented here are designed to address general questions, for example, about how changes in petroleum production might influence landscapes and the ecosystems they support. Simulated brine scars are placed without regard to site-specific topographic effects on the spread of brine spills in a particular landscape. This simplification might give somewhat unrealistic results if, for example, steep slopes in an area produced long, narrow spills. For a risk assessment that requires such situation-specific details, it may be more appropriate to use a mechanistic, transport model. For example, Paige et al. (2003) developed a model that uses GIS data describing oil pipelines, in combination with elevation data and stream network data, to predict exposure of high consequence areas (e.g. populated areas, areas unusually sensitive to environmental damage, and commercially navigable waterways) to pipeline spills.

The two models described here are also static—they do not simulate the temporal dynamics of brine spills. The well-complex model could be developed further as a tool for prospective ecological risk assessment by superimposing brine spills that occur at different times. This would require data on the temporal frequency of brine spills. Although we lack temporal data at our site, we do have qualitative information about mechanisms involved that might be used in the future to condition failure (spill) rates on local attributes. For example, the failure rate of pipelines at our study site is higher near roads, because vehicles push rocks into pipelines. Brine scars tend to expand through erosion from year to year, particularly on steeper slopes (greater than 8%, API, 1997). In addition, failure of brine-containing pipes is more common when pipe joints occur in depressions because of the internal corrosion from stagnant fluid. Newer, polymer pipelines fail less frequently than metal lines because they are usually buried and do not corrode.

With energy costs rising, there is mounting pressure to increase petroleum production, and to quantify the associated ecological risks. One strength of our spill-generator models is that they can be used to simulate oil fields with more wells than are currently in operation. We illustrated the spill-generator models with between 10% and 50% of area covered in brine scars. Although these percentages are higher than the percentage of brine-impacted area in the TPP (about 1%), we were unable to find statistics describing typical well and spill densities in petroleum production areas.

Future research is needed to better describe how the total area of brine spills increases with an increased number of oil wells. Unfortunately, brine spills in terrestrial ecosystems are poorly documented. This has been attributed to the focus of U.S. regulatory reporting requirements on crude oil spills in aquatic, and particularly marine, ecosystems (Fisher and Sublette, submitted). In a review of spills reported to the National Response Center (U.S.A.), Fisher and Sublette found it surprising that, ‘...even though pollution prevention, rational management and decision making regarding environmental issues requires information, little is known about the specific origins, causes, and magnitudes of pollution incidents that result from exploration and production activities.’

In conclusion, we anticipate that ecological risk assessment will be an important application of spill-generator models. For example, we have used the Poisson–gamma spill model, in combination with a population model, to quantify spill frequency and area thresholds leading to high risk of local extinction for the American badger (*Taxidea taxus*) (Efroymson et al., 2004; Jager et al., submitted).
ACKNOWLEDGEMENTS

This research was sponsored by the National Petroleum Technology Office, U.S. Department of Energy, and performed at Oak Ridge National Laboratory (ORNL). ORNL is managed by UT-Battelle, LLC, for the DOE under contract DE-AC05-00OR22725. We appreciate the spill data provided by Bryan Tapp (University of Tulsa) and his students. We thank Rick Schmoyer, Bill Hargrove, and an anonymous reviewer for providing technical reviews. Robert Washington-Allen was kind enough to review the proposed model for elevation-biased diffusion.

REFERENCES


