2016

Predicting cannabis cultivation on national forests using a rational choice framework

Frank H. Koch
USDA Forest Service, fhkoch@fs.fed.us

Jeffrey P. Prestemon
USDA Forest Service, jprestemon@fs.fed.us

Geoffrey H. Donovan
USDA Forest Service, gdonovan@fs.fed.us

Everett A. Hinkley
USDA Forest Service, ehinkley@fs.fed.us

John M. Chase
USDA Forest Service, jchase@fs.fed.us

Ecological Economics 129 (2016) 161–171

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Analysis

Predicting cannabis cultivation on national forests using a rational choice framework

Frank H. Koch [a,*], Jeffrey P. Prestemon [b], Geoffrey H. Donovan [c], Everett A. Hinkley [d], John M. Chase [e]

[a] Eastern Forest Environmental Threat Assessment Center, Southern Research Station, USDA Forest Service, PO Box 12254, Research Triangle Park, NC 27709, USA
[b] Forest Economics and Policy Work Unit, Southern Research Station, USDA Forest Service, PO Box 12254, Research Triangle Park, NC 27709, USA
[c] Goods, Services, and Values Program, Pacific Northwest Research Station, USDA Forest Service, 620 Southwest Main, Suite 400, Portland, OR 97205, USA
[d] Forest Service Engineering, USDA Forest Service, 201 14th Street Southwest, Washington, DC 20024, USA
[e] Forest Inventory and Analysis, Pacific Northwest Research Station, USDA Forest Service, 620 Southwest Main, Suite 400, Portland, OR 97205, USA

ARTICLE INFO

Article history:
Received 16 June 2015
Received in revised form 22 March 2016
Accepted 10 June 2016
Available online 22 June 2016

Keywords:
Cannabis
Illegal cultivation
National forests
Predictive model
Marijuana

ABSTRACT

Government agencies in the United States eradicated 10.3 million cannabis plants in 2010. Most (94%) of these plants were outdoor-grown, and 46% of those were discovered on federal lands, primarily on national forests in California, Oregon, and Washington. We developed models that reveal how drug markets, policies, and environmental conditions affect grow siting decisions. The models were built on a rational choice theoretical structure, and utilized data describing 2322 cannabis grow locations (2004–2012) and 9324 absence locations in the states' national forests. Predictor variables included cannabis market prices, law enforcement density, and socioeconomic, demographic, and environmental variables. We also used the models to construct regional maps of grow site likelihood. Significant predictors included marijuana street price and variables associated with grow site productivity (e.g., elevation and proximity to water), production costs, and risk of discovery. Overall, the pattern of grow site establishment on national forests is consistent with rational choice theory. In particular, growers consider cannabis prices and law enforcement when selecting sites. Ongoing adjustments in state cannabis laws could affect cultivation decisions on national forests. Any changes in cannabis policies can be reflected in our models to allow agencies to redirect interdiction resources and potentially increase discovery success.

Published by Elsevier B.V.

1. Introduction

In the United States, illegal cannabis cultivation on public lands is a major problem for land management agencies (Bouchard, 2007). In particular, national forests of the United States have experienced rising rates of illegal cultivation. The US Department of Justice's National Drug Intelligence Center (2011) reported that the rate of outdoor-grown cannabis seizures nationwide increased 150% from 2005 to 2010, fueled by apparent demand growth and profit-earning opportunities for domestic producers; of all outdoor-grown plants seized in 2010, 44% came from federal lands, primarily national forests. Large-scale producers may be motivated by the perception that domestic cultivation is less risky (i.e., in terms of detection by law enforcement) than importing cannabis across national borders (Barratt et al., 2012; Bouchard, 2007). Domestic producers also face low average costs, at $75 per pound, and can sell their output for up to $7000 per pound. Even in states such as Colorado and Oregon where cannabis possession, distribution, and production were recently legalized, illicit cultivation on national forests and other federal lands is likely to persist, either to supply states where cannabis is still prohibited (Roberts, 2014) or to avoid the taxes and regulations imposed on licensed growers.

Illegal grow operations endanger those who visit or work on national forests. They also cause extensive ecological damage and require costly clean-up (Liddick, 2010; Tynon and Chavez, 2006). Unfortunately, finding cannabis grow sites ("grows") is difficult given available enforcement resources, which must be applied to extensive areas of public land that may be suitable for grow operations (Chavez and Tynon, 2000). Therefore, law enforcement agencies need tools that can help them allocate scarce resources to improve rates of interdiction success. For example, they might employ mathematical models to predict where certain crimes – in this case, illegal cannabis cultivation – will occur in the future, a practice known as prospective hotspotting. Hotspotting techniques based on spatial patterns of historical crime occurrence data are widely used by law enforcement, but such techniques essentially assume that new crimes occur near where they happened previously (Caplan et al., 2011). This is also true of hotspotting methods (e.g., Bowers et al., 2004; Johnson and Bowers,

* Corresponding author.
E-mail addresses: fhkoch@fs.fed.us (F.H. Koch), jprestemon@fs.fed.us (J.P. Prestemon), gdonovan@fs.fed.us (G.H. Donovan), ehinkley@fs.fed.us (E.A. Hinkley), jchase@fs.fed.us (J.M. Chase).

http://dx.doi.org/10.1016/j.ecolecon.2016.06.013
0921-8009/Published by Elsevier B.V.

2004; Ratcliffe, 2004; Rossmo, 1999) that emphasize temporal as well as spatial patterns of past criminal activity. Recently proposed approaches (e.g., risk terrain modeling) supplement historical crime occurrence data with additional data on crime-related variables to better identify hotspots (Caplan et al., 2011; Wang et al., 2013). Nevertheless, these new methods are primarily predictive in purpose, and not inferential in the sense of uncovering and understanding the roles of important drivers of crime. With respect to illegal cannabis cultivation, we believe that a more effective approach could be developed not just from knowledge of recent grow locations, but also from an understanding of grower decisions. Models structured in this way are potentially more capable of handling shifts in the decision-making environment, for example due to spatial and temporal changes in the risks and rewards of a crime. They have the additional advantage of providing inference about the importance of various factors as aspects of an underlying theoretical framework of a crime.

Illegal cultivation on national forests can be explained using rational choice theory (Becker, 1968; Cornish and Clarke, 1986, 1987). Cannabis growers, like other criminal offenders, are rational agents (e.g., Akers, 1990), and they choose locations (or victims) based on the situational status of those locations (or victims). Rational choice theory befits the analysis of criminal events, in no small part because it adopts the premise that situational (i.e., environment-describing) variables can help to explain these events (Hirschi, 1986; Weisburd and Piquero, 2008). Evidence suggests that prospective criminals often behave as if they are rational (Nagin and Paternoster, 1993), especially with respect to crimes involving monetary gains, even when emotions enter into their decision-making (Paternoster and Simpson, 1996; van Gelder and de Vries, 2014). Furthermore, drug trafficking organizations (DTOs) are thought to dominate cannabis production in West Coast national forests (Weisheit, 2011), and the decision-making by these sorts of criminal groups would seem to be well represented by a rational choice model that defines the expected costs and benefits of crime commission (e.g., Desroches, 2005).

While rational choice theory provides an overall construct of criminal decision-making, other theories from criminology help to explain how costs and benefits come together to determine decision outcomes. For instance, routine activities theory (Cohen and Felson, 1979) asserts that many crimes occur due to the convergence of three conditions: a likely offender (someone who is able and motivated to commit a crime); a suitable target (depending on the type of crime, a person or location perceived to be vulnerable or conducive to the crime); and the absence of a capable guardian (a person or thing that – as opposed to an incapable guardian – serves as a deterrent to the crime). Thus, a key aspect of the environment that a potential offender faces is the presence of factors that make the offender more or less visible to capable guardians (e.g., Jeffery, 1977), including law enforcement. Because they influence how the offender perceives the likelihood of being caught and suffering consequences, visibility factors can affect the offender's decisions significantly. These effects can be complex, non-linear, and bi-directional, as illustrated by the example of a cannabis grower selecting a new cultivation site: site preparation often requires large quantities of supplies and equipment (e.g., PVC tubing for irrigation, tools and herbicides for removing native vegetation), so locations close to a road would logically be appealing, yet locations close to a road are also more likely to be discovered by law enforcement or forest visitors.

The environment can also include factors that affect the opportunity costs of being caught, including penalties for being caught (sentences or fines) and lost wages or work opportunities related to imprisonment (e.g., Aaltonen et al., 2013; Burdett et al., 2003; Gould et al., 2002), as well as the opportunity cost of time needed to carry out the criminal activity. The environment might further be described by higher-level socioeconomic factors governing the perceived rewards from crime commission. For example, the prices that can be obtained from the sale of illicit drugs are affected by aggregate demand for and supply of such drugs, which respond to public policies directed at both producers and consumers. Finally, the reward gained by a producer physically varies across space. Ultimately, because all of these environmental factors vary over space and time, the incentives for grow establishment also vary over these dimensions.

These concepts can be used to model illegal cannabis cultivation activities on national forests in the United States, by connecting grower decisions statistically to factors affecting cannabis production risks and rewards. Although our focus is illegal cannabis cultivation, this represents just one example from a class of problems where the factors that determine the spatial pattern of a phenomenon are uncertain and resistant to simple inference. Other examples might be predicting locations where an invasive species is likely to become established or identifying hotspots of illegal wildlife poaching or plant harvesting. In such cases, human activities (e.g., travel for recreation or commerce) often strongly influence the observed pattern (Gallardo et al., 2015), but the nature and degree of that influence may be difficult to ascertain because the data available to describe the pattern (e.g., reports of crime occurrence in the field) may be incomplete or otherwise biased. Therefore, another important objective of our work was to outline a conceptual approach that could be applied to this general class of problems.

2. Methods

Predicting cannabis grow locations resembles how ecologists model the geographic distributions of species based on occurrence data. The fundamental principle behind species distribution models is that spatial variation in species occurrence can be described using environmental factors (e.g., climate or topography) that also vary across the occupied space (Elith and Leathwick, 2009). Historically, ecologists have employed regression methods (e.g., generalized linear models, especially logistic regression) to predict species distributions and to explore ecological relationships between the underlying drivers (Austin, 2007; Elith and Leathwick, 2009). In recent years, regression-based approaches have increasingly been supplanted by methods adapted from machine learning and data-mining literature, including decision trees and decision-tree ensembles (e.g., boosted regression trees, random forests), artificial neural networks, maximum entropy models, and genetic algorithms (Elith et al., 2006). While these methods have documented advantages in terms of predictive success in some empirical applications, they are complex and often opaque (Elith and Leathwick, 2009), limiting their suitability for examining interactions among explanatory variables, including endogeneity. In particular, we were concerned about the potentially endogenous relationship between grow location and cannabis price: higher price may encourage more grows, but more grows may reduce price. Consequently, we chose to use regression methods (i.e., logit and probit regression) in our analyses that allowed us to address the potential endogeneity straightforwardly. Furthermore, logit and probit regression are commonly used in analyses involving rational choice, as detailed below.

2.1. Theoretical Framework

Becker (1968) provided a formal exposition of rational choice theory in terms of expected utility:

$$EU(C) = [1-\pi(\mathbf{z})]u(B)-\pi(\mathbf{z})u(A)-c(C), \quad (1)$$

where EU denotes expected utility, C denotes a criminal action, π is the perceived (by the criminal) probability of suffering a criminal sanction, $\mathbf{z}$ is a vector of exogenous variables affecting the probability, $u(B)$ is the utility gain from committing the crime, $u(A)$ is the utility loss from being caught, and $c(C)$ are the direct costs of committing the crime. The vector $\mathbf{z}$ may also include variables describing the presence of police or other capable guardians. The benefits of committing a crime depend on the size of the reward. In the case of a crime such as cannabis cultivation for sale to distributors, the benefits could be connected to the quantity, which is affected by the productivity of a grow site, and the subsequent value of the cannabis produced, which is governed by cannabis market prices. The direct costs of crime commission could include the opportunity cost of engaging in another activity (e.g., wage labor) and the costs of supplies.

Eq. (1) strictly applies to the decision-making process of an individual. Nonetheless, empirical tests of rational choice theory using aggregate (population-level) data on the occurrence of criminal events are possible given a site-specific expression of the decision criterion. This testing framework involves modeling the relationship between the spatial (or more properly, spatiotemporal) pattern of crime occurrence and the spatial (or spatiotemporal) pattern of expected benefits and costs. Briefly, the probability that a crime occurs at a given location i (suppressing the time dimension to simplify exposition) can be specified as a binary choice set (Greene, 1990, p. 662):

$$\begin{aligned} &Prob(a\ \ crime\ \ occurs) = Prob(C_i = 1) = F(\mathbf{x}_i, \boldsymbol{\beta}), \\ &Prob(a\ \ crime\ \ does\ \ not\, occur) = Prob(C_i = 0) = 1 - F(\mathbf{x}_i, \boldsymbol{\beta}), \end{aligned} \tag{2}$$

where $\mathbf{x}_i$ contains variables that measure the expected benefits and costs of crime commission and $\boldsymbol{\beta}$ is a static vector of parameters conformable to $\mathbf{x}_i$. Given data on crime occurrence, $C_i = [0,1]$, at each point i in a set of locations, a binary choice model can be estimated. To ensure that the probability of crime occurrence is bounded by 0 and 1, either the logistic or the normal cumulative distribution function is typically used to describe the function $F(\cdot)$. Binary choice models are usually estimated via the maximum likelihood method; the log-likelihood function for estimating the models is generally specified as

$$\max L(\boldsymbol{\beta}) = \sum_{i=1}^{N} C_i \log\left[1 - F\left(-\boldsymbol{\beta}'\mathbf{x}_i\right)\right] + (1 - C_i) \log\left[1 - F\left(-\boldsymbol{\beta}'\mathbf{x}_i\right)\right]. \tag{3}$$

2.2. Empirical Approach

We modeled the relationships in $F(\cdot)$ using a stratified sample of our study area, which consisted of all national forests in California, Oregon, and Washington (Fig. 1). From 2004 to 2012, 2322 illegal cannabis grows ($C_i = 1$) were discovered on these forests, the majority of them in California. The median number of plants per grow was 2134. Because the grow location data did not document absences ($C_i = 0$, locations where there was no evidence that a grow existed), we specified them using a different data source. The USDA Forest Service's Forest Inventory and Analysis (FIA) program maintains a series of permanent and regularly surveyed plots across all public and private US forest land (Reams et al., 2005). The plots are approximately 5 km apart. Many of these plots are in remote locations, so survey crews sometimes come across grows. To protect crew safety, if a grow is encountered on or near a plot location on national forest lands, the survey crew will leave the area immediately without measuring the plot. Conversely, it can be presumed that any plot that has been surveyed did not have a grow present ($C_i = 0$) at the time of survey. Thus, to describe absences in our sample, we used data from 9324 FIA plots surveyed between 2004 and 2008 on national forests (i.e., federal lands managed by the USDA Forest Service) of California, Oregon, and Washington. No other lands were considered in this study.

In our case, $Prob(C_i = 1)$ from Eq. (2) is the probability that a grow site is present in a location. However, our sample only included discovered grow sites, which represented some unknown fraction of all locations where grow sites were present during the study period. In turn, we assumed that the set of discovered grow sites was perfectly representative of the larger set of all grow sites (i.e., both discovered and undiscovered), such that all of the sites shared the key attributes that gave our models predictive power, even though they were estimated with only the discovered grow sites (and absences). While this may seem like a minor assumption, its significance is heightened by the

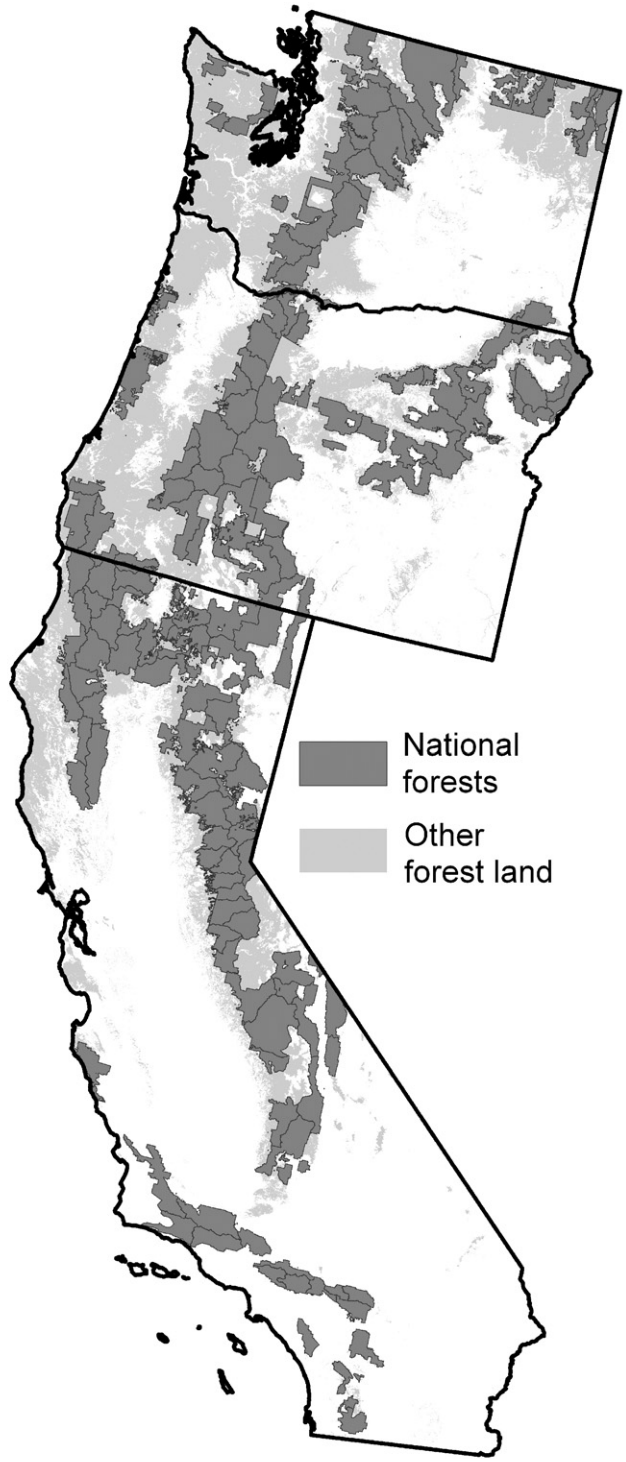

Fig. 1. Study area: national forests in California, Oregon, and Washington.

fact that the sample is response-based rather than random, since grow occurrences ($C_i = 1$) arrived in the sample only after they were discovered, which did not happen randomly. We presumed that our data set included an over-sample of grows relative to their true frequency of occurrence among all possible locations in the study region. Regardless, a response-based sample has important implications for the validity and consistency of the output probability estimates (Hsieh et al., 1985; Manski and Lerman, 1977; Xie and Manski, 1989). These implications relate to assumptions about the distribution (i.e., logistic or normal) chosen to describe $F(\cdot)$ and, in turn, the type of statistical model chosen to analyze the sample. If the occurrence process follows a logistic distribution, then the logit model produces unbiased model parameter estimates (except for the intercept term), in spite of any sampling bias. Alternatively, if the occurrence process follows a normal distribution, then a probit model is necessary, which, in the context of response-based sampling, requires a weighting scheme to avoid model bias. (In our particular case, we intended for the weighting scheme to account for any bias that violated our assumption about the sample's representativeness.) See Appendix A in the supplementary material for details about both weighted probit and unweighted logit model estimation.

We had no a priori knowledge of the true underlying probability distribution of grow occurrence in our data. Moreover, because the logistic and normal distributions are quite similar in shape, logit and probit models applied to the same data often yield similar results (Aldrich and Nelson, 1984), and various performance indicators (e.g., statistical goodness-of-fit measures) may not definitively show either the logit or probit model to be more appropriate for a particular data set. For these reasons, we estimated both unweighted logit and weighted probit models from our sample.

2.3. Model Development

We chose candidate explanatory variables for our models based on rational choice theory and conversations with law enforcement and public land managers. These variables are listed in Table 1 and are documented more fully in Appendix A. A few variables are noteworthy for how they were handled during model estimation. The first is cannabis price; because of both spatial and temporal gaps in the raw price data, we estimated prices for the entire set of observations via an imputation procedure (Rubin, 1987) described in Appendix A. Then, to account for potential endogeneity in the price variable, we implemented a control function (Hausman, 1978), using latitude as an instrument, in both the logit and probit models. We chose latitude as our instrument because cannabis prices in the US follow a predictable latitudinal gradient, decreasing steadily north of Mexico (Caulkins and Bond, 2012). Essentially, latitude represents a measure of the influence of Mexican cannabis imports on supply and demand, and thus on prices. (Note also that the ecological variables in the models account for geographic variation in suitable growing conditions, so latitude is not measuring this variation.) We included both the predicted price variable and a variable containing the residuals of the control function equation in the main model; the predicted price variable captured the true effect of cannabis price, while the control function residuals variable captured the endogeneity. Additionally, due to the uncertainty in the imputed estimates, we also developed alternative logit and probit models that lacked cannabis price as an explicit predictor. In those cases, we used dummy variables for the states of Oregon and Washington to account for lower price levels found in those states compared to California.

Illegal grows are discovered both deliberately (by law enforcement) and accidentally (by national forest visitors). In other words, discoveries are a function of where visitors go and where law enforcement personnel patrol, neither of which is likely to be distributed evenly or randomly across the landscape. Thus, grow discoveries in our data set were non-random, necessitating post-stratification of our data on grow presences and absences. We identified two other variables, law

Table 1
Candidate explanatory variables for the logit and probit models.

Variable	Description	Scale/reporting level	Data source/citation
Observation year	Year when grow was detected, or for absences, when FIA plot was surveyed	n/a	USDA Forest Service, Law Enforcement and Investigations; Forest Inventory and Analysis Program
Observation state	State (CA, OR, or WA) in which observation was recorded	n/a	Tele Atlas North America (2008)
Elevation[a]	Values from digital elevation model (DEM)	10 m for OR and WA; 30 m for CA (raster data)	US Geological Survey (2013a)
Percent slope[a]	Values from slope map generated from DEM	10 m for OR and WA; 30 m for CA (raster data)	Derived from elevation data
Aspect[b]	Values from aspect map generated from DEM	10 m for OR and WA; 30 m for CA (raster data)	Derived from elevation data
Average July precipitation	30-year climatological mean (1971–2000)	≈1 km (raster data)	PRISM Climate Group, Oregon State University (2006a)
Average maximum July temperature	30-year climatological mean (1971–2000)	≈1 km (raster data)	PRISM Climate Group, Oregon State University (2006b)
Distance to nearest water feature	Euclidean distance (m) from sample location	1:12,000–1:100,000 (vector data)	US Geological Survey (2013b)
Forest type	Forest type assigned by statistical model	250 m (raster data)	USDA Forest Service (2008)
Retail wage rate, weekly[c]	Annual average of the quarterly rate (current $)	By county	US Bureau of Labor Statistics (2013a)
Consumer price index[d]	All city average (1999 base)	National	US Bureau of Labor Statistics (2013b)
Unemployment rate[c]	As measured on July 1 of calendar year	By county	US Bureau of Labor Statistics (2013c)
Law enforcement density (a) one year prior and (b) two years prior	Number of sworn law enforcement officers divided by national forest area	By national forest	USDA Forest Service (2013)
Hispanic male population density[c]	Number of Hispanic males age 15–39 in a county, divided by county area	By county	US Bureau of the Census (2013a), 2010 Census estimates
Total population density[c]	Total county population, divided by county area	By county	US Bureau of the Census (2013b), 2010 Census estimates
Poverty rate[c]	Total county population living in poverty (%)	By county	US Bureau of the Census (2013c), 2010 Census estimates
Cannabis price[e]	Wholesale price (current $/lb)	By county	Western States Information Network (2004, 2006, 2008, 2010, 2012)
Gross state product[f]	All industry total economic output of the state (current $ million)	By state	US Bureau of Economic Analysis (2014a)
Gross domestic product deflator[g]	All sector total economic output (chained current $ billion, 2005 base)	National	US Bureau of Economic Analysis (2014b)

[a] Squared elevation and slope values were also included as candidate variables.
[b] Original aspect values were subjected to a cosine transformation.
[c] As reported for the observation year.
[d] Used to deflate retail wage rates and cannabis prices in the logit and probit models with price.
[e] Implemented in two of four models using a control function approach.
[f] Used in the imputation equations for the logit and probit models with price.
[g] Used to deflate cannabis prices in the logit and probit models with price, and gross state products to constant dollar levels in the price imputation equations.

enforcement density and number of backcountry visitor days, which we used to develop weighting schemes for probit model estimation (see Appendix A). Law enforcement density represented the impact of purposeful grow discovery on our sample, while the number of backcountry visitor days accounted for the impact of accidental discovery. The visitation measure was only applied to develop probit weights, and not to estimate the main logit or probit models, so it is not listed in Table 1. We omitted the visitation measure from the main models for technical reasons. Like law enforcement density (as described in Appendix A), visitation could be an endogenous factor in grow occurrence: visitors might avoid going to national forests (or parts of national forests) where cannabis grows have been discovered recently (i.e., grow establishment causes visitation changes), and growers may prefer places with lower visitor use (visitation affects grow establishment likelihood). Indeed, visitors may even be warned to avoid areas with recent grows by forest personnel. However, we could not specify a statistical model that included this possibly fine-scale spatiotemporal dynamic of grow siting and visitation using the National Visitor Use Monitoring (NVUM) data that served as our measure of visitor activity. The NVUM data (see Appendix A) were collected over a five-year window and applicable to entire national forests, so we could not develop temporally lagged measures for them as we were able to do with law enforcement density, for which we had annual data.

We estimated two logit and two weighted probit models – in each case, models both with and without price – using an initial randomly selected training data set of 60% of all observations, using all of the candidate variables (except price) listed in Table 1. Each model was parameterized uniquely through a selection process in which we typically dropped variables with p-values < 0.10. To evaluate the robustness of the selected models, we calculated and compared in-sample fit statistics (based on the 60% training sample of observations, sampled without replacement) and out-of-sample fit statistics (using the remaining 40% of observations). We then re-estimated the parameters for each selected model and evaluated in-sample and out-of-sample performance using nine additional random samples without replacement. Final models, reported in our tables of results and used in mapping, were estimated with 100% of observations.

2.4. Likelihood Maps

We used the four fitted models to create raster (gridded) maps of our study area, in which each map location (i.e., a map cell, as represented by its centroid point) in the area had an estimated likelihood of being a cannabis grow location. The maps had a spatial resolution of 250 m and were identical in extent. We first used GIS software (ESRI ArcGIS 10.1) to populate all map locations (cells) in the study area with values for the explanatory variables listed in Table 1. For the logit and probit models that included a control function for cannabis price, we used the corresponding linear regression equations (Tables B.1 and B.2 in Appendix B, respectively) to estimate a predicted cannabis price for each map location, which we used in place of the imputed price and control function residuals. We applied each fitted model equation to calculate output likelihoods for all study area locations based on the explanatory variables. Note that all map locations that fell outside national forest boundaries were omitted from the output likelihood maps (as well as the multi-attribute frontier map described below).

Because we did not know whether the distribution underlying our sample was logistic or normal, we could not determine which model form, logit or probit, would provide the most reliable output probabilities. Although model performance may be compared via validation, this could provide an inaccurate answer if – as was likely in our case – the available validation data (i.e., the hold-out data left after selecting the training sample) are subject to selection bias. Furthermore, while various goodness-of-fit measures have been proposed for logit and probit models, such as the pseudo-R^2 formulations detailed by Hagle and Mitchell (1992), they must also be interpreted cautiously since they are influenced by underlying distributional characteristics. Given our uncertainty about the most appropriate model, we used a Pareto frontier approach to construct a consensus model. The approach, also known as a multi-attribute frontier approach (Yemshanov et al., 2013), is described in Appendix A. A key advantage of the approach is that, unlike other commonly used combinatorial methods (e.g., linear weighted averaging), it requires no judgment about which component model is "best". Rather, when applied in a mapping context, each map location is ranked objectively against the other locations according to its probability (or likelihood) values from all of the component models. The output from this approach was a single map of likelihood rank values, scaled between 0 and 1.

3. Results

3.1. Predictive Equations

Tables 2 and 3 summarize, respectively, the logit equations excluding and including the imputed cannabis price and the residuals of the control function. The logit model without price as a predictor (Table 2) was highly significant against a null model (p-value < 0.001). (No equivalent test of model significance was available for either the logit or probit model that included a control function for price.) In both logit models, nearly all explanatory variables were highly significantly different from zero (p-values < 0.001), the only exception being average July precipitation in the logit model with the control function for price (Table 3). The variables in the control function for cannabis price (Table B.1 in Appendix B) had p-values that were usually <0.05. Latitude, which served as the price instrument in the function, was highly significant (p-value < 0.001) and negatively signed, indicating a generally falling price going north in the study area.

Notably, the intercept estimates shown in Tables 2 and 3 have not been corrected for bias arising from the response-based sample. We were unable to correct the intercepts because we did not know the nature of this bias, i.e., we did not know how the response-based sample differed from the population (of all locations). Unfortunately, our inability to correct the intercepts prevented us from estimating true grow probabilities, which rely on the intercept values. However, with the logistic distribution, the rank, in terms of probability or likelihood of grow occurrence for all sample observations (and therefore also individuals in the population), can be ordered from lowest to highest; this rank ordering is preserved despite the intercept bias, and so may be used in creating output maps (see Section 3.2).

Setting aside the intercept issue, the parameter estimates for both logit models suggest some trends with respect to cannabis cultivation

Table 2
Parameter estimates for the logit model without price.

Variable	Coefficient	SE	p > |t|
Aspect	−0.352	0.049	0.000
Percent slope	0.0161	0.0065	0.013
Distance to nearest water feature	−0.00169	0.00026	0.000
Unemployment rate	−0.1608	0.0229	0.000
Elevation ∗ elevation	−9.18E−08	3.85E−09	0.000
Average July precipitation	−0.000287	0.000076	0.000
Slope ∗ slope	−0.000232	0.000076	0.002
Poverty rate	0.100	0.024	0.000
Law enforcement density 1 year prior	−1.31	0.22	0.000
Law enforcement density 2 years prior	2.07	0.22	0.000
State = Oregon (dummy variable)	−2.95	0.35	0.000
State = Washington (dummy variable)	−2.68	0.46	0.000
Year	0.181	0.028	0.000
Intercept	−353	57	0.000
Observations	11,578		
Random effects parameter estimate (county FIPS)	1.14	0.25	0.00
Wald χ^2 statistic (13 d.f.)	1095.64		0.00

Note: "SE" is standard error.

Table 3
Parameter estimates for the logit model with price.

Variable	Coefficient	SE	p > |t|
Aspect	−0.346	0.056	0.000
Percent slope	−0.0204	0.0031	0.000
Distance to nearest water feature	−0.00196	0.00026	0.000
Unemployment rate	−0.342	0.035	0.000
Elevation ∗ elevation	−6.60E−08	4.67E−09	0.000
Average July precipitation	−1.46E−04	9.00E−05	0.104
Retail wage rate	−0.0236	0.0037	0.000
Hispanic male population density	−113	17	0.000
Total population density	13.4	1.9	0.000
Law enforcement density 1 year prior	−4.15	0.46	0.000
Law enforcement density 2 years prior	3.62	0.29	0.000
Year	0.944	0.103	0.000
Price (imputed price)	7.50E−03	1.05E−03	0.000
Price equation residuals (from imputed price)	−8.12E−03	1.07E−03	0.000
Intercept	−1909	211	0.000
Observations	11,578		
Random effects parameter estimate (county FIPS)	1.39	0.14	

Note: "SE" is standard error.

on national forests. For instance, although very flat (i.e., close to zero slope) sites are not necessarily favored, possibly because irrigation water is most easily transported on sloped land, very steep slopes are also not favored. South-facing aspects are favored, as indicated by a negative sign on this variable (directly south-facing is the most negative value of the cosine of aspect). This finding fits with the idea that south-facing slopes provide a more productive grow environment due to greater sunlight availability. Sites that are closer to rivers or other freshwater sources and that have lower mid-summer precipitation are more favorable. This supports the notion that California has a better climate for growing cannabis than Oregon and Washington (which are both wetter and cloudier), as well as the notion that the preferred sunny and dry grow sites are typically dependent on access to water for irrigation.

The poverty rate of all individuals was included as a predictor in the logit model without price (Table 2), suggesting that opportunity costs associated with being caught are lower in places with higher poverty. In the logit model with price (Table 3), the poverty rate was replaced by the (negatively signed) retail wage rate, similarly suggesting that opportunity costs associated with being caught are lower in low-wage times and locations. Higher unemployment rates appear to be negatively related to growing decisions, after accounting for low-wage conditions, suggesting that places with the weakest labor markets are not viewed as favorable places for cannabis siting. This effect was unexpected based on rational choice connections to opportunity costs. However, it may reflect local demand factors over and above supply (i.e., grow establishment) factors not already indexed by the poverty variable. For instance, low-wealth markets may provide fewer selling opportunities and thus weaker incentives for local production.

Two related variables, Hispanic male population density and total population density, had opposite signs, which was unexpected. We speculate that Hispanic male and total population densities are related to siting in distinct ways that may stem from differential effects of these variables on cannabis prices. Regardless, a hypothesis that Hispanic males are more numerous in places favorable for grow siting (as incapable guardians in the routine activities perspective; see Cohen and Felson, 1979) was rejected. With respect to law enforcement, the one-year lag of law enforcement officer density showed a negative effect while the two-year lag showed a positive effect. Law enforcement activity in the most recent year has a negative influence on grow site establishment, while the two-year lag variable, with a positively signed parameter estimate, may be acting as an additional proxy variable for site favorability.

In the logit model without price (Table 2), grow likelihood was negatively related to a site occurring in Oregon or Washington, which is consistent with a lower frequency of discovered and reported grows in these two states. Furthermore, although price was not explicitly included in the model, the dummy variables for Oregon and Washington may have implicitly captured cannabis price differences between these states and California (Oregon and Washington prices are lower, on average). In contrast, the logit model including price (Table 3) suggests that site establishment is unrelated to the state where cannabis might be grown, after accounting for variation in price. Both logit models exhibited an overall positive time trend in grow likelihood, when other variables were held constant. Finally, the price variable shown in Table 3 was positively related to grow siting likelihood: growers are more likely to cultivate cannabis on national forests at times and in places where cannabis prices are higher. The negative sign on the residuals from the logit model's control function (available from the authors) indicates that the residuals successfully extracted the confounding endogenous component of the relationship between grow site likelihood and price.

As with their logit counterparts, the explanatory variables in the probit equations excluding (Table 4) and including a control function for cannabis price (Table 5) were highly significant. The imputed price control function estimates (Table B.2 in Appendix B), as was the case for the logit control function, demonstrated a highly significant (p-value < 0.001) effect of latitude on prices, and other variables were generally significant with p-values < 0.05. Overall, the individual parameter estimates for both probit models were consistent with the logit model estimates. Notably, young adult Hispanic male population density was positively related to grow site establishment in the probit model with the price control function, while overall population density was apparently unrelated. This suggests that grows are more likely in places with more young adult Hispanic males, contradicting the finding from the logit model with price. As in the logit model estimate, the cannabis price variable had a positive coefficient, indicating a higher propensity for grow establishment where prices are higher, while the negative sign on the residuals from the control function (available from the authors) indicates successful control for the endogenous component of price in the main model. In the probit model with price, we also found that grows are more likely in Washington, even after accounting for price; this is in contrast to the logit model with price, in which state was not a significant determinant (and was dropped during model selection) after accounting for price and other included variables.

Table 6 reports the average predictive ability of the four alternative models in terms of in-sample and out-of-sample performance. With respect to both the logit and probit models, a location was predicted to be a grow site (and labeled as a "1" in Table 6) based on an output probability of 0.5 or greater; any location meeting this threshold was more likely to be predicted to be a grow site (i.e., labeled as a "1") than to not be a grow site (i.e., labeled as a "0"). Notably, the 0.5 threshold maximized the number of correct predictions, which was important

Table 4
Parameter estimates for the probit model without price.

Variable	Coefficient	SE	p > |t|
Aspect	−0.159	0.028	0.000
Distance to nearest water feature	−0.00117	0.00020	0.000
Elevation ∗ elevation	−2.59E−08	4.44E−09	0.000
Average maximum July temperature	0.000529	0.000205	0.010
Average July precipitation	−0.000180	0.000052	0.001
Poverty rate	0.0409	0.0118	0.001
Law enforcement density 1 year prior	−1.12	0.25	0.000
Law enforcement density 2 years prior	1.13	0.26	0.000
State = Oregon (dummy variable)	−0.906	0.153	0.000
State = Washington (dummy variable)	−0.397	0.221	0.073
Intercept	−2.51	1.25	0.04
Observations	11,578		
Wald χ^2 statistic (10 d.f.)	588.71		0.000

Note: "SE" is standard error.

Table 5
Parameter estimates for the probit model with price.

Variable	Coefficient	SE	p > |t|
Aspect	−0.159	0.031	0.000
Distance to nearest water feature	−0.00125	0.00021	0.000
Elevation ∗ elevation	−1.42E−08	4.74E−09	0.003
Average maximum July temperature	0.00101	0.00024	0.000
Average July precipitation	−2.23E−04	7.64E−05	0.003
Slope ∗ slope	−3.53E−05	1.88E−05	0.060
Retail wage rate	−0.00156	0.00083	0.062
Hispanic male population density	1.05	0.48	0.030
Law enforcement density 1 year prior	−1.15	0.23	0.000
Law enforcement density 2 years prior	1.19	0.25	0.000
State = Washington (dummy variable)	0.498	0.178	0.005
Year	0.0674	0.0297	0.024
Price (imputed price)	7.30E−04	2.13E−04	0.001
Price equation residuals (from imputed price)	−8.10E−04	2.21E−04	0.000
Intercept	−139.7	59.8	0.019
Observations	11,578		

Note: "SE" is standard error.

when evaluating out-of-sample performance and comparing in-sample and out-of-sample predictions.

For all four models, out-of-sample performance was nearly identical to in-sample performance. With respect to the individual models, the logit model without price produced many fewer false presences ($C_i = 1$ predicted but $C_i = 0$ observed) and false absences ($C_i = 0$ predicted but $C_i = 1$ observed) than correctly classified presences and absences. By comparison, the logit model with price had even fewer false presences, but a higher proportion of false absences. The lower classification success of the two probit models resulted from the applied weighting scheme, which kept the output probabilities comparatively low, and based on the 0.5 probability threshold, yielded high rates of false absences. Nonetheless, we must reiterate that because we do not know how the response-based sample departed from the true mix of presences and absences in the full population of potential grow sites (i.e., the nature of the bias in the sample), the output probabilities should only be interpreted in a relative sense. Thus, the most appropriate way to employ the models is to rank all potential locations from statistically least to most likely; when used only for ranking, the logit and probit models are relatively consistent, as demonstrated by the output likelihood maps.

Table 6
Predicted vs. actual "confusion" matrices for model predictions, in-sample and out-of-sample, where 0 corresponds to a cannabis grow absence and a 1 to a cannabis grow presence at a sample location. In-sample statistics are averages across ten random 60% training data sets, while out-of-sample statistics are averages across the corresponding 40% validation data sets.

		Actual	Predicted	
			0	1
Logit without price	In-sample	0	5281.5	311.1
		1	400.9	984.2
	Out-of-sample	0	3515.4	216.0
		1	273.0	663.9
Logit with price (endogeneity corrected)	In-sample	0	5348.1	244.5
		1	746.4	638.7
	Out-of-sample	0	3563.8	167.6
		1	506.8	430.1
Probit without price	In-sample	0	5589.6	3.0
		1	1365.9	19.2
	Out-of-sample	0	3729.6	1.8
		1	926.0	10.9
Probit with price (endogeneity corrected)	In-sample	0	5545.7	46.9
		1	1332.2	52.9
	Out-of-sample	0	3700.4	31.0
		1	902.5	34.4

3.2. Output Maps

Fig. 2 shows a subset of each model's likelihood map (and the multi-attribute frontier ranking map; Fig. 2e) for a portion of the study area. Due to law enforcement and public safety concerns, we are unable to disclose the particular area portrayed in the maps. Despite their limited geographic scope, the subset maps illustrate some of the trends suggested by the fitted model parameter estimates (Tables 2–5) with respect to grow site selection. For example, all of the subset maps show comparatively higher likelihood values in areas near rivers or other freshwater sources, which is consistent with the notion that access to water for irrigation is a key requirement for cannabis cultivation. In addition, areas on south-facing slopes also exhibit higher likelihood values, supporting the notion that the greater available sunlight of southern aspects is conducive to cultivation. Unfortunately, the subset maps cannot depict the influence of variables operating at broader spatial scales (e.g., mid-summer precipitation), but they clearly support hypotheses regarding some of the main geographic drivers of grow site selection.

The subset map for the logit model without price (Fig. 2a) is visually distinct from the subset maps for the other models (Fig. 2b–d), each of which exhibits comparatively lower likelihood values at all map locations (i.e., map cells). On the other hand, the subset map for the logit model without price is the most visually similar to the map for the multi-attribute frontier method (Fig. 2e), suggesting that it had the greatest influence on the multi-attribute frontier ranking process (at least for a subset of the study area). The disparity between the probit models and the logit model without price has a straightforward explanation: the maximum output likelihood value (and thus the range of output likelihoods) under each of the probit models (≈0.76 for probit with price, ≈0.73 for probit without price) was much smaller than the maximum value under the logit model without price (≈0.97). The difference between the two logit models is more difficult to explain, but it may stem from the models having slightly different suites of explanatory variables as well as different coefficients for their shared variables.

More importantly, when evaluated in strictly relative terms (i.e., based on how they ranked map locations according to their likelihood of hosting grows), the four models behaved similarly. For instance, within each of the four subset maps (Fig. 2a–d), the highest likelihood values occur along the map's south-western edge, while the lowest values occur in the south-eastern corner as well as along a line running southwest from the center of the map's northern edge. Thus, in the context of determining the relative likelihood of a location being a grow site – rather than the absolute likelihood, which is difficult to verify – any differences between the four models appear to be outweighed by the general similarity of their output maps' spatial patterns.

The output map values (i.e., for the full study area) were not normally distributed, so we calculated Spearman rank correlations (Table 7) between the likelihood maps for the four fitted models as well as the map of multi-attribute frontier ranks. With respect to the fitted models, the maps for the logit model without price and the probit model with price displayed the lowest inter-correlation, but they were all highly correlated ($r > 0.87$), further downplaying the differences between them in terms of their likelihood. Moreover, each model was more highly correlated with the multi-attribute frontier ranking than with any of the other fitted models. These results seem to confirm that the multi-attribute frontier ranking operated as expected: contrary to the visual evidence in Fig. 2, no single "criterion" (i.e., fitted model) was favored over any other. Pairwise scatterplots (Fig. B.1 in Appendix B) reveal other details about the relationships between the output maps. Most notably, the frontier ranking map appears to have a roughly linear relationship with the two logit models, especially the logit model with price. In contrast, the relationship is more curvilinear for the probit models, such that moderate likelihoods under both probit models

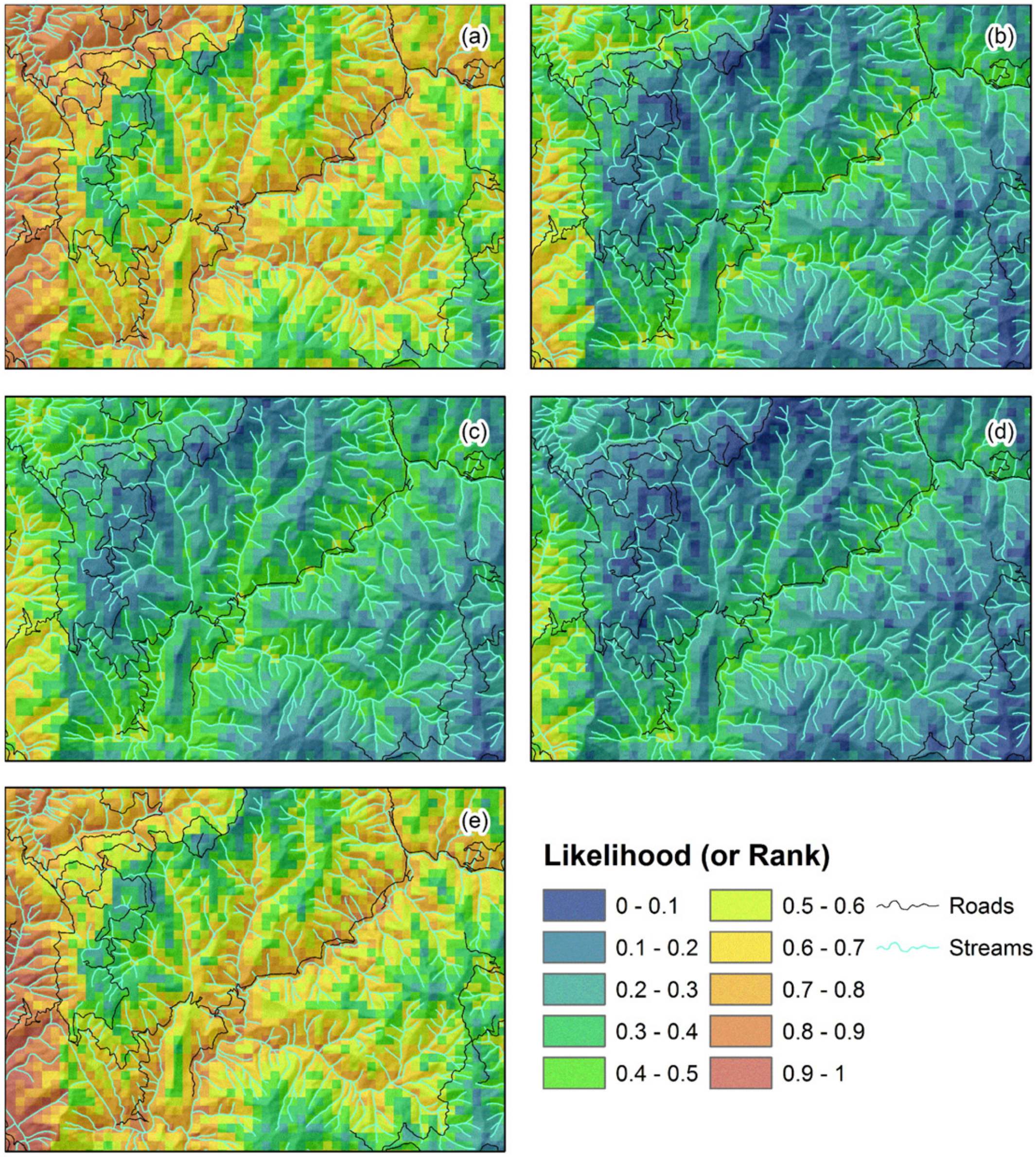

Fig. 2. Output likelihood maps for a portion of the study area: (a) logit model without price; (b) logit model with price; (c) probit model without price; (d) probit model with price; (e) multi-attribute frontier ranking. The frontier rankings have been rescaled between 0 and 1 to make them comparable to the likelihoods estimated with the individual models.

(0.2–0.4) correspond to moderately high ranks (0.6–0.9) under the multi-attribute frontier ranking.

4. Discussion and Conclusions

Outdoor grows on public lands comprise nearly half of all discoveries of cannabis production in the United States, and most of those grows are found on the national forests of the West Coast. In our analyses, we found that cannabis cultivation decisions fit a rational choice explanation of crime occurrences (Cornish and Clarke, 1986). This finding is consistent with conclusions by Bouchard et al. (2013), who examined grow siting in British Columbia (Canada) using a smaller data set. In general, grow sites are likely to be concentrated in locations with higher potential productivity (i.e., south-facing slopes, which have greater light availability, and lower elevations, which provide higher temperatures), and in times and places with higher cannabis prices. Additionally, our findings substantiate the notion that a prospective offender's rational choices relate to the opportunity costs of crime,

Table 7
Spearman rank correlations between the output maps for the four fitted models and the multi-attribute frontier rankings. Correlations are based on values for all locations (map cells) in the study area; locations outside the national forests are excluded.

	Logit without price	Logit with price	Probit without price	Probit with price	Multi-attribute frontier rank
Logit without price	1				
Logit with price	0.9171	1			
Probit without price	0.9488	0.8771	1		
Probit with price	0.8708	0.9103	0.9448	1	
Multi-attribute frontier rank	0.9559	0.9628	0.9647	0.9626	1

wherein worse economic conditions (e.g., lower wages and higher poverty) are linked to lower expected costs of crime commission. Growers also respond negatively to law enforcement pressure, consistent with rational choice theory and with the idea of capable guardians derived from routine activities theory (Cohen and Felson, 1979).

We demonstrated with our modeling that grow siting decisions can be translated into maps of grow site likelihoods (or likelihood ranks) that are based on a theory of criminal decision-making, and not simply based on environmental factors or data documenting previous grow site locations. Although theory-based crime models are common, they are rarely translated into spatial outputs. Instead, the crime hotspotting methods (which involve predictive algorithms that sometimes include information on recent and nearby crimes) used by many law enforcement agencies are generally retrospective in nature, not firmly connected to theories in criminology (Caplan et al., 2011). The typical hotspotting approach can be effective in urban settings, where crimes are often clustered tightly in time and space. Yet, we contend that such hotspotting works less well for crimes such as outdoor cannabis cultivation or wildland arson, which are linked to biophysical variables that are distributed widely across space.

In any event, our maps are practical outputs that can immediately assist decision makers in prioritizing resources to deal with illegal grows. We believe the maps would best be used in conjunction with other information, such as the budget to support interdiction activities or locations where law enforcement resources can be readily deployed. This additional information should influence how the likelihoods (or likelihood ranks) are interpreted and utilized. For instance, given a small overall response budget, locations with very high likelihoods (i.e., that are deemed very likely to be grow sites) would almost certainly be targeted first, but the threshold value at which locations are judged high priority might vary depending on their proximity to available response resources.

Furthermore, the theoretical foundation of the underlying models facilitates adaptive management, which we believe is critical in these circumstances. For instance, our models can illustrate, cartographically, how changes in legal, economic, and socio-demographic conditions, and in law enforcement allocations, may shift the balance between benefits and costs for growers. Alternatively, law enforcement and public land managers could use our models to apply strategies that effectively raise the arrest risk perceptions of growers (e.g., Nagin, 1998), thereby increasing their expected costs. These include targeted use of remote aerial surveillance technologies and the chemical monitoring of waterways to detect upstream grow site locations. Similarly, the models could be used to understand how changes in cannabis market prices, labor markets, and human populations are likely to alter grow siting on public lands.

The models can also be used to explore how the decriminalization and legalization of cannabis production and consumption that is currently occurring in many states, including states studied here, might affect illegal outdoor cultivation on public lands. Some evidence suggests that demand for illicit cannabis will persist in states with legal markets (e.g., Bremner and Del Giudice, 2014; Gurman, 2014). Nevertheless, legalized production has the potential to take market share from illegal producers, putting downward pressure on prices for illicit cannabis (Caulkins and Bond, 2012), and this effect could be modeled and illustrated as likelihood shifts on the landscape. Legalization and decriminalization may also drive up cannabis market demand, putting upward pressure on prices, particularly because of the potential for attracting drug tourism. These price increases can be similarly modeled and translated into shifts in grow likelihood across national forests. The eventual impacts of these legalization and decriminalization processes on illegal cultivation decisions are highly uncertain (e.g., Caulkins et al., 2012), particularly since much of this cultivation is on public lands in states where these legality shifts are still ongoing and notably remain in conflict with federal drug laws. Despite this uncertainty, the rational choice framework underlying our models grants us the flexibility to explore a variety of plausible future scenarios, and in turn, to predict how illegal cannabis cultivation decisions on national forests are likely to change in response to future policy alterations.

Although we strived to include a comprehensive set of explanatory variables in our models, we were forced into difficult choices about some predictors due to data limitations or unclear relationships with grow site likelihood. Foremost, we did not include road proximity in our models because it had an unknown but possibly non-linear relationship with the bias in the response-based sample; as noted earlier, road proximity affects visibility but also the costs of providing supplies to grow sites and transporting harvested cannabis from the sites. However, we included other variables in the models that capture at least part of the variation that would be explained by road proximity (e.g., stream proximity and slope), so the effects of its omission may be muted.

As noted earlier, a five-year reporting window for data on visitor use (i.e., the NVUM data) prevented us from implementing lagged visitation measures to address possible endogeneity of visitor use and grow discovery. Moreover, visitor use data were not available at fine spatial scales, which would be desirable for model estimation; growers probably seek sites within forests that have lower visibility or discovery likelihood. We also had concerns about spatial scale with respect to several of the socioeconomic variables. These variables (e.g., unemployment rate) were usually estimated from data reported at the county level, and we felt this was an appropriate analytical scale. Nonetheless, a potential grow site might not only be affected by socioeconomic conditions in the county in which it is located, but in neighboring counties as well, especially if the counties are part of a large metropolitan area. An alternative approach would be to describe the socioeconomic variables more smoothly, for instance by weighting their values by inverse distance from a location of interest. Without justification for a more sophisticated weighting scheme, we instead opted to assign full weight to the county containing a potential grow site, and zero weights to all other counties. Clearly, additional research is needed to refine approaches for relating illegal cannabis grows to socioeconomic conditions based on crime theory.

Finally, there were some potentially influential factors for which we could not account, simply because of a lack of data. For example, competition and conflict between regional drug trafficking organizations (DTOs) are likely to influence grow site decisions, but verifiable information on such interactions is not usually available to analysts. The same can be said about differing perceptions among growers of their likelihood of being caught. We contend that these factors are idiosyncratic and therefore included in the random prediction error. Moreover, the diverse set of variables that we chose, motivated by crime theory and recommendations from law enforcement, may have provided some ability to account for the kinds of heuristic biases and not-completely-rational decision-making associated with complex decisions, particularly those involving high-consequence but low-probability events (e.g., grow discovery and arrest). We are optimistic that our models accounted for such biases in producing parameter estimates and thus a grow likelihood for each location. Still, we recognize that better fitting models might be found if more work were done. Ideas might be available from theories in economics, statistics, psychology, and sociology that could be incorporated into future modeling of cannabis grow site locations.

Sidestepping these study-specific issues, our modeling approach is readily adaptable to problems other than illegal cannabis cultivation. A rational choice framework may not be applicable when analyzing, for example, the human-assisted spread of a new and poorly known forest pest. Nevertheless, the other main aspects of our approach (i.e., estimation of both logit and probit models, followed by multi-attribute frontier ranking to combine the outputs) provide a practical way to make predictions in the face of uncertainty about the factors affecting the phenomenon of interest.

Acknowledgements

We thank John Stogner, Kier Klepzig, and Gregory Ruark for helpful comments on an earlier version of this article. We also thank Forest Service Law Enforcement and Investigations (LEI) for providing us with the data that served as the foundation for this work.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ecolecon.2016.06.013.

References

Aaltonen, M., Macdonald, J.M., Martikainen, P., Kivivuori, J., 2013. Examining the generality of the unemployment–crime association. Criminology 51, 561–594. http://dx.doi.org/10.1111/1745-9125.12012.

Akers, R.L., 1990. Rational choice, deterrence, and social learning theory in criminology: the path not taken. J. Crim. Law Crim. 81, 653–676.

Aldrich, J.H., Nelson, F.D., 1984. Linear probability, logit, and probit models. Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 07-045. Sage Publications, Newbury Park, California.

Austin, M., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol. Model. 200, 1–19. http://dx.doi.org/10.1016/j.ecolmodel.2006.07.005.

Barratt, M.J., Bouchard, M., Decorte, T., Frank, V.A., Hakkarainen, P., Lenton, S., Malm, A., Nguyen, H., Potter, G.R., 2012. Understanding global patterns of domestic cannabis cultivation. Drugs Alcohol Today 12, 213–221. http://dx.doi.org/10.1108/17459261211286627.

Becker, G.S., 1968. Crime and punishment: an economic approach. J. Polit. Econ. 76, 169–217.

Bouchard, M., 2007. A capture–recapture model to estimate the size of criminal populations and the risks of detection in a marijuana cultivation industry. J. Quant. Criminol. 23, 221–241. http://dx.doi.org/10.1007/s10940-007-9027-1.

Bouchard, M., Beauregard, E., Kalacska, M., 2013. Journey to grow: linking process to outcome in target site selection for cannabis cultivation. J. Res. Crime Delinq. 50, 33–52. http://dx.doi.org/10.1177/0022427811418001.

Bowers, K.J., Johnson, S.D., Pease, K., 2004. Prospective hotspotting: the future of crime mapping? Brit. J. Criminol. 44, 641–658. http://dx.doi.org/10.1093/bjc/azh036.

Bremner, B., Del Giudice, V., 2014. Legal weed's strange economics in Colorado. Bloomberg Businessweek, 9 January 2014. Retrieved 15 December 2014 (http://www.businessweek.com/articles/2014-01-09/colorado-legal-marijuanas-strange-economics).

Burdett, K., Lagos, R., Wright, R., 2003. Crime, inequality, and unemployment. Am. Econ. Rev. 93, 1764–1777. http://dx.doi.org/10.1257/000282803322655536.

Caplan, J.M., Kennedy, L.W., Miller, J., 2011. Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting. Justice Q. 28, 360–381. http://dx.doi.org/10.1080/07418825.2010.486037.

Caulkins, J.P., Bond, B.M., 2012. Marijuana price gradients: implications for exports and export-generated tax revenue for California after legalization. J. Drug Issues 42, 28–45. http://dx.doi.org/10.1177/0022042612436650.

Caulkins, J.P., Coulson, C.C., Farber, C., Vesely, J.V., 2012. Marijuana legalization: certainty, impossibility, both, or neither? J. Drug Policy Anal. 5, 1–27. http://dx.doi.org/10.1515/1941-2851.1035.

Chavez, D.J., Tynon, J.F., 2000. Triage law enforcement: societal impacts on National Forests in the West. Environ. Manag. 26, 403–407. http://dx.doi.org/10.1007/s002670010097.

Cohen, L., Felson, M., 1979. Social change and crime rate trends: a routine activity approach. Am. Sociol. Rev. 44, 588–608.

Cornish, D.B., Clarke, R.V., 1986. The Reasoning Criminal: Rational Choice Perspectives on Offending. Springer-Verlag, New York.

Cornish, D.B., Clarke, R.V., 1987. Understanding crime displacement: an application of rational choice theory. Criminology 25, 901–916. http://dx.doi.org/10.1111/j.1745-9125.1987.tb00826.x.

Desroches, F.J., 2005. The Crime That Pays: Drug Trafficking and Organized Crime in Canada. Canadian Scholars Press, Toronto.

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697. http://dx.doi.org/10.1146/annurev.ecolsys.110308.120159.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151. http://dx.doi.org/10.1111/j.2006.0906-7590.04596.x.

Gallardo, B., Zieritz, A., Aldridge, D.C., 2015. The importance of the human footprint in shaping the global distribution of terrestrial, freshwater and marine invaders. PLoS One 10, e0125801. http://dx.doi.org/10.1371/journal.pone.0125801.

Gould, E.D., Weinberg, B.A., Mustard, D.B., 2002. Crime rates and local labor market opportunities in the United States: 1979–1997. Rev. Econ. Stat. 84, 45–61. http://dx.doi.org/10.1162/003465302317331919.

Greene, W.H., 1990. Econometric Analysis. MacMillan, New York.

Gurman, S., 2014. Legal, illegal marijuana sales coexist in Colorado. The Spokesman-Review Online, 5 April 2014 (Retrieved 29 September 2014) http://www.spokesman.com/stories/2014/apr/05/legal-illegal-marijuana-sales-coexist-in-colorado/.

Hagle, T.M., Mitchell, G.E., 1992. Goodness-of-fit measures for probit and logit. Am. J. Polit. Sci. 36, 762–784. http://dx.doi.org/10.2307/2111590.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271. http://dx.doi.org/10.2307/1913827.

Hirschi, T., 1986. On the compatibility of rational choice and social control theories of crime. In: Cornish, D.B., Clarke, R.V. (Eds.), The Reasoning Criminal: Rational Choice Perspectives on Offending. Springer-Verlag, New York, pp. 105–118.

Hsieh, D.A., Manski, C.F., McFadden, D., 1985. Estimation of response probabilities from augmented retrospective observations. J. Am. Stat. Assoc. 80, 651–662. http://dx.doi.org/10.1080/01621459.1985.10478165.

Jeffery, C.R., 1977. Crime Prevention through Environmental Design. Sage Publications, Beverly Hills, California.

Johnson, S.D., Bowers, K.J., 2004. The burglary as a clue to the future: the beginnings of prospective hot-spotting. Eur. J. Criminol. 1, 237–255. http://dx.doi.org/10.1177/1477370804041252.

Liddick, D., 2010. The traffic in garbage and hazardous wastes: an overview. Trends Organized Crime 13, 134–146. http://dx.doi.org/10.1007/s12117-009-9089-6.

Manski, C.F., Lerman, S.R., 1977. The estimation of choice probabilities from choice based samples. Econometrica 45, 1977–1988. http://dx.doi.org/10.2307/1914121.

Nagin, D.S., 1998. Criminal deterrence research at the outset of the twenty-first century. In: Tonry, M. (Ed.)Crime and Justice: A Review of Research vol. 23. University of Chicago Press, Chicago, pp. 1–42.

Nagin, D.S., Paternoster, R., 1993. Enduring individual differences in rational choice theories of crime. Law Soc. Rev. 27, 467–496. http://dx.doi.org/10.2307/3054102.

Paternoster, R., Simpson, S., 1996. Sanction threats and appeals to morality: testing a rational choice model of corporate crime. Law Soc. Rev. 30, 549–584. http://dx.doi.org/10.2307/3054128.

PRISM Climate Group, Oregon State University, 2006a. United States average monthly precipitation, 1971–2000. (Retrieved 13 September 2012) http://prism.oregonstate.edu.

PRISM Climate Group, Oregon State University, 2006b. United States average monthly temperature, 1971–2000. (Retrieved 13 September 2012) http://prism.oregonstate.edu.

Ratcliffe, J.H., 2004. The hotspot matrix: a framework for the spatio-temporal targeting of crime reduction. Police Pract. Res. 5, 5–23. http://dx.doi.org/10.180/1561426042000191305.

Reams, G.A., Smith, W.D., Hansen, M.A., Bechtold, W.A., Roesch, F.A., Moisen, G.G., 2005. The Forest Inventory and Analysis sampling frame. In: Bechtold, W.A., Patterson, P.L. (Eds.), The Enhanced Forest Inventory and Analysis Program – National Sampling Design and Estimation Procedures, General Technical Report SRS-80. US Department of Agriculture, Forest Service, Southern Research Station, Asheville, North Carolina, pp. 11–26.

Roberts, D., 2014. Colorado's unregulated marijuana grow sites persist despite legal 'green rush'. The Guardian, 23 May 2014. (Retrieved 28 May 2015) http://www.theguardian.com/world/2014/may/23/colorado-unregulated-marijuana-green-rush.

Rossmo, D.K., 1999. Geographic Profiling. CRC Press, Boca Raton, Florida.

Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, New York.

Tele Atlas North America, 2008. U.S. States [Geospatial Data in Vector Digital Format]. ESRI, Redlands, California.

Tynon, J.F., Chavez, D.J., 2006. Crime in national forests: a call for research. J. For. 104, 154–157.

US Bureau of Economic Analysis, 2014a. Gross domestic product by state (millions of current dollars), levels, all industry total. (Retrieved 2 May 2014) http://www.bea.gov/regional/index.htm.

US Bureau of Economic Analysis, 2014b. Current-dollar and "real" gross domestic product. (Retrieved 20 January 2014) http://www.bea.gov/national/xls/gdplev.xls.

US Bureau of Labor Statistics, 2013a. Weekly wage rate, quarterly average, service providing industry, private sector, all establishments, 2002-2012 (second quarter). (Retrieved 26 January 2013) http://www.bls.gov/data/.

US Bureau of Labor Statistics, 2013b. Consumer price index, all urban consumers, current series. (Retrieved 13 January 2013) http://www.bls.gov/data/.

US Bureau of Labor Statistics, 2013c. Monthly unemployment rate, by county. (Retrieved 23 January 2013) http://data.bls.gov/cgi-bin/dsrv.

US Bureau of the Census, 2013a. Population estimates. (Retrieved 23 January 2013) http://www.census.gov/popest/.

US Bureau of the Census, 2013b. Table 1. Land area, population, and density for states and counties: 1990. (Retrieved 26 January 2013) http://www.census.gov/population/censusdata/90den_stco.txt.

US Bureau of the Census, 2013c. Small area income and poverty estimates, state and county data. (Retrieved 27 January 2013) http://www.census.gov/did/www/saipe/data/statecounty/.

US Department of Justice, National Drug Intelligence Center, 2011r. National Drug Threat Assessment. Product No. 2011-Q0317-001. (Retrieved 27 August 2014) http://www.justice.gov/archive/ndic/pubs44/44849/44849p.pdf.

US Geological Survey, 2013a. National elevation dataset. (Retrieved 11 April 2013) http://ned.usgs.gov.

US Geological Survey, 2013b. National hydrography dataset. (Retrieved 12 April 2013) http://nhd.usgs.gov.

USDA Forest Service, 2008. National forest type dataset. (Retrieved 23 April 2013) http://data.fs.usda.gov/geodata/rastergateway/forest_type/index.php.

USDA Forest Service, 2013. Law enforcement activity on national forests. Unpublished Raw Data, Obtained by Special Request From USDA Forest Service, Law Enforcement and Investigations.

van Gelder, J.-L., de Vries, R.E., 2014. Rational misbehavior? Evaluating an integrated dual-process model of criminal decision making. J. Quant. Criminol. 30, 1–27. http://dx.doi.org/10.1007/s10940-012-9192-8.

Wang, D., Ding, W., Lo, H., Morabito, M., Chen, P., Salazar, J., Stepinski, T., 2013. Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns. Comput. Environ. Urban 39, 93–106. http://dx.doi.org/10.1016/j.compenvurbsys.2013.01.008.

Weisburd, D., Piquero, A.R., 2008. How well do criminologists explain crime? Statistical modeling in published studies. Crime Justice 37, 453–502. http://dx.doi.org/10.1086/524284.

Weisheit, R.A., 2011. Cannabis cultivation in the United States. In: Decorte, T., Potter, G.R., Bouchard, M. (Eds.), World Wide Weed: Global Trends in Cannabis Cultivation and Its Control. Ashgate, Surrey, United Kingdom, pp. 145–162.

Western States Information Network, 2004. Illegal Drug Price & Purity Guide 2004. Western States Information Network, Sacramento, California.

Western States Information Network, 2006. Illegal Drug Price & Purity Guide 2006. Western States Information Network, Sacramento, California.

Western States Information Network, 2008. Illegal Drug Price & Purity Guide 2008. Western States Information Network, Sacramento, California.

Western States Information Network, 2010. Illegal Drug Price & Purity Guide 2010. Western States Information Network, Sacramento, California.

Western States Information Network, 2012. Illegal Drug Price & Purity Guide 2012. Western States Information Network, Sacramento, California.

Xie, Y., Manski, C.F., 1989. The logit model and response-based samples. Sociol. Methods Res. 17, 283–302. http://dx.doi.org/10.1177/0049124189017003003.

Yemshanov, D., Koch, F.H., Ben-Haim, Y., Downing, M., Sapio, F., Siltanen, M., 2013. A new multicriteria risk mapping approach based on a multiattribute frontier concept. Risk Anal. 33, 1694–1709. http://dx.doi.org/10.1111/risa.12013.