Energy Footprint Model (EFM)
User's Document

Steven L. Garman
USGS Geosciences and Environmental Change Science Center
Denver Federal Center, Denver, CO
2/24/2017

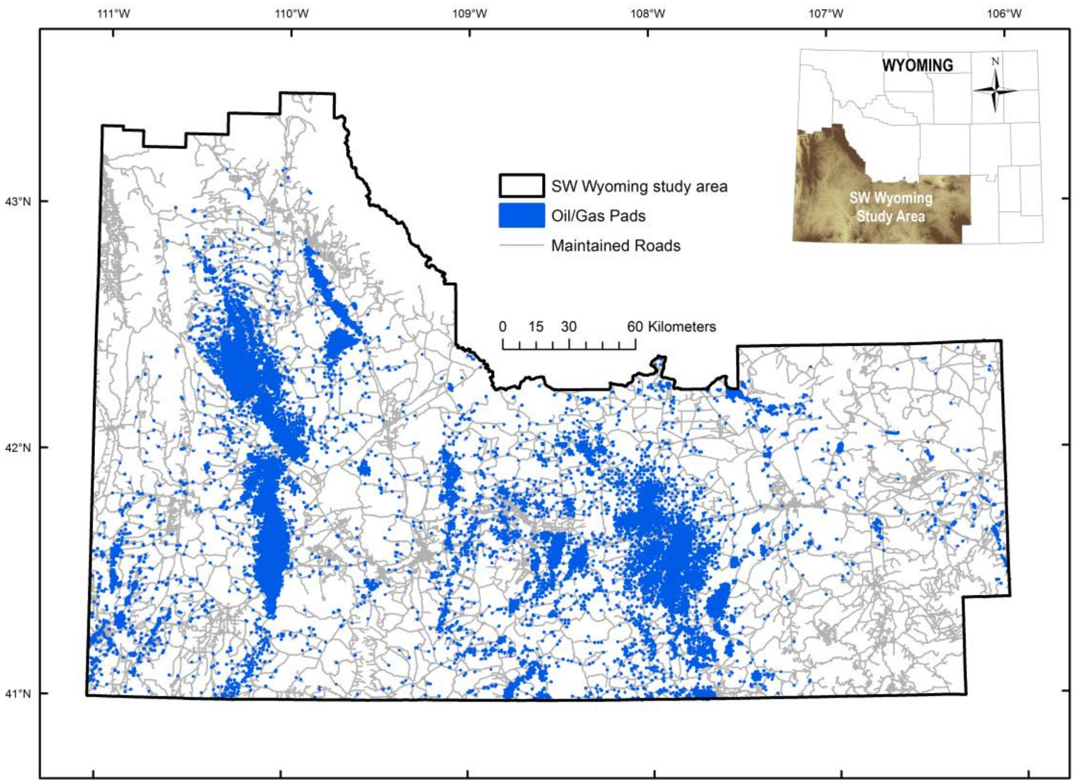

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The Energy Footprint Model (EFM) is a customized, research simulation model developed to forecast the footprint of oil/gas development in southwestern (SW) Wyoming using contemporary or novel energy build-out designs. It uses geospatial data layers representing the baseline landscape to guide the location of future development (e.g., Fig. 1), and outputs maps of baseline and simulated oil/gas pads and roads (e.g., Fig. 2) that can be used to assess and compare effects on biophysical attributes. The Energy Footprint Model Technical Document describes the logic, parameterization, and testing of the model. This document describes how to run the model, the required inputs, model outputs, and system requirements.

MODEL EXECUTION AND INPUTS

The EFM model was specifically designed to run from a command script for efficient processing of a large number of simulations across multiple CPUs. For this reason, there is no GUI interface. To execute EFM.exe, you must navigate to the directory where the program resides, then type EFM followed by the list of required arguments (file names and numerically coded arguments). Typically, a .bat file is used to execute a large number of simulations, especially given the long list of arguments. An example of such a file (cmd.bat) with contemporary input files is shown below. Pathnames of files can be used; otherwise, the files must reside in the directory where EFM.exe is being executed. When using a .bat file, navigate to the directory with EFM.exe and type the name of the.bat file (which needs to reside in the same directory) to run EFM.exe. The sequentially ordered files and argument codes are described in Table 1 and in the subsequent section.

E.g., cmd.bat – the following would be a single line

```
efm rdtype.b rdcenter.in rdendpts.in rdid.b padgridpat.in
sectionauproj.in pads.b padinfo.in sagebrush.b nontraverse.b
coreareas.b subsurf.b surfowner.b lekpts.in dem.b ausections.in
buildbas.in none none sectionattr.in sectionpwlist.in
padpatcoords.in grtspts.in  prioritysids.in lekperim.b
areapatchids.b rdverts.in -48 1 50 50 30 0.25 1
```

Four output data files are written to disk, and used to produce the pad and road shapefiles at the end of a simulation replication (described in Outputs). The generation of these shapefiles, however, requires 2 python scripts, 2 static shapefiles, and a static data file, all of which must reside in the directory where EFM.exe is being executed. The required data file, **baserdinfovn5.in**, contains information about baseline roads. The baseline pad shapefile, **basepadv5.***, and baseline road shapefile, **baserdv5.***, are merged with simulated information to produce composite shapefiles of roads and of pads (i.e., time-stamped baseline and simulated oil/gas infrastructure). Python scripts **Dopads.py** and

Dordslb.py produce the pad and road shapefiles, respectively, and are called from inside the footprint model. In the header of each python script is the working and output directories; these must be modified to reflect the directory where EFM.exe is being executed.

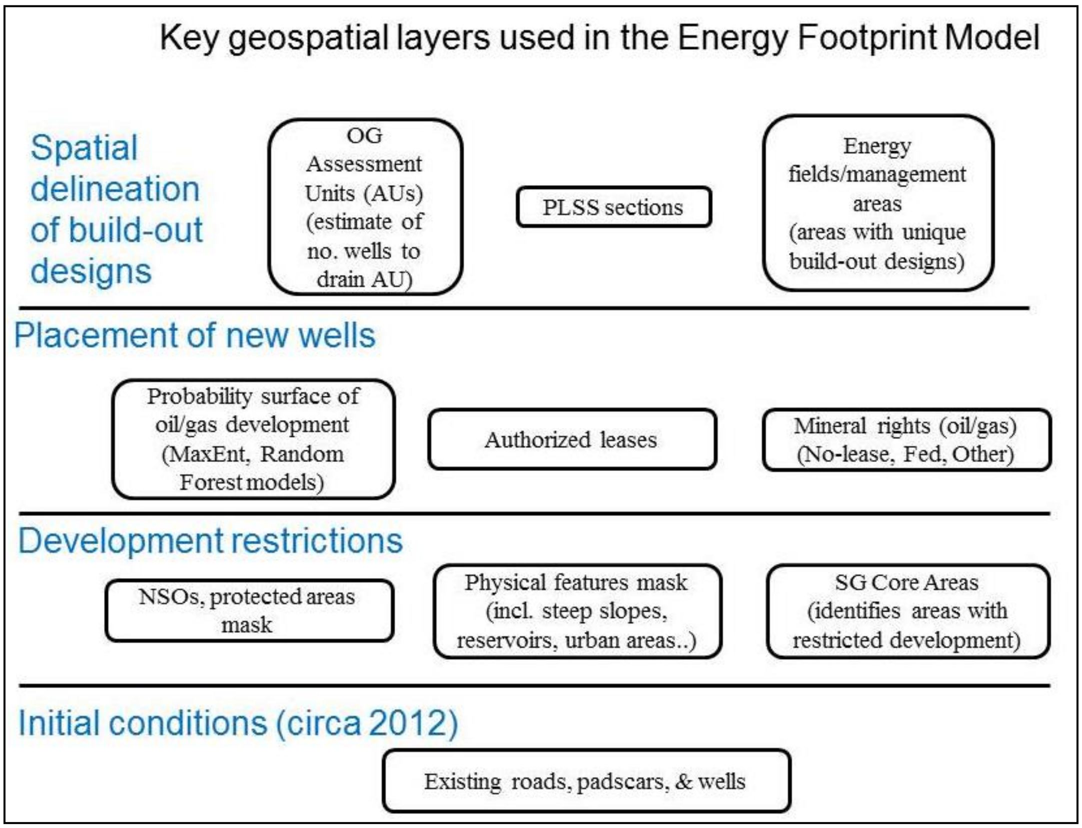

Figure 1. Overview of the geospatial data layers used in the Energy Footprint Model.

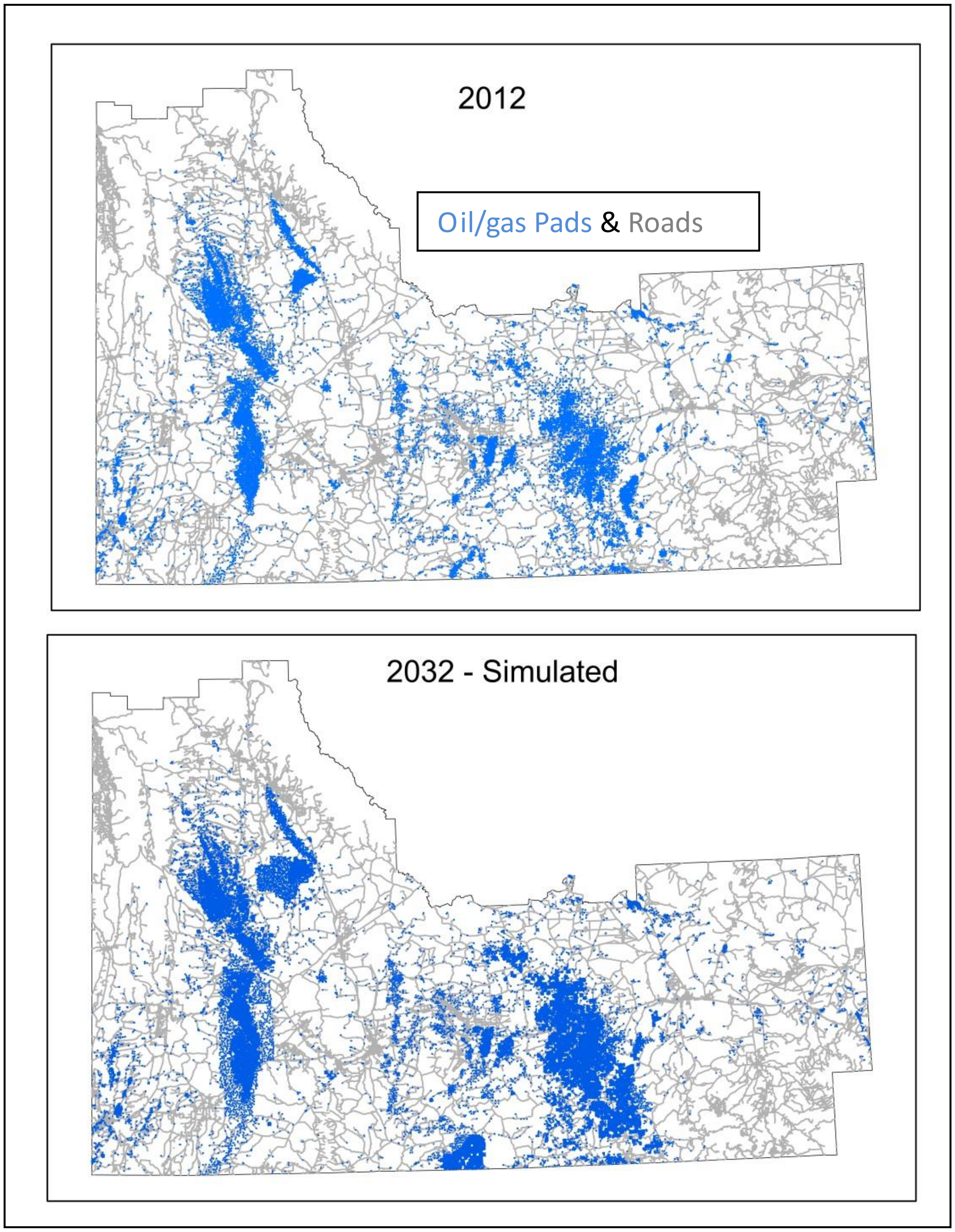

Figure 2. Example of a simulated oil/gas scenario in the SW Wyoming study area. Top map is baseline conditions. Bottom map is a 20-yr simulation of well and pad densities currently being considered for key energy fields.

Acronyms used in the following table and in subsequent sections:

AU – USGS Oil/Gas Assessment Unit

NSO – administrative no-surface occupancy (e.g., a sage-grouse lek)

PID – pad identifier [ID]. Unique number for initial and simulated pads.

Project (area) – A recognized oil/gas project area (oil/gas field) in SW Wyoming

RID – road identifier [ID]. Unique number for initial and simulated roads.

SID – section identifier [ID]. Unique number for every PLSS section.

UTM – Universal Transverse Mercator (all input data are projected to NAD 83, UTM Zone 12).

UTMN, UTME – North, East UTM coordinates

row, col – row and column of internally stored grid maps, respectively. Input maps are 30-m grids georegistered to the SW Wyoming Area of Interest. Each element of the grid is indexed by a row and a column number.

Table 1. Description of the sequentially ordered arguments required by the Energy Footprint Model (EFM). Baseline refers to initial conditions (2012). As a convention, input text files have an '.in' suffix. Some arguments can be 'none', meaning they are not used; none is case sensitive. The structure of files and explanation of coded arguments is provided in the subsequent section.

Arg. #	File name or code	Format	Grain (m)	Description	Values(s)
1	rdtype.b	B (binary)	30	Map of baseline road types	1 – interstate 2 – divided highways 3 – secondary routes 4 – road class 3 5 – road class 4 (oil/gas) 7 – interchanges
2	rdcenter.in	ASCII	Na	UTM coordinates of center-line vertices of baseline roads	Na
3	rdendpts.in	ASCII	Na	UTM coordinates of end points of each baseline road segment	Na
4	rdid.b	B	30	Map of baseline roads indexed by RID	1-94491
5	padgridpat.in	ASCII	30	Rows and columns of digitized pads. Data are based on a 30-m resolution grid	Na
6	sectionauproj.in	ASCII	Na	Pad and well attributes for each section by AU x Project combinations	Na
7	pads.b	B	30	Map of baseline pads, indexed by PID	1-17375
8	padinfo.in	ASCII	Na	Attributes of baseline pads, indexed by PID	Na
9	sagebrush.b	B	30	Map of % cover of sagebrush	Integer: 0-101, where 101 is a mask and set to 0
10	nontraverse.b	B	30	Map of where pads and roads can not occur. NSO stipulations, lek perimeters, protected avian nest sites, steep slopes, human development, water impoundments,..	1
11	coreareas.b	B	30	Map of sage-grouse core areas	1-14 (accession coding)
12	subsurf.b	B	30	Map of subsurface oil/gas rights (masks-out non rights)	0 - none 1 – public 2 – private
13	surfowner.b	B	30	Map of surface ownership	1 – federal 2 – private 3 - state
14	lekpts.in	ASCII	point	UTM coordinates of sage-grouse lek centers	Na
15	dem.b	B	30	Digital elevation model	1711-420 meters

Table 1. Contd.

Arg. #	File name or code	Format	Grain (m)	Description	Values(s)
16	ausections.in	ASCII	Na	No. of sections and a list of sections (SIDs) in each AU x Project combination	Na
17	buildbas.in; emulates future proposed efforts	ASCII	Na	List of the primary build-out parameters for each AU x Project combination of interest	Na
18	*none* or e.g., combo.in	ASCII	Na	A file indicating the combinations of AU x Project build-out designs in #17 to use for a 'new' AU x Project (a synthetic Project Area). Primarily used to combine multiple projects areas to simulate a larger continuous oil/gas AU.	Na
19	*none* or e.g., buildfreq.in	ASCII	Na	A file that indicates the frequency in which different sets of build-out specs are used within a simulation. Specifies the annual probability and also the exact no. of time steps for each spec.	
20	sectionattr.in	ASCII	Na	Attributes of each section, including its 8 neighboring sections, size, mineral rights, and ownership.	Na
21	sectionpwlist.in	ASCII	Na	List of PIDs and number of active wells in each section	Na
22	padpatcoords.in	ASCII	Na	Relativized UTM coordinates of digitized pads in the pad palettes	Na
23	grtspts.in	ASCII	Na	GRTS points (UTM coordinates) for each section.	Na
24	prioritysids.in	ASCII	Na	SIDs of sections that are given higher priority for development. This information is used for experimental purposes which are only accessible via code changes.	Na
25	lekperim.b	B	30	Map of sage-grouse lek perimeters	1
26	areapatchids.b	B	30	A map of landscape areas or patches that can be connected by roads without traversing non-developable areas; used to expedite the delineation of long, convoluted roads	Patch IDs from 1-n.
27	rdverts.in	ASCII	Points	UTM of vertices every 30 m for baseline roads	Na

Table 1. Contd.

Arg. #	File name or code	Format	Grain (m)	Description	Values(s)
28	Any negative integer	Na	Na	Random number seed	Na
29	≥ 1	Na	Na	Starting replication no.	Na
30	≥ 1 but equal to or larger than argument #29	Na	Na	Ending replication no. Start and end rep no. can be any sequence (e.g., 1 2; 5 5; 5 10). The replication number is used as a suffix in the output shapefiles.	Na
31	≥ 1 up to the duration (#32) of a simulation	Na	Na	Frequency in which dump* and roadfo* files are stored on disk. These files are used in generating the output shapefiles.	
32	≥ 1	Na	Na	The duration (number of years) of a simulation replication	Na
33	Fraction from 0 to 1.0	Na	Na	Annual probability of using deactivated pads for the establishment of new wells.	Na
34	1, 2, 3, 4, 5, or 6	Na	Na	Coded section selection methods that use different weightings to derive section-selection probabilities	Na

STRUCTURE OF INPUT FILES

Binary Files

Custom binary files are used for gridded geospatial data layers. The pre-processing program, crinput.exe, transforms an ArcGIS ASCII file to a binary format that contains a header that lists the number of rows and columns, grain size, and upper-left UTM coordinates of the binary file, and a data segment that contains the row, column, and value of non-zero elements of the gridded map. The binary format expedites ingesting the many maps used in the footprint model.   All geospatial information is geo-registered to a master AOI grid to ensure proper alignment before translation to binary formats. The program, crinput.exe, creates binary files; e.g., crinput <inputgrid.asc> < output.b>. Binary files can be translated back to ArcGIS grid files by first using the program binascii.exe to generate a grid ASCII format from a binary file; e.g., binascii <output.b> <newgrid.asc>. This ASCII file is then converted to a grid file using ArcGIS commands.

ASCII Files and Argument Codes

ASCII-formatted input files contain information on baseline conditions.   Most of this information is updated throughout a simulation as new pads and roads are established on the landscape. As part of the Monte Carlo replications, baseline conditions are re-initialized internally, so files are only ingested at program initiation (with 1 exception).

Formats of ASCII files are list below. Argument number is listed (same as in Table 1) followed by the name of the file. Data fields of files are annotated, but it is important to note that there are actually NO column headers in the files, with 1 noted exception. For complex formats, an example of file content is provided. Unless baseline conditions change, files pertaining to baseline roads, pads, sections, AUs and project areas do not need to be modified for future simulations. There are 3 user-provided inputs that can change among simulation studies, depending on the questions being addressed. These include build-out designs (#17), the designation of synthetic AU x Project areas (#18), and the use of multiple designs within a simulation (#19). Also, numerically coded arguments pertaining to the number and duration of replications, data recording frequency, frequency of use of inactive pads, and selection of the section weighting method will vary depending on the questions being addressed.

#2. rdcenter.in - List of the UTM coordinates of the center-line points of baseline roads, indexed by the grid element number (which is the row no. times the column no.).

Grid element (row * col)	UTMN	UTME
…	…	…

#3. rdendpts.in - List of the UTM coordinates of the end points of baseline road segments for each RID .

UTMN	UTME	RID
…	…	…

#5. padgridpat.in - Digitized pads are stored as 30-m grid cells and have a unique number (PatternID). This file provides the 30-m representation of each pad across all palettes (currently 8 palettes). The maximum number of rows and columns bound the extent of a pad. The subsequent list of row and column numbers delineates the actual pad surface, at a 30-m resolution. All row and column numbers are relative; the width of a pad is the number of columns (from 1 to n) and the height is the number of rows (from 1 to n). The actual surface is similarly referenced by relative row and column numbers. When establishing a pad, the model overlays a pad pattern onto the landscape, and records the location of a pad in row and column numbers of the landscape.

No. of palettes			
No. of entries in palette #1			
“ “ “ “ palette #2			
…			
No. of patterns that follow for palette #1			
PatternID	Max. no. of rows	Max. no. of cols (columns)	No. of row-col combinations that follow
Row#	Col#		
…	…		
PatternID	Max. no. of rows	Max. no. of cols	No. of row-col combinations that follow
Row#	Col#		
…	…		
Etc.			
No. of patterns that follow for palette #2			
PatternID	Max. no. of rows	Max. no. of cols	No. of row-col combinations that follow
Row#	Col#		
…	…		
Etc.			

#6. sectionauproj.in - Baseline pad and well information for each section.

Total number of sections								
SID	No. AU x Project combination							
Total # of pads	# of active pads	# of inactive pads	Total # of wells	# active wells	# inactive wells	Total pad area (ha)	Surface area remaining (ha) [developable area minus existing pad area].	
AU #	Project #							
#pads	# active pads	# inactive pads	#wells	# active wells	# inactive wells	Pad area (ha)	Bottom hole area consumed (acres)	Bottom hole area remaining (acres)
Repeat the 2 lines above OR Next SID	**Repeat the 2 lines above** OR Next No. of AU x Project combinations							
Etc.								

#8. padinfo.in – Well and other information for baseline pads.

ID	PID	SID	Ha	Type – G, O, S, POG	Start	Stop	Nwells	Status	Nactive	Ninactive	Fedmin	Surface ownership	AU	Project	N	E
ID	PID	Start	Stop	Status; 1, 0	Production code (PG, PO, PA…)	Formation	Well status (O,G, I)	AU	Project	SID	Bottom hole area (acres)					
…																
…																

The first row is pad #1, where;

ID is a master ID in the original 2012 pad database (not used but retained for reference)
PID is the ID used in the model
SID is the section number containing the pad
HA is pad size in hectares,
Type – Gas (G), oil (O), storage (S), Possible Oil/gas (POG)
Start and Stop are the calendar year when the pad was established and the year when it became inactive (else 0)
Nwells is the no. of wells on the pad
Status is 1 if active else 0
Nactive and Ninactive are no. of active and inactive wells, respectively
Fedmin is mineral rights and is 0 if none, 1 if public, 2 if private
Surface ownership is 1 if federal, 2 if private, 3 if state
AU is the assessment unit no.
Project is the coded no. of the project area
N and E are UTM coordinates of the pad centroid

The second row contains information for a well on the pad (if any), and is repeated for every well on a pad, where;

ID and PID as defined above
Start and Stop are years when the well was established and when it became inactive (else 0)
Status is 1 if active, else 0
Production code; PG - producing gas; PO - producing oil; PA permanently abandoned
Formation is the name of the oil/gas play
Well status is O if oil, G if Gas, I if injector
AU, Project, and SID as defined above
Bottom hole area is an estimate of the area drained by the well (acres)

Following the well information for the first pad, row 1 is repeated for the next pad, etc

#14. lekpts.in - List of the UTM coordinates of every lek point in SW Wyoming.

UTME	UTMN
…	…

#16. ausections.in - List of the sections that occur within an AU x Project combination. This list determines the sections that may be develop for an AU x Project area.

Total no. of AUs
Total no. of Projects
No. of AU x Project combinations (N)
AU #, Project #, No. of sections within AU x Project
… repeat above line for the N combinations
Au#, Project#, no. of sections that follow
SID
SID
…
AU#, Project#, no of sections that follow
SID
Etc.

#17 A build-out design file.

AU #	Project #	Total no. of wells	Total bottom hole area to develop (acres)	No. of wells per year	Bottom hole area per well (acres)	No. of wells per pad	Pad palette no.	Always zero	Annual probability of using this specification	Begin year (calendar yr) when this spec should be used	End year (calendar yr)

e.g.,

AU #	Project #	Total no. of wells	Total bottom hole area to develop (acres)	No. of wells per year	Bottom hole area per well (acres)	No. of wells per pad	Pad palette no.	Always zero	Annual probability of using this specification	Begin year (calendar yr) when this spec should be used	End year (calendar yr) to terminate the use of this spec
9	3	3500	35000	262	10	16	6	0	1	2017	3000
9	3	0	0	0	10	15	6	0	1	2017	3000
21	20	164	6560	61	40	4	2	0	1	2017	3000
21	20	0	0	0	20	3	2	0	1	2017	3000
21	20	0	0	0	60	2	1	0	1	2017	3000
…											

The first entry for each AU x Project combination <u>must</u> indicate the total no. of wells to be established, the total bottom hole area to develop, and the no. of wells to establish each year. Subsequent specifications for the AU x Project combination must specify the AU and Project codes, and all other information to the right of and including Bottom hole area per well; other fields are ignored. In the above example, the AU=9 x Project=3 combination has 2 well-pad options – the first is 16 wells per pad, the second is 15 wells per pad. The AU=21 x Project=20 combination has 3 well-pad options consisting of 4, 3, and 2 wells per pad and that differ in bottom hole area per well. It is best to zero out the fields that are not used to avoid confusion. The Pad Palette number refers to the distribution of digitized pads. Each distribution has a different mean and variance of pad sizes (see EFM Technical Document). The user must specify which distribution to randomly select from for each well-pad specification. Annual probability of using a specification is a feature that allows the user to vary the frequency that a well-pad combination is used. If this value is 1 for all combinations, the model determines if the first listed well-pad combination will fit in a section selected for development. If not, the model sequentially evaluates well-pad combinations in the build-out design until a combination fits. If the annual probability is other than 1, the same sequential processing occurs but the model randomly determines, based on the probability, if a well-pad combination is included in this processing in the current time step. Currently, annual probably ranges from 0 to 1 in 0.1 intervals. The Begin and End year determines the simulation years that a specification is invoked.

#18 – Combinations of existing build-out specifications for a synthetic AU X Project Area; e.g., combo.in

AU #	Project #	No. of Au x Project areas that follow (N)	Total no. of wells	Annual allotment of wells	Total bottom hole area to develop (acres)	Begin year (calendar yr) when this set of specs are implemented	End year – calendar yr when specs are deactivated
AU #	Project #	Begin Yr (calendar yr) of spec	End Yr (calendar yr) of spec				
Repeat above line for N-1 entries							
AU #	Project #	No. of Au x Project areas that follow (N)	Total no. of wells	Annual allotment of wells	Total bottom hole area to develop (acres)	Begin year (calendar yr) when this set of specs are implemented	End year – calendar yr when specs are deactivated
AU #	Project #	Begin Yr (calendar yr) of spec	End Yr (calendar yr) of spec				
Repeat above line for N-1 entries							
…							

e.g.,

2	51	5	9000	60	720000	2022	3000
2	1	1999	1999				
2	16	1999	1999				
2	24	1999	1999				
2	12	1999	1999				
2	50	1999	1999				
11	52	3	13125	60	1050000	2023	3000
11	1	1999	1999				
11	50	1999	1999				
11	2	1999	1999				
…							

A synthetic project area is simply a merger of other project areas (actually the sections of each project area) within an AU, treating all of these project areas as a single, unique project area. This is a way to simulate development in the larger AUs which tend to span numerous project areas and areas outside of designated project areas. These outside areas are actually coded as area 50, so synthetic units can be coded to include non-project areas. In the above example, synthetic Project #51 and #52 are created from well-pad configurations specified in the primary build-out specification file (i.e., within a file like #19 above). Specifically the no. of wells per pad, the bottom hole area of each well, pad palette no., and probability of use (all other attributes are ignored) are acquired from the primary design file and become the design parameters of a synthetic AU X Project area. For AU=2 X Project=51, 5 sets of specifications are to be used to establish a maximum of 9000 wells with an annual allotment of 60, for a total developed area of 720,000 acres starting in 2002 and ending in 3000. The corresponding 5 Au X Project specifications would be listed in the primary build-out design file. Only those specifications with a start and stop year of 1999 would be included in this synthetic AU x Project area. Effectively, this 1999 coding indicates specifications to be used for synthetic areas. These start and stop years are trumped by start and stop years specified for the synthetic AU X Project.

#19 – Combinations of different build-out designs; e.g., buildfreq.in

File name of the primary set of build-out specifications	Probability of annual use	None or name of file designating synthetic project area specifications	No. of time steps to use the first build-out specs	No. of time steps to use the second build-out specs	No. of time steps to use the thrid build-out specs
File name of the 2nd set of build-out specifications	Probability of annual use	None or name of file designating synthetic project area specifications			
File name of the 3rd set of build-out specifications	Probability of annual use	None or name of file designating synthetic project area specifications			

e.g.,

buildbas.in	0.934	combos.in	28	1	1
builddir.in	0.033	combos.in			
buildhor.in	0.033	combos.in			

This file specifies the frequency with which different build-out designs are invoked during a simulation run. This is an experimental way to control the use of different drilling technologies. In the example above, buildbas.in approximates what has been proposed in recent planning documents for future development in SW Wyoming. The designs called buildir.in and buildhor.in are experimental and emphasize the wide-spread use of directional and horizontal wells, respectively. In this experimental set up, buildbas.in is used 93.4% of the time in a simulation, while the others only 3.3% of the time. The model randomly determines the simulation year to invoke each design. Specifying the no. of time steps somewhat duplicates these percentages; however, the combination of the 2 ensures that a design is invoked according to the user-provided frequency before termination of the simulation. Due to coding constraints, there has to be 3 build-out specs specified in this file even if 1 or more are not used.

#20. sectionattr.in – List of information for each section.

Total number of sections										
SID #1	SID of 8 neighbors (8 values)	Lease	Fed0	Fed1	Fed2	Prob	Ha1	Ha3	Nlha	Propnl
Repeat above for every SID										

Lease is percentage (integer) of a section with known sub-surface mineral leases
Fed0 is percentage (integer) of a section with no federal mineral rights
Fed1 and Fed2 are the percentage (integer) of the section with public and private mineral rights, respectively
Prob is the estimated oil/gas potential as a probability
Ha1 is total surface area of a section
Ha3 is the developable surface area of a section (area without NSO restrictions)
Nlha is the area (ha) of withdrawn leases
Propnl is Nlha as a proportion of the section

#21. sectionpwlist.in - List of the baseline pads (PIDs) on each section (SIDs).

Total number of sections	
SID #1	No. of pads (may be zero)
PID	
…	
SID #2	No. of pads
Etc.	

#22. padpatcoords.in – List of the relativized centroid and boundary vertices of each digitized pad (unique PatternIDs).

No. of patterns			
PatternID #1	No. of UTM coordinates that follow	Relativized north UTM coordinate of centroid	Relativized east UTM coordinate of centroid
North displacement	East displacement		
…	…		
PatternID #2	No. of UTM coordinates that follow	Relativized north UTM coordinate of centroid	Relativized east UTM coordinate of centroid
Etc.			

#23 – grtspts.in - GRTS (Generalized Random Tessellation Stratified) points are generated for each section (SID) and used to randomly locate pads. A large sample is generated per section to ensure a degree of spatial balance (the main feature of GRTS points) even if pads are already present.

Total number of sections	
SID #1	No. of GRTS pts that follow
UTMN	UTME
…	…
SID #2	No. of GRTS pts that follow
UTMN	UTME
Etc.	…

#24. prioritysids.in - List of the SIDs designated to receive higher priority. Used for experimental purposes.

SID
…

#27. rdverts.in - UTM coordinates of vertices every 30 m for every baseline road segment.

RID	UTMN	UTME
…	…	…

#34 – Section weighting methods. Six weighting methods are possible:

Code=1. Oil/gas potential was estimated based on historical patterns of development. Using this map, a probability of development is associated with each section. This weighting method relies only on this probability. The probability is scaled from 0 to 5.35 (exponential increase between a probability of 0 and 1). Each section within an AU x Project area that can be developed is assigned a scaled value. The sum of all values is used to develop a cumulative frequency distribution, where unique intervals of the distribution are associated with each section. A random draw from the distribution determines the section to be developed. This continues until the annual allotment of wells has been established. This is repeated each year for each AU x Project area.

Code=2. This method relies only on the development frequency of a section’s 8 neighbors. This approach subscribes to the logic that where there is proven oil/gas reserves, there’s bound to be more nearby. Also, it is common for energy fields to develop in clusters due to the availability of infrastructure. The weighting method counts the number of neighbors with pads and wells, for a maximum of 8. Similarly, the sum of scores is used to develop a cumulative frequency distribution which is used to randomly select sections for development.

Code=3. This method simply uses the extent of known leases and oil/gas subsurface mineral rights. A section is scored using the sum of the proportions of leases and of mineral rights. Similar to other methods, this score is summed and used to create cumulative frequency distributions which in turn are used in the random selection of sections to develop.

Code=4. This combines the scores of the first 3 methods, and similarly randomly selects from the derived cumulative frequency distribution to determine sections for development.

Code=5. This method assumes that all sections have equal weights. Effectively, this is the Null model that you can't really predict where future development will occur. This approach has utility for certain experimental assessments.

Code=6. This code invokes a random selection of methods 1-4 for each year of a simulation. This randomization accounts for uncertainty in how sections maybe selected for development over time.

OUTPUT FILES

Data Files

There are 4 data files written to disk that record the location of simulated pads and roads, and attributes of all pads and roads (baseline and simulated). All files are appended with the replication number and some also contain the output calendar year. Replicate data are used to generate the inputs to the python scripts that generate the final pad and road shapefiles. Inside the model, the pad centroid (actual landscape UTM coordinates) and the relativized boundary coordinates of the assigned digitized pad pattern are merged to create UTM coordinates of pad polygons. Attributes of each pad are stripped from output files and summarized in an interim file that is subsequently used in the python script to attribute the final pad shapefile. Similarly, road vertices are accessed to generate road segments, and attributes of each road are summarized in an interim file used to populate attributes in the final road shapefile.

Output data files and shapefiles are described below.

1. dumpp#YYYY contains information about every pad, including the number of wells established on the pad. The "#" is the replication number; YYYY is the calendar year.

PID	AU #	Proj #	SID	Start	Stop	Pad Status	Total no. wells	No. active wells	Lease (%)	Public mineral rights (%)	Oil/gas potential
For an active well: Start	For an active well: Stop	For an active well: Bottom hole area (acres)									
…											

The first row contains information about a pad.

AU # and Proj # are numeric codes for the AU and Project area where the pad resides.
SID is the ID of the section that contains the pad
Start and Stop are in calendar years and indicate the year the pad was created and when it became inactive (Stop = 0 if still active).
Pad status is 1 if active, else 0
Total no. wells is the total number of oil/gas wells established on the pad
No. active wells is the no. of active wells at the end of the simulation
Lease (%) is the percentage of the section with oil/gas leases
Public mineral rights (%) is the percentage of the section with public subsurface mineral rights
Oil/gas potential is the estimated oil/gas potential as a probability (0.0 – 1.0)

If a pad lacks active wells, the next row would be a repeat of the first row for the next pad. Otherwise, subsequent rows will list the active wells (1 row for each active well), and consist of the calendar year the well was established (Start) and when it became inactive (Stop; =0 if still active), and the Bottom-hole area (acres) of the well (specified in a build-out design).

After listing all the active wells on a pad, the first row is repeated for the next pad, etc..

2. padpattern# - contains information on the pad pattern selected for each simulated pad, and the location of a simulated pad on the landscape. The "#" is the replication number.

PID	Palette pattern no. (PatternID)	Start year (calendar yr)	Centroid, UTM N	Centroid, UTME	Status; 1 for active; else 0
…	…	…	…	…	…

PatternID cross-walks with the IDs in padpatcoords.in (argument #22, described above).
Start year is the calendar year the pad was established
Centroid UTM coordinates, North and East
Status designates active or inactive at the end of the simulation

3. roadfo#YYYY – contains the status of road segments that have changed status since the last time this file was output during the simulation. Unlike the other output files, information is recorded for baseline and simulated roads. The "#" is the replication number; YYYY is the calendar year.

RID	Status; 0= inactive, 1 = active	PID	Calendar year
…	…	…	…

RID is the road-segment ID
Status indicates if a road became active or inactive
PID is the pad ID associated with the road segment and that motivated changing the status of the road
Calendar year is the year the road changed status

4. rdlines#.csv – contains road vertices of simulated roads in UTM coordinates. The "#" is the replication number. This is a csv file that can be displayed in ArcGIS without transformation. The labels shown in the first line are actually listed in the first line of this file.

N	E	RDID	YR	BUFFER	CCODE
UTMN	UTME	RID	Start (calendar yr)	Buffer size (m)	Always =1
Repeat the above line for all vertices of this RID, then repeat for all subsequent RIDs					

UTMN and UTME coordinates are listed first

Simulated RIDs start at 1 but during processing of this file are transformed to sequential numbering starting with 1+ the maximum RID of the baseline roads

YR is the calendar year the road was created

Buffer is the buffer distance (m) used to estimate road width, and is 0.5 times the estimated width of each road type

Shapefiles

Pad and road shapefiles are created at the end of a simulation replication using the 4 data files listed above, and the files noted at the beginning of the **Model Execution and Inputs** section. Shapefiles are named padrep#YYYY, rdrep#YYYY, and rdbufrep#YYYY, where '#' is the replication number and YYYY is the calendar year. Pads are represented as polygons in padrep#YYYY. The rdrep#YYYY shape file is a line file, and rdbufrep#YYYY contains buffered roads to facilitate estimating actual road-surface area. Attributes of each shapefile are described below.

1. Attributes of a pad shapefile (padrep#YYYY).

ACCESS	Unique identifier (PID) – access is short for accession
AU	Oil/gas Assessment Unit code (coding of AUs)
PROJ	Project code (coding of energy fields)
SID	Section code (SID)
START	Calendar year the pad was established
STOP	Calendar year the pad became inactive. If still active, Stop=0
ACTIVE	1 if active, else 0
NWELLS	Total no. of wells on the pad
ACWELLS	No. of active wells on the pad
Area_Ha	Size of pad (ha)

2. Attributes of road shapefile - line format (rdrep#YYYY).

rdid	Unique identifier (RID)
type	Coded type – see argument #1 in Table 1.
width	Estimated road width (m)
buffer	Buffer size (m) which is one-half the road width
yr	Calendar year the road was created; however, year of baseline roads is unknown and set to 2009
active	1 if active, else 0
yrinact	0 or the year the road became inactive
Len_km	Length of a road (km)

3. Attributes of road shapefile - buffered format (rdbufrep#YYYY).

rdid	Unique identifier (RID)
type	Coded type – see argument #1 in Table 1.
width	Estimated road width (m)
buffer	Buffer size (m) which is one-half the road width
yr	Calendar year the road was created; however, year of baseline roads is unknown and set to 2009
active	1 if active, else 0
yrinact	0 or the year the road became inactive
Len_km	Length of a road (km)
BUFF_DIST	Repeat of buffer
Area_Ha	Area (ha) of a buffered road

POST-PROCESSING

The shapefiles are the primary outputs or results from the model, and are intended to be used in post-processing procedures to estimate the effects of oil/gas infrastructure on biophysical attributes, such as amount of surface disturbance, wildlife populations, and wildlife habitat. For the latter 2 assessments, infrastructure attributes (e.g., pad nos., road density) used to model effects on wildlife species can be extracted from the shapefiles to estimate expected impacts of a simulation run. The overall intent of the footprint model is to compare these impacts among a range of proposed and alternative build-out designs as a way to identify oil/gas development approaches that at least minimize environmental impacts. Post-processing procedures have been developed to assess impacts to certain wildlife species. However, these procedures continue to change with the inclusion of additional species, as new models of species' impact are reported in the literature, and in an attempt to accelerate the speed of assessments. For some users, a vector version of the data is too costly to use in analyses in terms of time, or too difficult given preferred assessment methods. In such cases, vector files are converted to rasters (30 m and larger grain sizes) to expedite analyses. For all of these reasons, post-processing procedures are not provided, but left up to a user to develop according to their needs.

COMPILATION, SYSTEM REQUIREMENTS, & SOURCE CODE

The EFM model was developed in the Microsoft Visual C++ environment (vers10.0, © 2010 Microsoft Corporation) as a 64-bit stand-alone simulation model. In the Visual C++ environment, the 64-bit configuration and 'Compile as C code' options must be selected since all source is written in C and uses 64-bit precision. Otherwise, a standard Application project is sufficient to compile and link the code. The model source code is generic C code; thus it can be compiled and linked using any 64-bit C compiler and linker. The model relies on 2 python scripts to create the pad and road shapefiles. In addition to a 64-bit operating system, arc.py must be installed and accessible.

The executable only runs on a 64-bit platform, and can require large amounts of computer memory. The storage of new pads and roads and some processing steps dynamically allocate memory; thus memory requirements increase with development intensity and over the course of a simulation. For instance, a 10-yr simulation in a small project area may only require 4-6 GB of memory. Based on experience, a 30-yr simulation that established 30,000 wells over the entire 7.7 MHA SW Wyoming study area required 13-40 GB of memory. An intense development of 170,000 wells exceeded 40 GB of memory and could only be run on machines with 96 GB of memory.

The source code is listed below, organized by general function. A description of processing steps is included within each file, and the general use of variables and structures in include files are described at the top of the file.

Function	Source code
Main module – controls initiation and Monte Carlo replications	efm.c
Routines to ingest binary, geospatial data layers	readareapatchesb.c, readdemb.c, readfedmb.c, readlekgridb.c, readnontraverseb.c, readpadsb.c, readrdidb.c, readroadsb.c, readsageb.c, readsecpad.c, readsgb.c, readsurfb.c
Routines to ingest ASCII tabular lists	readanchorpts.c, readauslist.c, readfillr.c, readginfo.c, readlek.c, readmsect.c ,readmspace.c , readpadcords.c, readpadinfo.c, readpadinfo2.c, readpatterns.c
Routines to initialize global structures, variables, and to open output files	init.c, reinit.c, openoutputfiles.c
Routines related to the derivation of new oil/gas roads that connect a new oil/gas pad to the existing road network, and to road deactivation	assignroads.c, checklines.c, checkrd.c, checkrdactive.c, checkrdnew.c, curve.c, datarecord.c, distance.c, donut.c, drawlinemp.c, drawsline.c, drawslinep.c, drawslinepcheck.c, drawslinepmod.c, findnearestcheck.c, findrdcells.c, padrdoverlap.c, processrds.c, rdgrid.c, rdgridup.c, roadskirt.c, setrdcells.c, sort.c, trig.c, updateanchorpts.c

Function	Source code
Routines that control the establishment of pads and roads on the landscape	establish.c, findlekdistance.c, locate.c, locatefunc.c, processcorepad.c, selectsections.c, selsecopts.c, setwells.c, simdevelop.c
Routines to update global variables with new pad and road information	masterupdate.c, perform.c, printcorearea.c, recordbha.c, sdcorearea.c, storespec.c, updaterdstore.c
Functions to translate between row and columns of grids and UTM coordinates	convtopts.c, convtorc.c
Routines that processes the oil/gas build-out designs	fillmod.c, redofill.c
Routines and python scripts used to create the output pad and road shapefiles	padshapefile.c, rdshapefile.c, dopads.py, dordslb.py
Random number generators	ran2.c, ran22.c, ran23.c
Misc. functions	deactivatewells.c, dumppads.c, findactualproj.c, pause.c, valid.c
Include files containing global structures and variables	binary.h, createpads.h, curve.h, efm.h, fnames.h, geo.h, padalloc.h, padpat.h, plss.h, rdend.h, rdgrid.h, rdshapefil.h, reduce.h, roads.h, TBHA.h
Pre-Processing Programs: Routines to create customized binary files from ArcGIS ASCII files, and to create ArcGIS ASCII files from customized binary files	crinput.c, binascii.c