Energy Footprint Model (EFM)
Technical Document

Steven L. Garman
USGS Geosciences and Environmental Change Science Center
Denver Federal Center, Denver, CO
2/14/2017

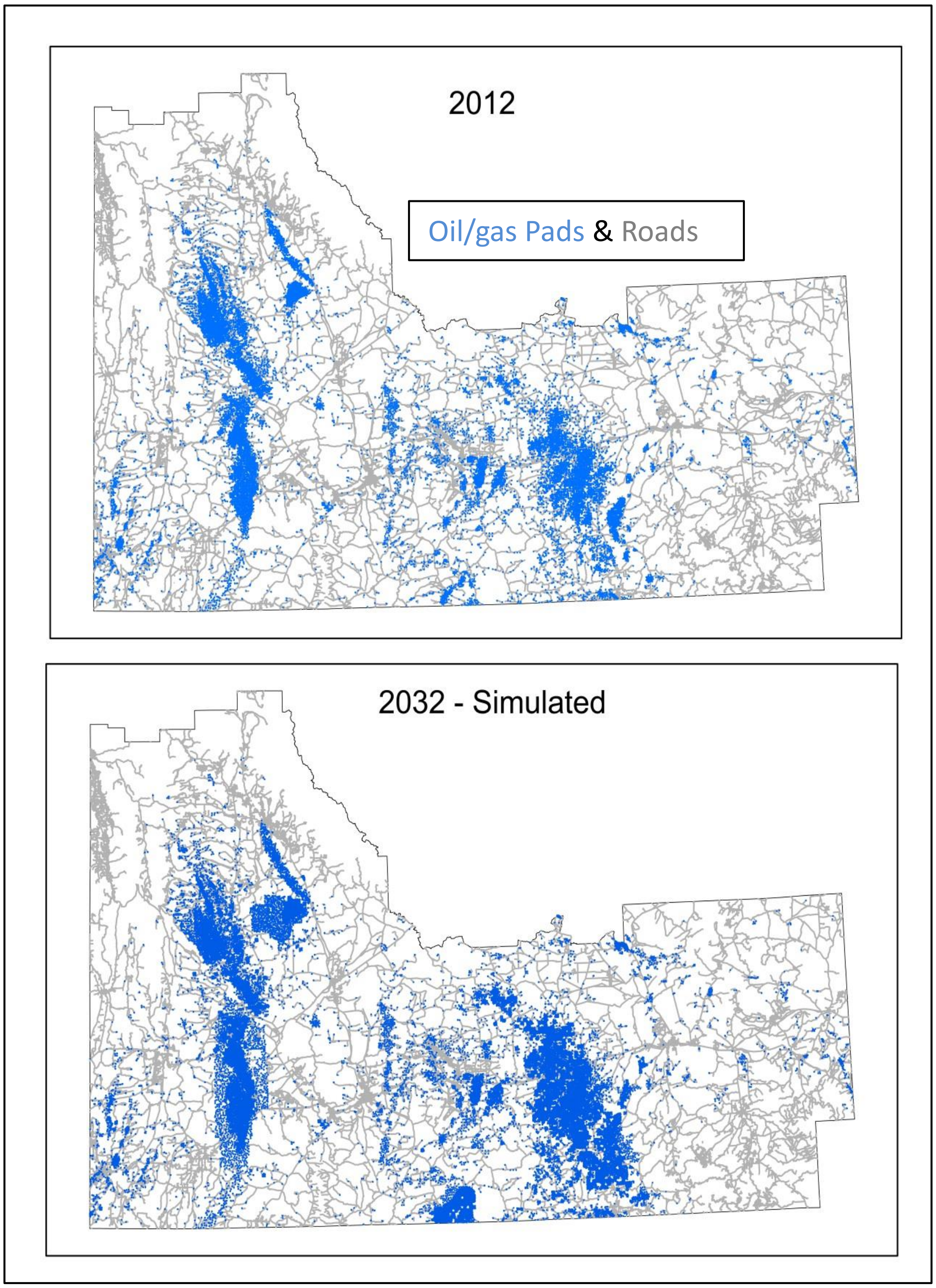

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Conservation management of the sagebrush (*Artemisia spp.)* rangeland in southwestern (SW) Wyoming is topical given the extent of foreseeable oil and gas development even in the face of short-term declines in fossil-fuel prices. The 77,000 km^2 region of SW Wyoming (the 5 southwestern counties) spans the Greater Green River Basin which is the largest onshore federal natural gas reserve in the conterminous U.S. (US Departments of the Interior, Agriculture, and Energy, 2008), contains a significant portion of the remaining intact sagebrush steppe (Knick et al. 2003), and supports some of the largest populations of sagebrush-associated wildlife species (Bowen et al. 2009) and wild ungulates in the U.S. (Sawyer et al. 2005). About 67 % (3.2 MH) of the sagebrush habitat in SW Wyoming overlies federal mineral estates available for energy development (from Homer et al. 2012, http://www.blm.gov/wy/st/en/resources/public_room/gis/datagis/state/state-own.html]. In addition to the 24,000 oil and gas wells drilled since the 1900's (Biewick and Wilson 2014), 7 federal-estate natural gas developments in early implementation or permitting phases propose drilling of 22,000 wells in conventional, coal-bed, and other unconventional oil and gas plays over the next 20+ yrs (US Bureau of Land Management 2015). The removal of habitat by oil and gas infrastructure, behavioral avoidance of infrastructure by wildlife species, and infrastructure-mediated impacts to survival and reproduction due to altered predator (Steenhof et al. 1993) and disease dynamics (Taylor et al. 2013) fragments sagebrush habitat (Walston et al. 2009) and elevates risk to the long-term sustainability of resident wildlife populations (Naugle et al. 2011; Holloran et al. 2010; Doherty et al. 2008).

An interagency effort, called the Wyoming Landscape Conservation Initiative (WLCI), was formed to provide science-based information to support policy decisions, to facilitate strategic landscape-scale conservation planning and long-term monitoring of natural resources, and to provide technical assistance on conservation projects in SW Wyoming (Bowen et al. 2009). Mechanistic studies of the effects of oil/gas development on wildlife and ecosystem properties comprise one part the science efforts of WLCI. To apply mechanistic findings to future possible development, there was a need for a research tool with the ability to forecast oil/gas development under different assumptions of development intensity and technologies. The Energy Footprint Model (EFM) was developed specifically to satisfy this need. The EFM forecasts the footprint of alternative oil/gas build-out designs across the entire WLCI study area. Model results are designed to be used in various assessments of biophysical responses, such as amount of surface disturbance, effects on wildlife species' populations, and wildlife habitat using empirical results of WLCI mechanistic studies and other literature findings. The overall utility of the footprint model is that it can provide a comparison of standard and novel oil/gas build-out designs and ultimately identify plausible development approaches that at least minimize ecological impacts. The collective results from research applications of the EFM thus have the potential to inform Federal land-managers and decision-makers in SW Wyoming who are tasked with promoting energy-development designs with the least amount of environmental consequences.

The EFM is a customized, stand-alone simulation model developed to implement contemporary and novel oil/gas development strategies on the landscape in an as realistic manner as possible. The Energy Footprint Model User's Document describes model operation, inputs, outputs, and system requirements. This document describes the logic, parameterization, and testing of the footprint model.

EFM MODEL

Build-out Designs

The EFM was designed to simulate the establishment of pads and roads on the SW Wyoming landscape based on user-provided build-out designs (Fig. 1). Build-out designs are ASCII files with a specific format (see input argument #17 in EFM User's Document) that designates the annual number of wells to establish, the number of wells per pad, and the duration of a simulation. Also, the bottom-hole spacing requirements of each well must be specified and implies the type of well bore type (vertical, directional, horizontal). Build-out designs are specific to mapped energy-development areas (Fig. 2), referred to as project areas (14,495 km^2 or 19% of SW Wyoming), which are largely concerted efforts by private industry to develop federal and state fluid-mineral leases. Each project area is assigned a unique number (Table 1). Distinct within-project management areas that have specific restrictions or proposed build-out designs are treated as a separate project area. The inter-project areas are assigned a 'project' number of 50 which enables their inclusion in broader-based simulations. The most recent version of the model also uses USGS Oil/Gas Assessment Units (AUs) as a spatial designation in build-out designs to accommodate applications directed specifically at AUs. Assessment Units are large-area delineations of oil/gas reservoirs of different depths that are known or suspected to have extractable fluid-mineral resources (e.g., Fig. 3). There are 29 AUs within the WLCI with many extensively overlapping since each represents reservoirs at different depths. Each AU is also assigned a unique number (Table 2). The numerical combinations of AUs and Project areas are embedded in the spatial maps the model uses to guide development, and allow the user to be very specific spatially about where build-out designs are applied.

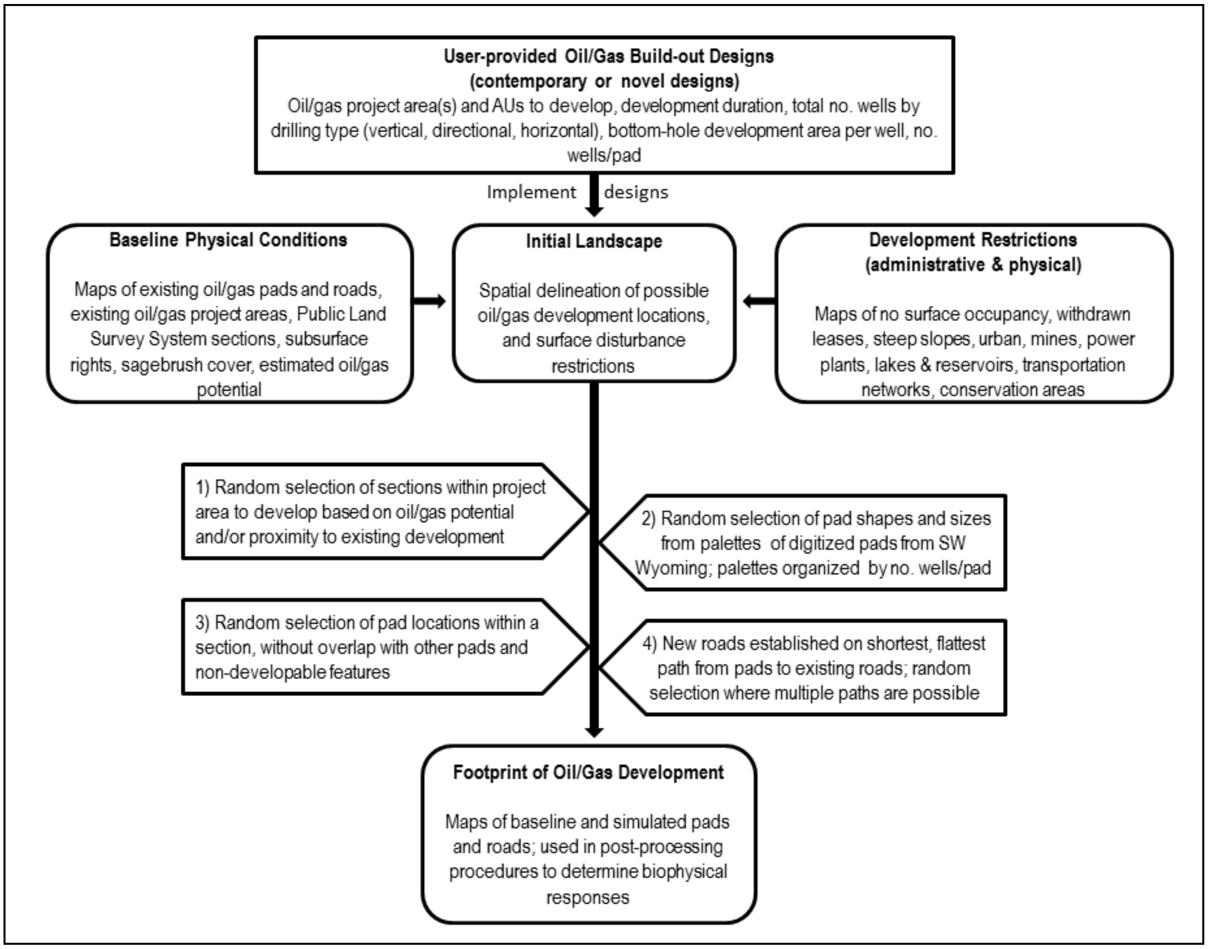

Figure 1. Schematic overview of the SW Wyoming Energy Footprint Model. User-provided build-out designs specify the general location and intensity of oil/gas development starting with the initial landscape which contains current (2012) development patterns and areas where development is prohibited or restricted. Procedures labeled 1-4 are stochastic (random) processes that determine the sections developed within a project area, the location of pads within a section and their shape and size, and the location of new oil/gas roads mediated by development restrictions. These 4 procedures occur on an annual time step until the specified numbers of wells are established. Simulation replicates repeat the implementation of designs starting with a unique random-number seed and the initial landscape. Each replicate results in a map of pads and roads that is used in post-processing procedures to evaluate the effects of infrastructure on biophysical attributes.

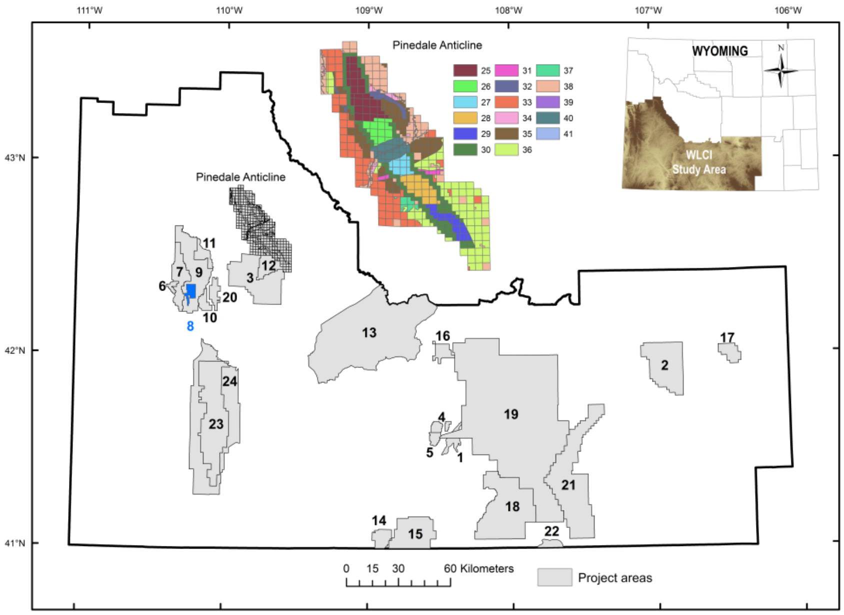

Figure 2. Designated project areas within the WLCI study area. Management units in the Pinedale Anticline project are treated as separate project areas. All project areas are uniquely numbered.

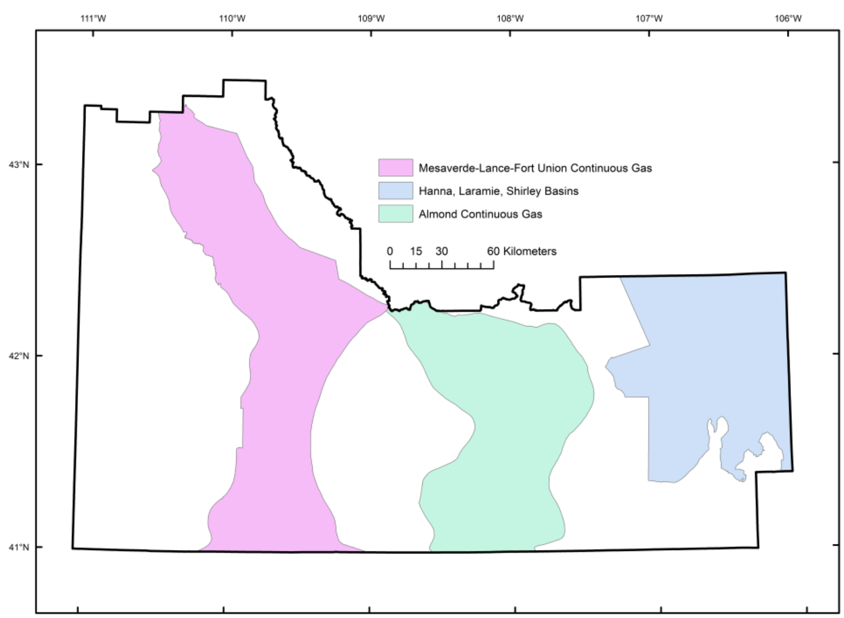

Figure 3. Examples of USGS Oil/Gas Assessment Units (AU) within the WLCI study area.

Table 1. Designated project areas in SW Wyoming (see Fig. 2).

Code	Project name – management unit
1	Table Rock
2	Seminoe Road
3	Normally Pressured Lance
4	Monel Arch – Arch Unit
5	Monel Arch – Monell Unit
6	LaBarge Infill – A
7	LaBarge Infill – B
8	LaBarge Infill – C
9	LaBarge Infill – D
10	LaBarge Infill – E
11	LaBarge Infill – F
12	Jonah Infill
13	Jack Morrow Hills
14	Horseshoe Basin
15	Hiawatha
16	Luman Ridge
17	Hanna Draw
18	Desolation Flats
19	Continental Divide-Creston
20	Bird Canyon
21	Atlantic Rim
22	South Baggs
23	Moxa Core
24	Moxa Infill (flank)
25	Pinedale Anticline - Core Area with Development Areas
26	Pinedale Anticline - Core Area with Development Areas
27	Pinedale Anticline - Core Area with Development Areas
28	Pinedale Anticline - Core Area with Development Areas
29	Pinedale Anticline - Core Area with Development Areas
30	Pinedale Anticline - Potential Development Area
31	Pinedale Anticline - Lander Trail
32	Pinedale Anticline – Mesa Breaks
33	Pinedale Anticline – Unleased Federal Minerals
34	Pinedale Anticline - Sensitive Viewshed
35	Pinedale Anticline - Big Game Winter Range and Greater Sage-Grouse Strutting and Nesting Habitat
36	Pinedale Anticline - Greater Sage-Grouse Strutting and Nesting Habitat
37	Pinedale Anticline - Ross Butte/Blue Rim
38	Pinedale Anticline - Non-Federal Lands
39	Pinedale Anticline - River
40	Pinedale Anticline - River Corridor
41	Pinedale Anticline - Core Area with Development Areas
50	Area outside of designated project areas

Table 2. USGS oil/gas Assessment Units (AU) overlapping SW Wyoming.

Code	AU (non-development designation)
1	Almond Continuous Gas
2	Hilliard-Baxter-Mancos Continuous Gas
3	Fort Union Coalbed Gas
4	Mesaverde Coalbed Gas - Hypothetical
5	Lance-Fort Union Continuous Gas
6	Lewis Continuous Gas
7	Mesaverde Coalbed Gas
8	Mowry Continuous Gas
9	Mesaverde-Lance-Fort Union Continuous Gas
10	Niobrara Continuous Oil
11	Rock Springs-Ericson Continuous Gas
12	Wyoming Thrust Belt Provence
13	Shirley Basin Provence
14	(Wind River Range – no development)
15	(No AU overlap)
16	Tensleep-Casper Conventional Oil and Gas
17	Wasatch-Green River Coalbed Gas
18	Wasatch-Green River Continuous Gas
19	Mesaverde-Lance-Fort Union Conventional Oil and Gas
20	Mesozoic-Cenozoic Conventional Oil and Gas
21	Mowry Conventional Oil and Gas
22	Niobrara Continuous Gas
23	Sub-Cretaceous Conventional Oil and Gas
24	Fort Union Coalbed Gas
25	Hilliard-Baxter-Mancos Conventional Oil and Gas
26	Lance Coalbed Gas
27	Lance-Fort Union Conventional Oil and Gas
28	Lewis Conventional Oil and Gas
29	Mesaverde Conventional Oil and Gas

Build-out designs can be fairly simple, where an AU x Project area combination can have 1 type of well and 1 well-pad configuration (no. wells/pad). Designs can be more complex, however. Multiple well types and well-pad configurations can be specified for the same area. For instance, well-configurations could range from 8/pad to 1/pad for the same. When establishing wells, the remaining bottom-hole area for an AU within a section is evaluated first. Depending on the remaining area, 8 wells may be too many. In such cases, the model goes down the listed order of well-pad configurations to determine which, if any, could be established in the section. This ensures that the AU area within the section is fully developed. Well-pad configurations can be constant but the types of wells may vary; e.g., 2 directional wells/pad may be specified first, followed by 2 vertical wells/pad with different bottom-hole areas of development. Again, the model determines if the directional wells can be established, and if not, if the 2 vertical wells will fit. The sequential order of well-pad configurations or well types used in the build-out design determines the order in which they are evaluated, with the first listed design being evaluated first. There is also a parameter called Annual Probability of Spec Use that indicates the frequency a design specification is considered.   The default is 1, which means the model uses the sequential ordering, as described above, to establish wells. If a specification (e.g., 7 wells/pad) has a probability < 1, the model still sequentially evaluates the ability to implement specifications, but randomly determines if this specification should be considered that simulation year. This is a way to emulate uncertainty in how specifications are implemented, and is of value for certain experimental studies. However, most applications should simply set this value to 1.

Design specifications are time-stamped, meaning the calendar year of a simulation is used to turn the specification on and off.   This allows the user to emulate development that is slated to commence *n* years from the start of the simulation.

Simulating development within an entire AU became a feature in the latest version of the model. For this reason, an approach was developed to accommodate AU build-out specifications without major code changes. As a result, this approach is relatively odd, but functions. This approach creates a synthetic project area within an AU by combining any number of the actual project areas that are within the AU. The project area labeled as #50 (outside of existing project areas) can be included to complete the AU area. A user-provided file details the project areas to merge into a synthetic AU x Project combination (see input argument #18 in EFM User's Document). This file (user-defined name, but currently use the name combo.in) lists the codes of project areas to merge and a unique time-stamp, and the build-out specifications such as bottom-hole area, total and annual number of wells, and start and stop years for the entire AU. The project-level well-pad configurations (which may differ among projects or be the same) are included in the build-out design file, and time-stamped using a unique year such as 1999, which is also specified in the combo.in file. The merge process internally creates the build-out design to be applied to the AU. An additional feature is that the merger may not occur until *n* years after starting a simulation. In the interim, development of the individual project areas may be simulated using project-specific designs. For instance, current production in a project area may be expected to only last *n*-1 years, after which development across the larger AU may be expected to ramp up. Setting the time-stamps of the project-specific designs to 2013 and (2013+n)-1 would terminate project-centric development the year before the development emphasis shifted to an entire AU.

Another feature is the ability to use multiple sets of build-out designs within a simulation. For instance, the collection of ongoing and proposed project-level development approaches in SW Wyoming was stored as a baseline build-out. Although the frequency of proposed directional well technology has increased over recent years, the majority of foreseeable drilling is still vertical. To assess the implications of new drilling technologies, the baseline design was modified to include substantially more directional wells (where deemed possible) and to include substantially more horizontal wells. These 2 modifications comprise the directional and horizontal build-outs, respectively. All 3 build-outs are used in a simulation where the frequency of each are randomly varied over the course of a simulation. Again, a user-provided file (e.g., buildfreq.in) specifies the build-out files and their frequency or probability of annual use (see input argument #19 in EFM User's Document).

The current version of the model emphasizes establishing the total specified production level, which as a surrogate, means the total specified bottom-hole area of wells. For experimental purposes, holding the production level constant but varying numbers and types of well bores allows interesting and useful comparisons of trade-offs in other biophysical properties. For instance, it is obvious that decreasing production will have less effect on biophysical properties. Answering the question, how can production be maintained while decrease impacts, is more topical and practical, and largely the motivating reason for developing the EFM. For each AU x Project combination, the average bottom-hole area of wells (total bottom-hole area to develop/total no. of wells) is first determined then the product of this value and the requested no. of wells per year is derived as the targeted bottom-hole area to develop each year. Due to how the multi well-pad specifications are selected each time step (described above), the specified total no. of wells and the total annual allotment of wells may not actually be achieved; it's the total specified targeted development (total bottom-hole area) that matters the most. The model ensures that the bottom-hole area targets are achieved, and will do so regardless of requested well numbers. Only when you have 1 well-pad configuration are you guaranteed that the model will establish the specified total number and annual allotment of wells. There is an additional consideration. In some cases you may specify a large number of wells and an extensive development area and activate this specification latter in a simulation. Based on the specified annual allotment of wells and thus the average annual development target, there may be insufficient time to achieve total development area. Using the annual average of development, the model calculates the expected total amount for these cases and uses this total to help guide the selection of well-pad configurations to achieve this total by the end of the simulation.

Initial Landscape

Build-out designs are applied to the initial landscape, which due to the availability of key geospatial data layers, is set at 2012. For this reason, the model always start simulations at year 2013. The initial landscape is generated from baseline physical conditions and development restrictions (Fig. 1). Baseline physical conditions consist of infrastructure conditions derived from digitized wells and pads (circa 2012) (Garman and McBeth 2015), and roads (circa 2009) (O'Donnell et al. 2014) [Fig. 4]. Pads include the area with well heads, pumps, and storage facilities, and the surrounding area that was scraped during pad construction. Paved, maintained highways and oil/gas roads are included in the roads layer. Maps of pads and roads are stored as 30-m grids inside the model along with tabular information pertaining to the year they were established, and for pads, numbers of wells and status of each well. Pad grids contain the unique pad ID assigned to baseline pads. Delineating new roads to connect new pads to the existing network is one of the more time-consuming operations in the model. To expedite processing, a variety of information is used for baseline roads, such as the road type, estimates of width (Table 3), and road ID (unique identifier). Additionally, center-line (every 30-m), end-point, and anchor-point (1 point on a road segment within a 30-m grid cell) vertices of road segments are ingested. The 30-m road grids help to quickly locate a road. Internally, a virtual 10-m grid is generated using road vertices and used to locate new road pathways. A map of Public Land Survey System (PLSS) sections is key since development in SW Wyoming is focused on sections. The number of wells allowed within a section is based on the actual section area and the bottom-hole area of wells. Each section is also labeled to identify membership with 1 or more AU x Project-area combinations. Additionally, current leases and subsurface mineral rights (http://www.blm.gov/wy/st/en/resources/public_room/gis/datagis/state/state-own.html) are recorded for each section to determine their availability for oil and gas development. Development within Greater sage-grouse (*Centrocercus urophasianus*) conservation areas is restricted by the amount of disturbance to sagebrush. A map of sagebrush cover (Homer et al. 2012) is used to determine disturbance levels with these areas. Using similar methods described by Copeland et al. (2009), I developed a boosted Maxent model (Phillips and Dudik 2008) of oil/gas potential based on well data up to 2012 (Biewick and Wilson, 2014) and wells that produced for ≥ 20 yrs (n=8759). A 480-iteration model resulted in an AUC = 0.6485. The predicted oil/gas potential estimates are averaged for each section and used to identify areas most likely to have future development (Fig. 5).

Combined with the physical baseline are known restrictions which limit the areas that can be developed (Fig. 1). Restrictions include administrative no-surface occupancy (NSO) designations (http://www.blm.gov/wy/st/en/resources/public_room/gis/datagis/state/NSO.html) which protect areas of special conservation concern (e.g., buffered Greater Sage-grouse leks and raptor nests, historic trails, riparian corridors). Other areas where energy development is prohibited or unlikely (e.g., National Parks, surface mines, population centers, reservoirs, lakes, interstate highways, >25 % slopes) are aggregated with the NSO designations and collectively referred to as the non-traverse layer (Fig. 6). The model prohibits the overlap of new roads and pads with areas in the non-traverse layer. Greater sage-grouse (hereafter sage-grouse) conservation areas, known as core areas (SGCAs) (Fig. 7), are large tracts of land designated to protect the species' breeding (leks) and nesting areas, and have special development restrictions (State of Wyoming 2011). The model establishes pads inside of mapped SGCAs (https://wgfd.wyo.gov/Habitat/Sage-Grouse-Management) in accordance with the restrictions on pad density (1 pad per 259 ha) and spacing (>986 m from a lek), and surface disturbance (≤5 % surface disturbance of sagebrush within a 6.5 km buffer around a new pad and around leks within this buffer). All maps ingested by the model are 30-m grids.

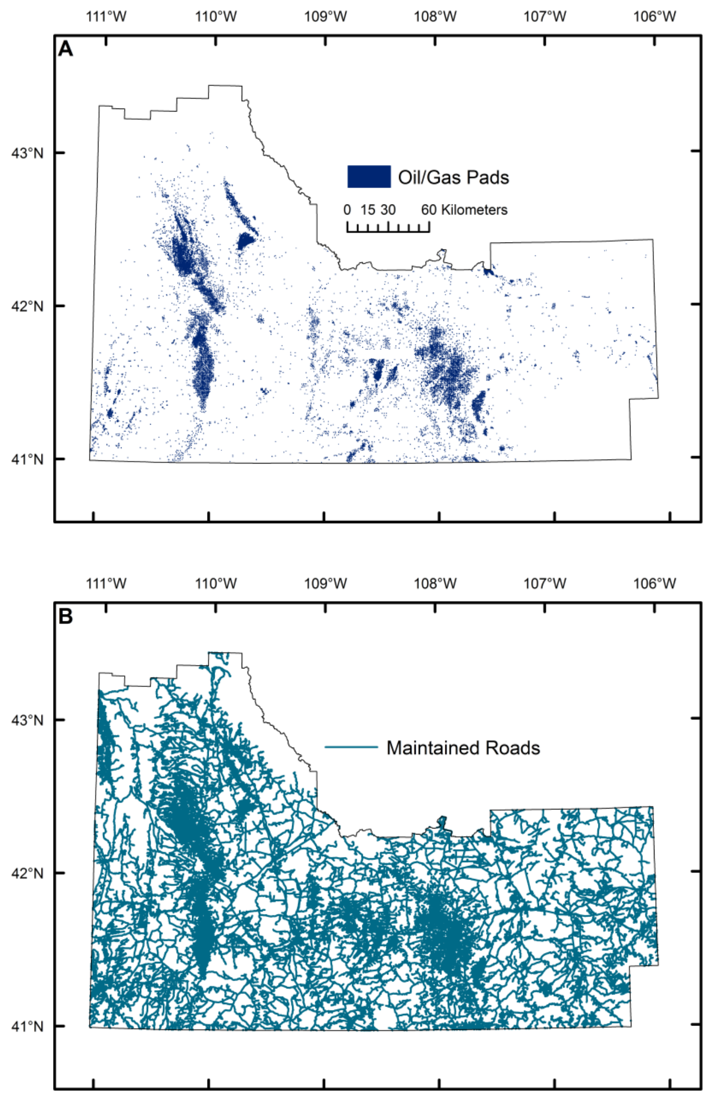

Figure 4. Baseline pads (A) and maintained roads (B) in the WLCI study area.

Table 3. Mean width estimates by road type used in the Energy Footprint Model. Widths were estimated from aerial photography and empirical measures.

Road type	Width (m)
1 - Interstates	80
2 - Divided highways	40
7 - Interchanges	40
3 - Secondary routes	15
4 - Road Class 3	10
5 - Road class 4 (oil/gas)	10

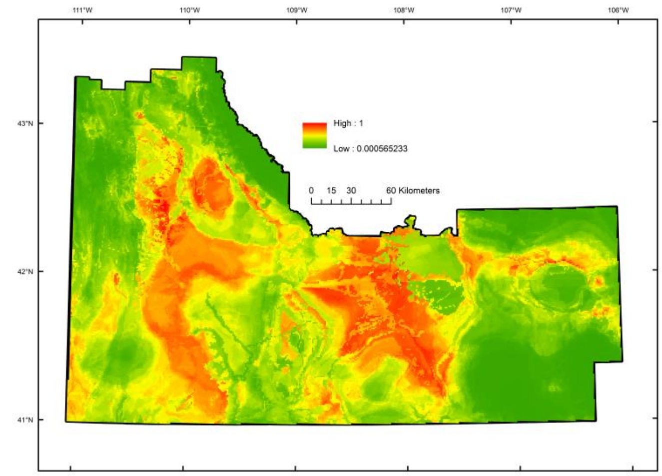

Figure 5. Estimated oil/gas potential in the WLCI study area based on historical patterns of development.

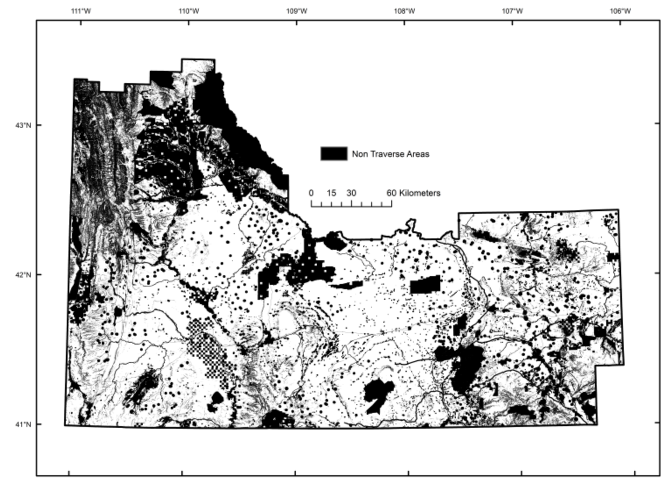

Figure 6. Non-traverse map showing areas off-limits to oil/gas development.

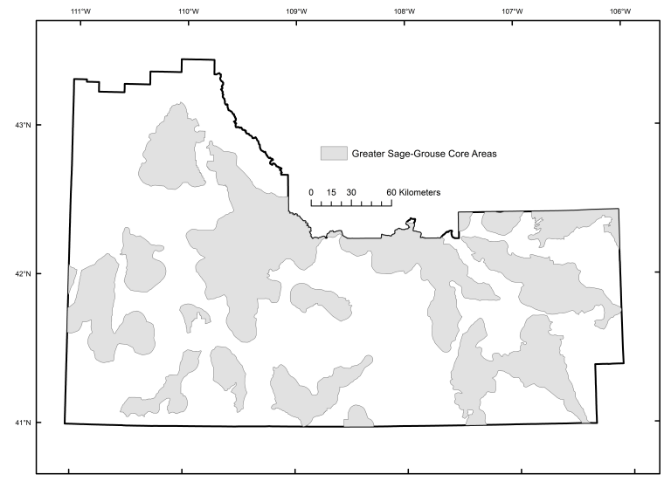

Figure 7. Greater Sage-Grouse core areas in the WLCI study area.

Establishment of Pads and Roads

The model annually develops individual sections within specified project areas over the development period or until the specified number of wells are established. This processing involves 4 basic steps, all of which have a random component (Fig. 1). In each time step, the model cycles through the AU X Project area combinations, sequentially processing all the associated project areas. This cycling continues for the specified duration of a simulation.

Step 1 – Random selection of sections. For each project area the first step is to generate weights for each section and to use these weights to randomly select sections for development. Weighting options include: 1) the probability of oil and gas potential, 2) the proportion of adjacent sections with energy development, 3) the proportion of existing leases and subsurface fluid-mineral rights, 4) a combination of oil/gas potential and neighborhood development, or 5) equal weighting among all sections. The user can choose which option to use in a simulation. Additionally, there is an option to randomly vary annually among the first 4 methods.

> The first method translates the oil/gas potential values originally from 0 to 1.0 to values scaled by an exponential model which produce values from 0 to 5.35. Scaled values of each section available for development in a project area are summed and used to develop a cumulative frequency distribution from 0.0 to 1.0 with discrete intervals of the distribution assigned to each section based on its relative contribution to the distribution. A random draw from the distribution determines the section to be immediately developed following steps 2-4 (Fig. 1). Random selection of subsequent sections continues until the annual allotment of wells has been established.
>
> The second weighting option scores each section according to the number of its 8 neighbors with development; this emulates the clustering patterns that often develop following initial discoveries of productive oil/gas plays. Each section is scored by the number of neighboring sections (0 – 8) with development. Similar to the oil/gas weighting, the sum of scores is used to create a cumulative frequency distribution where intervals are associated with individual sections, and random draws determine the section to immediately develop.
>
> The third method uses the proportion of a section with existing leases and subsurface oil/gas rights as the section score, and creates a cumulative frequency distribution from which random draws determine the section to immediately develop.

The fourth method uses the first 3 scoring methods to produce the cumulative frequency distribution, and random draws from the distribution determines the section(s) to be developed. This method effectively assumes that the dispersion of development is a complex process that can be motivated by multiple factors.

The fifth method simply assumes that all sections have equal weights, and is a useful approach in certain experimental research applications.

Each of the first 4 weighting methods is logical and has merit. However, only using 1 for an entire simulation assumes wide-spread consistency in how sections may be chosen for development. To account for uncertainty, the user can randomly choose among the 4 methods each time step during a simulation. Currently, selecting weighting option 6 (see input argument #34, EFM User's Document) results in a simple random selection of options 1-4 annually.

Step 2 – Random selection of existing pads or a pad pattern. After selecting a section, there are 2 possible pathways to selecting pads. The user can choose to establish new wells on existing, inactive pads that never had wells in the AU being processed (see input argument #33 in EFM User's Document). Changing the status of a pad is discussed below. This option is implemented as an annual probability, and if > 0.0, randomly activated based on the probability value. If activated, inactive pads are first searched for in each section. If one is found, it may be re-activated and used. A key issue is that if wells of the AU being processed (AU of the new wells) are on the pad and the pad is inactive, it is assumed that the AU reservoir is no long capable of further extraction and is not selected. Currently, a limitation of this option is that pad size is not evaluated to determine if it is large enough to support the well-pad configuration being established. If this option is activated, all possible inactive pads are used then the procedure to create new pads is invoked.

The other pathway is the creation of new pads. This involves the random selection of a pad pattern, defined by shape and size. To enhance the realism of simulations, uniquely numbered pad patterns are selected from palettes of actual digitized pads (Garman and McBeth 2015). Each palette is a distribution of pads sizes organized by the number of wells per pad (1-2, 3-5, 6-8, etc.); however pad sizes and shapes can vary considerably within each distribution (Table 4). Although pad patterns are organized by well density, the user actually specifies the palette distribution to use for each well within the build-out design, allowing the user to determine an appropriate size range. For instance, larger pads are required for a single horizontal well compared to 1 vertical well; thus palette 2 or 3 is more appropriate than palette 1. The number of pads with directional and horizontal wells in the digitized set was limited; thus pads were not organized by well-bore type.

Table 4. Description of the digitized pad palettes used in the Energy Footprint Model.

Palette no.	No. wells/pad	Pad size (ha)		No. of pads	Comment
		Range	Mean (SD)		
1	1-2	>0.4 ≤2.9	1.46 (0.55)	4355	
2	3-5	>1.3 ≤4.4	2.43 (0.76)	356	
3	6-8	>1.8 ≤6.5	3.60 (1.21)	56	
4	9-11	>2.8 ≤7.5	4.79 (1.42)	16	
5	12-15	>4.7 ≤7.6	5.90 (0.74)	16	
6	16-25	>3.9 ≤12.9	6.37 (2.25)	26	
7	>25	>5.2 ≤14.2	8.04 (2.33)	17	
8	64*	7.1-7.9	7.54 (0.27)	24	*About 18 acre pads, to accommodate proposed high-density well configurations in certain project areas in SW Wyoming

There is actually a pair of files that record the dimensions of each pattern. One stores the relative pattern (shape and size) of pads as 30-m grid cells. The width and height of a pattern is delineated as the actual number of rows and columns of the bounding box of the pattern. The rows and columns of the actual pad surface are determined relative to the bounding box and recorded after the dimensions of the bounding box. For instance, a pad pattern may have a bounding box of 4 rows by 5 columns. The 4 and 5 are listed first. The pad surface may be rows 1-3 and columns 1-3, and row 4 column 5. The unique number, and the bounding box and surface area rows and columns are repeated for the entire set of patterns across all palettes and stored in this 1 file. The other file stores the actual UTM (Universal Transverse Mercator) coordinates of the pattern centroid and of the pad perimeter. Similarly, coordinates are in relative units that delineate the direction and magnitude of displacement from the centroid. The unique pattern number, centroid coordinates, and perimeter coordinates for each pattern across all palettes are stored in this 1 file. The number of patterns in each palette is fixed, so the model knows the range of entries for each palette when performing a random selection.

Step 3 – Random pad placement. Pads are randomly located within a section without overlap with the bottom-hole area of existing wells, existing pads, and with non-traverse areas. The random selection relies on GRTS (Generalized Random Tessellation Stratified) points that were generated for every section in pre-processing procedures. The GRTS points encourage spatial balance of locations to avoid clumping and systematic spacing within sections which rarely occurs on the contemporary landscape. A point is randomly selected, then the bottom-hole area associated with the wells being established on the pad is used as a search area around the point. If this area overlaps an existing pad, it is considered to be too close to the bottom-hole area of wells on the pad and another point is selected. This continues until a suitable location is found or all points have been scrutinized. If the latter, then processing backtracks to Step 1 (Fig. 1) to find another section. Otherwise, the 30-m grid version of the pattern is overlaid using the selected point as the pattern centroid. Overlap with an existing pad or with non-traverse areas is determined. If there is overlap, another point is selected. If there is no overlap, the pattern is 'burned' into the landscape by adding it to the dynamic pad-grid layer. In this grid layer, the actual rows and columns of a pad surface contain the unique pad ID (an accession number) assigned to every new pad. This unique pad ID is stored along with the corresponding pad-pattern number, the number of wells, and the year the pad and wells were established in a pad data structure. This structure is later used to generate the output vector versions of baseline and simulated pads. All wells on a new pad are considered drilled and operational the year the pad is established.

Step 4 – Creating roads. Once a pad is established, a road is generated if necessary. Pads are allowed to overlap existing oil/gas roads which emulate what occurs on the landscape. In these cases, it is likely that the oil/gas operator already maintains the road to access other pads, and placing a pad across an existing road affords cost savings. If a new pad overlaps an existing oil/gas road, then no road is generated. Otherwise, there are 2 algorithms used to find a road pathway between a new pad and the existing road network. In each, a new road can't connect to an interstate or an exchange. After generating a new road, grid files are updated with road type and the unique road ID, and anchor points and establishment year are recorded in structures later used to create the road shapefiles.

> The simplest method involves finding the shortest, flattest path. The closest road cell to the new pad is first determined. A straight line is drawn from the anchor point of the road cell to the edge or to the center of a pad. Based on historical development patterns, roads appear to sometimes skirt the edge of pads and continue into the energy field to connect to other pads, while other times a road noticeably is developed across a pad and the terminus is sometimes extended to connect to subsequent pads. To emulate these observations, there is a 25% change that roads skirt the edge of a pad and a 75% they extend onto a pad. The terminus of the straight line is thus first randomly determined then the line is drawn. If the line overlaps non-traverse areas, another road cell is selected and the straight line procedure is repeated. In short, if there is a clear shot from the existing network to the pad, this approach will quickly find it. Once found, elevation change along a 250-m buffer around the straight line is scrutinized to find the flattest pathway, starting at the anchor point of an existing road and progressing towards the pad in 10-m steps (using the virtual road grid). If there are multiple pathways, one is randomly selected. The cells that include the new pathway are updated in the road grids (with a unique road ID), and are flagged for a subsequent vectorization procedure. This procedure uses each triad of cells of the road pathway to determine the orientation of a

road which is then used to convert the pathway into a vector. A triad of cells refers to 3 contiguous cells starting with the anchor-point cell. This cell plus the next 2 containing the pathway are used to determine the direction of the road (north, south, northwest, east, etc.). A line from the anchor point to the 2nd cell of the triad is adjusted within the 2nd cell so that it enters the 3rd cell in the correct direction without a sharp 45- or 90-degree bend. This adjustment effectively involves translating the 10-m cell to 1 x 1 meters and selecting the 1-m grid elements that the road should pass through to achieve the direction into the last cell and look like a real road (a rule set was developed to achieve this look). The UTM coordinates of the initial anchor point and of the selected 1-m grid centers of the second cell are stored as interim center-line vertices. The next step creates another triad using the last 2 cells of the first triad and the next cell of the road pathway. The same operations are performed, then the triad is adjusted , etc. This continues until the pad is reached. Sometimes the pad is overshot, in which case, the line coordinates are adjusted to end at the pads edge or center, whichever approach was selected. The coordinates of this road line is further adjusted using a 3-point smoothing method (3-point sequences are used to derive a mean which is applied to the 2nd point of the sequence) to get rid of kinks that sometimes occur. The smoothed coordinates become the final center-line vertices of the new road and stored for later use in generating the road shapefiles.

The second method for deriving roads is more complex computationally. If a road can't be established with the first method it is often because of large obstructions (large patches of non-traverse areas) such as steep slopes, or large lakes and reservoirs. In other words, a road must detour considerably around these obstructions to connect pads with the existing network. The model exhausts the first method before jumping to this second method which can be very time consumptive. In this method, the closest 30-m road cell to a new pad is first determined (30-m grid is used because of processing speed). Starting at the road cell, an expanding concentric square is used to derive the Euclidean distance from the road cell to every cell in the square. The square expands outward 1 row and 1 column at a time, and the minimum distance from the road cell to every cell within the square is re-calculated and stored as the cell value. This continues until the pad cell is included within the square. The result is a Euclidean distance matrix from the road to the pad. This procedure is repeated starting at the pad and terminating whenever the road cell is within the square, producing a second distance matrix. Where the 2 matrices overlap, the cell values (minimum distances) are summed and stored. The pad and road cell should have the same distance value, otherwise the shortest distance was not derived. If these distance values are equal, then every cell with the same stored distance value as the pad and road cell represents the shortest pathway between the pad and road. There is a check during the expanding square to determine if another road cell is closer than the originally selected location. If so, it becomes the focal cell of interest and the expanding square is adjusted. Because of processing time, roads are not adjusted for shape and form. The center line vertices of cells are stored as the anchor points of new roads and used to create the road shapefiles. Although computational expensive, this algorithm generates roads that can navigate around numerous obstacles.

To expedite overall processing when using the second method, connected areas on the landscape are determined in pre-processing procedures and used in the road generation procedure (see input argument #26 in EFM User's Document). A connected area is where a road can connect to every grid cell without passing over a non-traverse area. In essence, the connected area layer is a 30-m patch map created using the non-traverse map, where each patch is uniquely coded. When locating a new pad, the underlying patch ID must match the underlying patch ID of the initially selected road cell, otherwise it is obvious that the road can't connect to the pad. Road cells are selected until 1 is within the same patch as the pad (same patch ID), then the expanding concentric square operation commences.

The maximum extent of the expanding square is fixed at 14 km. Although this a long way to connect to 1 pad, it affords the opportunity for subsequent development along the length of the road.

Deactivating and Activating Pads and Roads

Annually, wells, pads and roads can become inactive (deactivated), emulating the longevity of well production. Using the pad-scar (Garman and McBeth 2015), and well (Biewick and Wilson 2014) databases for SW Wyoming, estimates of well longevity were derived and used to model deactivation. Newly drilled wells can come up dry and are immediately abandoned. In recent times, the rate of new, dry wells is appreciable small. To emulate at least a few wells being immediately abandoned, the probability of deactivating a well the year it is drilled is set to 0.000000000110205. The annual probability of deactivation of wells after their first year is set to 0.0051778379. After 30 years, the probability is set to 0.1. After 50 and 60 years the probability is set to 0.3 and 0.5, respectively. These values seem relatively low given well age, but older, producing wells do appear in the databases. To ensure wells are eventually deactivated, the probability of deactivation increases to 0.8 for wells ≥ 80 yrs. When a well is deactivated, the simulation year is recorded and stored in data structures that track the wells on a pad.

When all wells on a pad are deactivated, the pad is automatically deactivated. Currently, the model does not emulate restoration of disturbed areas. In SW Wyoming, natural recruitment is perhaps the only way disturbed areas are revegetated to native conditions, and this may take 50+ years. To avoid guessing at restoration efforts and rates, a pad is simply tagged as deactivated. Pad status is recorded in the final shapefiles which enables a user to treat deactivated pads in any manner.

Deactivating roads is difficult given the extent of network connections that must be examined to determine if a road could be used to reach other active pads. When a pad is deactivated, the model determines if there is only 1 road segment leading to the pad and if the road segment has any branching roads or connects to other pads. If there are no branching segments or connections to other pads, the road is deactivated, else it remains active.

If a deactivated pad is re-used, the pad status is changed to reflect this and the status of a deactivated road segment that connects to the pad is also changed.

Creating Shapefiles

Steps 1-4 (Fig. 1) are repeated for the duration of the simulation or until all wells have been established. At the end of a simulation, attributes of road and pads are written to disk for storage. The model, however, summarizes these attributes and calls 2 python scripts that uses these summaries, 2 baseline shapefiles, and the vector version of the pad patterns to generate composite (baseline and simulated) pad and road shapefiles. Shapefiles are time-stamped indicating the year pads and roads were created and when they became inactive. The pad shapefile also contains the total number of wells, the number of active wells, and pad size. One of the road shapefiles is a line file that additionally contains the road type. The other road shapefile is a buffered version of the line format, and includes the buffer width and the surface area of each road segment. These shapefiles are the primary model results and are intended to be used in post-processing procedures to evaluate infrastructure effects on biophysical properties.

MONTE CARLO SIMULATIONS

The Energy Footprint Model was designed for Monte Carlo simulations. Creating scripts to perform numerous simulation experiments and replications are address in the EFM User's Document. In short, input arguments control the number of replications and the sequential replicate numbering of output files (the suffix of output data files and shapefiles includes the replication number). The Monte Carlo replications are controlled by the main module of the EFM, including re-initialization of the numerous data structures that store pad and roads attributes, and of gridded maps. With 1 exception, the input files are ingested only at program initiation; thus re-initiation is relatively rapid. There is one special feature of the re-initialization process. The GRTS points used to randomly locate pads in sections do not change because of the large number of points (total of ca. 2.8 million). GRTS points are sequentially ordered to maintain spatial balance as points are 'used'. In the first run of a simulation, GRTS points are accessed sequentially starting at the first point. During each re-initialization, the starting location is randomly determined, and points are shifted in sequence with wrap-around placing the selected starting point at the top of the list. There can be up to 261 GRTS points in a section, so roughly, it would take 261 replications before duplicate sequences are generated. For the experimental applications conducted to date (see end of document), 10-40 replications per design have been typical.

MODEL TESTING

The Energy Footprint Model (EFM) simulates oil/gas infrastructure patterns based on the specifications of a build-out design. When specifications are constrained (e.g., 1 well/pad) and the development area is very specific (e.g., 10 specific sections within a project area), the model effectively becomes deterministic in that the exact number of specified wells and pads are established in the specified locations. However, when build-out specs include a random component (i.e., variable frequency of individual specifications) and when the area of development within a project is large and unspecified, the model uses random procedures to emulate possible footprint patterns. Although the bottom-hole area of development is the same among replications, the number of wells and pads may vary along with the spatial pattern of the footprint. A question that often arises is how realistic are the simulated patterns generated with these random procedures. This is a difficult question to address for several reasons. The EFM was specifically developed as a research tool to foster tradeoff assessments of unique and otherwise untested designs. When simulating future, novel build-out designs there are no contemporary patterns for a benchmark comparison. Additionally, USGS Energy Team members and members of the Bureau of Land Management (BLM) Oil/Gas Division have insisted that future development locations and patterns can not be predicted with any accuracy, and are likely to substantially differ from historical trends. The random components of the model effectively account for all of this uncertainty. Thus, the EFM provides at least futures that are as plausible as using any other method to forecast future patterns.

Evaluating the ability of the EFM to predict historical patterns can, however, provide insight into the underlying behavior of the model. An issue, though, is how much historical information should be provided to the model when making predictions of the past. For instance, if build-out designs are constructed to emulate the known progression of developed sections and their well densities, model predictions are almost guaranteed to match historical patterns. An alternative is to use as little information as possible to parameterize an historic build-out design, and use general measures of pattern similarity to determine how well model predictions can match historical patterns. This at least illustrates the scale at which model predictions begin to emulate general patterns on the contemporary landscape.

Simulation Design

The Continental Divide-Creston (CDC) project area (4323.5 km^2) and a surrounding 10-km buffer was selected for testing the ability of the EFM to emulate historical patterns (Fig. 8A). In 2012, there were 3,252 pads and 4,227 wells in CDC. About half of the wells was established since 2001 when the recent boom in oil/gas production started. Pads and roads established during and before 2001 formed the baseline conditions (Fig. 8B). Development since 2001 was simulated and compared to the actual infrastructure pattern that occurred during this period.

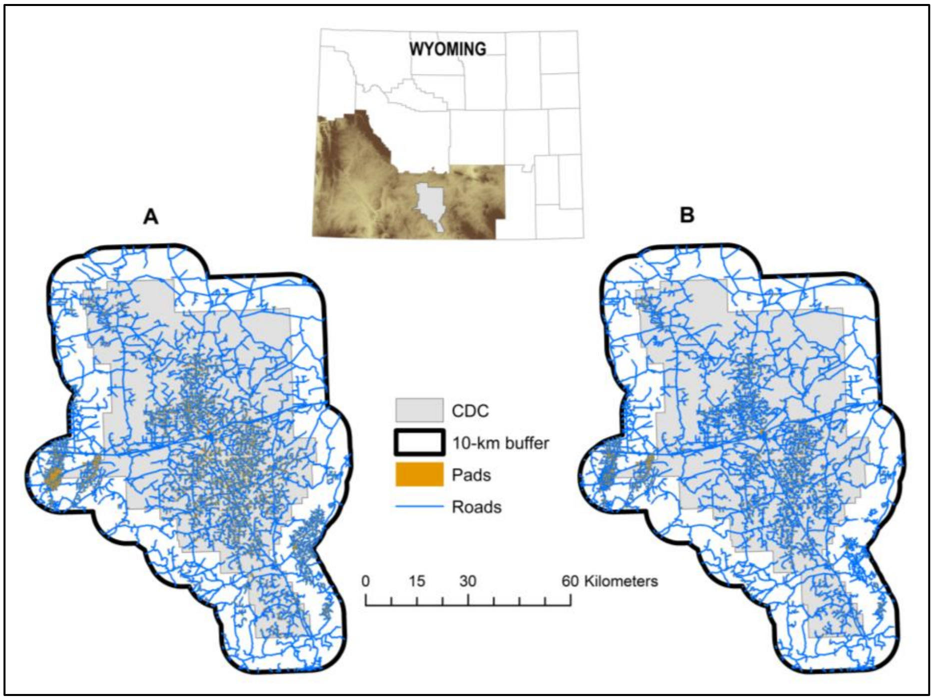

Figure 8. The Continental Divide-Creston (CDC) project area with a 10-km buffer. **A** shows the 2012 infrastructure. **B** shows the 2001 baseline (removal of roads and pads developed after 2001) used in the assessment of the EFM.

Pads are time-tagged and were easily separated into baseline and future conditions. The WLCI road layer (O'Donnell et al. 2014) lacks time-tags of development. Thus, road segments that only connected to pads established after 2001 were identified and manually removed to create the baseline road layer. Of the 13,366 original road segments, 1,719 segments were eliminated. A build-out design was developed to simulate the 1,586 pads and 2,487 wells that historically were established from 2002 to 2012. About 80% of historic well-pad configurations were 1 well/pad, with most of the remaining pads containing 2-4 wells, but 40 pads supported 8-13 wells. Well-pad configurations were constructed to match the frequency of occurrence of historic configurations, and assumed a 40-acre bottom-hole area per well. Starting with the 2001 baseline conditions, infrastructure patterns were simulated over an 11-yr period within the CDC project area. A low level of 'background' development (200 wells using 1 well/pad) was simulated in the buffered area. Twenty simulation replications were performed using the preferred section-selection method which is the randomization of the 4 selection methods (see Step 1 above). Modeled pads and roads were rasterized using a 10-m grain size to facilitate similarity assessments with a 10-m raster version of actual development since 2001.

Assessment Method

A multiple-resolution fit method (Costanza 1989) was used to evaluate similarity between simulated and actual patterns of pads and of roads. Typically, this method applies standard fit procedures over a number of spatial resolutions. The way the fit changes with resolution is used to interpret model performance. As used here, fit is defined as the number of similar responses in 2 separate maps within the same window location. In raster maps, a moving window starts at the size of a grid element (the spatial grain of the maps). At this scale, the fit is either 1 (both maps have the same value) or 0 (maps lack the same value). At larger window sizes, fit between 2 maps is determined by the number of grid cells with the same non-background value. Vertical overlap of grid values is not a requirement at larger window sizes; thus this approach evaluates general spatial patterns not exact patterns. Each window size is moved across the maps to derive a weighted fit metric scaled from 0 (no fit) to 1 (perfect fit). A metric is often generated to indicate overall fit. Here, the fit metrics by window size were simply plotted to determine the scale at which simulated patterns emulated the general patterns of historical development. Fit measures were generated separately for pads and for roads.

Results & Discussion

An example of a simulated pattern and historical infrastructure development from 2002 to 2012 is shown in Fig. 9. Trends in fit metrics with window sizes are shown in Fig. 10. A noticeable difference between the simulation example and historical pattern is the tendency for the simulation to cluster pads and to widely distribute them across the study area (Fig. 9). In part, the clustering may reflect inexact bottom-hole values for simulated wells. A larger bottom area per well would require development of more sections and thus decrease the clumping of pads. The wide dispersion of developed sections results from the lack of spatial specificity of development. Without guidance to concentrate development within such a large project area, the model relies on randomization to select sections. Although there are small clusters of developed sections in the example pattern (the larger clusters of pads), overall the pad clusters have the appearance of being spatially random. Fit and scale relationships were similar for pads and roads although fit was slightly higher for pads (Fig. 10). In general, results indicate that simulation results do not, and probably never can, exactly match historical patterns. For instance, at window sizes ≤ 3 km there was less than 0.5 (50%) fit. Matching broad-scale historical patterns, however, began at about a 6-7 km window size which is about the size of individual energy fields (Fig. 10). At ≥10 km, simulated and actual broad-scale patterns were mostly identical.

Exactly matching historical oil/gas development patterns likely requires detailed information to guide the locations of development, and on well and pad densities. The assessment conducted here lacked specificity of spatial patterns. Using only well and pad numbers, however, the EFM generated broad-scale patterns similar to historical development, at least at the scale of individual energy fields. The slightly lower fit of roads than pads to historical patterns suggests that road design and placement could be improved. Although BLM guidance (major Federal ownership of fluid-mineral rights in SW Wyoming) for new pads is the development of the shortest possible connecting road, oil/gas operators are likely to promote road placement that is the most cost-effect and optimal in terms of servicing the most number of pad. Although the road-development logic of the EFM could be improved, the generally good fit to large-scale historical patterns is acceptable in the context of research studies of future plausible oil/gas patterns.

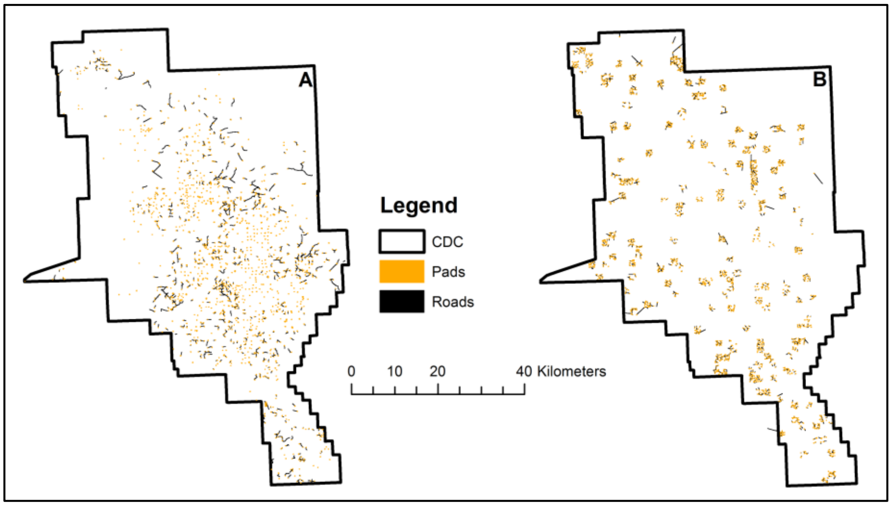

Figure 9. Historic and simulated development in the CDC project area from 2002 to 2012. A is historic pads and roads. B is an example of simulated pads and roads.

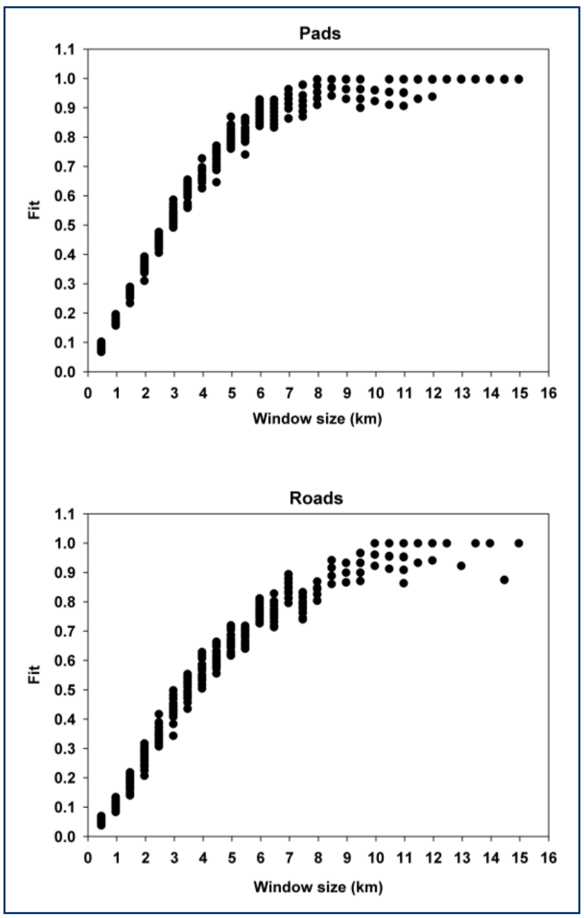

Figure 10. Fit assessment of simulated (n=20) and historical oil/gas pads and roads in the CDC project area from 2002 to 2012. Window size is the edge length of a square analysis window. Fit is determined by the number of 10-m grid cells within a moving window that represented pads or that represented roads in both the simulated and historical maps.

ACKNOWLEDGMENT

Funding for this study was provided by the USGS Land Change Science program and the Wyoming Landscape Conservation Initiative (USGS Ecosystems Mission Area). Any use of trade or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

LITERATURE CITED

Biewick, L. R. H., & Wilson, A. B. (2014). Energy Map of Southwestern Wyoming, Part B—Oil and Gas, Oil Shale, Uranium, and Solar. U.S. Geological Survey Data Series 843. http://pubs.usgs.gov/ds/843. Accessed 2 December 2014.

Bowen, Z. H., Aldridge, C. L., Anderson, P. J., Chong, G. W., Drummond, M. A., **et al.** (2009). U.S. Geological Survey Science Strategy for the Wyoming Landscape Conservation Initiative. U.S. Geological Survey Scientific Investigations Report 2008-5195, 26 p. http://pubs.usgs.gov/sir/2008/5195. Accessed 10 February 2013.

Copeland, H. E., Doherty, K. E., Naugle, D. E., Pocewicz, A., & Kiesecker, J. M. (2009). Mapping oil and gas development potential in the US Intermountain West and estimating impacts to species. *PLoS One*, *4*(10), 1-7.

Costanza, R. (1989). Model goodness of fit: a multiple resolution procedure. *Ecological Modelling*, *47*, 199-215.

Doherty, K. E., Naugle, D. E., Walker, B. L., & Graham, J. M. (2008). Greater sage-grouse winter habitat selection and energy development. *Journal of Wildlife Management, 72*(1), 187-195.

Garman, S. L., & McBeth, J. L. (2015). Digital Representation of oil and natural gas well pad scars in southwest Wyoming-2012 update. **U.S. Geological Survey Data Series 934 [abs.].** http://dx.doi.org/10.3133/ds934**. Accessed 12 September 2014.**

Holloran, M. J., Kaiser, R. C., & Hubert, W. A. (2010). Yearling greater sage-grouse response to energy development in Wyoming. *Journal of Wildlife Management, 74*(1), 65-72.

Homer, C. G., Aldridge, C. L., Meyer, D. K., and Schell, S. J.. (2012). Multi-scale remote sensing sagebrush characterization with regression trees over Wyoming, USA: Laying a foundation for monitoring. International Journal of Applied Earth Observation and Geoinformation 14: 233-244.

Knick, S. T., Dobkin, D. S., Rotenberry, J. T., Schroeder, M. A., Vander Haegen, W. M., & van Riper III, C. (2003). Teetering on the edge or too late? Conservation and research issues for avifuana of sagebrush habitats. *Condor, 105*(4), 611-634.

Naugle, D. E., Doherty, K. E., Walker, B. L., Holloran, M. J., & Copeland, H. E. (2011). Energy development and greater sage-grouse. In S. T. Knick, & J. W. Connelly (Eds.), *Greater Sage-Grouse: Ecology and Conservation of a Landscape Species and Its Habitats* (pp. 489-505). Studies in Avian Biology Series (vol. 38). University of California Press, Berkeley, California.

O'Donnell, M. S., Fancher, T. S., Freeman, A. T., Zeigler, A. E., Bowen, Z. H., & Aldridge, C. L. (2014). Large scale Wyoming transportation data—A resource planning tool. U.S. Geological Survey Data Series 821, 21 p. http://dx.doi.org/10.3133/ds821. Accessed 12 December 2014.

Phillips, S. L., and Dudik, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.

Sawyer, H., Kauffman, M. J., & Nielson, R. M. (2009). Influence of well pad activity on winter habitat selection patterns of mule deer. *Journal of Wildlife Management, 73*(7), 1052-1061.

Sawyer, H., Lindzey, F., & McWhirter, D. (2005). Mule deer and pronghorn migration in western Wyoming. *Wildlife Society Bulletin, 33*(4), 1266-1273.

State of Wyoming. (2011). State of Wyoming Executive Department, Executive Order 2011-5, Greater sage-grouse core area protection. Office of the Governor of Wyoming, Cheyenne, USA.

Steenhof, K., Kochert, M. N., & Roppe, J. A. (1993). Nesting by raptors and common ravens on electrical transmission line towers. *Journal of Wildlife Management, 57*(2), 271-281.

Taylor, R. L., Tack, J. D., Naugle, D. E., & Mills, L. S. (2013). Combined effects of energy development and disease on Greater Sage-grouse. *PLoS One 8*(8), 1-10.

US Bureau of Land Management. (2015). NEPA HOTSHEET, 3 August 2015, BLM Wyoming Land Use Plans and Proposed Projects.

US Departments of the Interior, Agriculture, and Energy. (2008). Inventory of onshore federal oil and natural gas resources and restrictions to their development: phase III inventory- onshore United States. BLM/WO/GI-03/002+3100/REV08. http://www.blm.gov/epca. Accessed 10 May 2013.

Walston, L. J., Cantwell, B. L., & Krummel, J. R. (2009). Quantifying spatiotemporal changes in a sagebrush ecosystem in relation to energy development. *Ecography, 32*(6), 943-952.

EFM RESEARCH & PUBLICATIONS

Garman, S. L. 2015. Forecasting and evaluating patterns of energy development in southwestern Wyoming. WLCI Fact Sheet 7, 2 p. http://pubs.usgs.gov/wlci/fs/7/. IP-059845.

Garman, S. L. in review. A Simulation Framework for Assessing Physical and Wildlife Impacts of Oil and Gas Development Scenarios in Southwestern Wyoming.

Garman, S. L., and S. Germaine. In prep. Effects of alternative oil/gas development in southwestern Wyoming on pygmy rabbits (*Brachylagus idahoensis).*

Garman, S. L., and S. Hawkins. In prep. USGS oil and gas multi-resource assessments: an application using a meta-model of the Energy Footprint Model.

Heinrichs, J, M. O'Donnell, C. Aldridge, S. Garman, C. Homer, and N. Schumaker. In prep. Greater sage-grouse responses to climate change, future energy development in Wyoming.