Probabilistic Modelling and Complex Energy Landscape Design and Planning: An Experimental Approach

Timothy MURTHA[1], Richard HAMMOND[2], Brian ORLAND[2] and Lacey GOLDBERG[2]
[1]PENN STATE UNIVERSITY, Pennsylvania/USA · murtha@psu.edu
[2]PENN STATE UNIVERSITY, Pennsylvania/USA

Abstract

Spatially explicit and temporally sensitive probabilistic modelling and environmental simulation offer advantages to large scale landscape planning and design, especially in the context of energy exploration. Using a case study from North America, we investigate the strengths and appropriate uses of probabilistic modelling for regional decision making. The case study explores shale gas extraction from the Marcellus Shale region. We not only demonstrate the usefulness for integrating these models in the design and planning process, but also offer some guidance for the appropriate scales and applications.

1 Introduction

This is a descriptive paper of a complex process. Here, we report the results of an experimental approach to applying spatially relevant probabilistic modelling and environmental simulation to systems thinking in the design and planning process. The length of this paper does not allow for in depth discussion about the specifics of our approach. Instead we focus on the application of the results in design and planning. Another paper will discuss our modelling approach with greater detail. Here, we examine the results of an adapted maximum entropy approach and software (PHILLIPS et al. 2004), originally designed for species habitat modeling and large-scale biodiversity conservation. We use this approach to investigate the complex dynamics of an intensive energy landscape in the United States, i.e., shale gas exploration in Pennsylvania. Using an iterative approach, we adapted techniques first applied by JOHNSON (2010) to project the form and distribution of Pennsylvania's future energy extraction landscape. The purpose of the simulations is not to accurately predict specific locations of natural gas well pads or extraction sites, but to identify design and planning scale sub-regional patterns. The goal is to also project potential futures on landscapes where exploration and extraction activities are likely to occur in the next 15-20 years and to quantify the infrastructure needed to support the activeties, e.g., road and pipeline construction. It is a spatially and temporally sensitive approach to modelling.

Relying on data published by the *Pennsylvania Department of Conservation and Natural Resources*, we first model future projections of gas exploration activities using current observations of the presence of drilling from 2009 to 2014. Similar to JOHNSON's report (2010) these observations were compared to basic environmental and geological data, such

Buhmann, E., Ervin, S. M. & Pietsch, M. (Eds.) (2015): Peer Reviewed Proceedings of Digital Landscape Architecture 2015 at Anhalt University of Applied Sciences. © Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach. ISBN 978-3-87907-555-3.

as depth to the Marcellus deposit and thickness of the deposit measured across our region of interest. The goal was to interpolate a probability grid throughout the region that highlights similar conditions to observed locations (*i.e.,* permitted well pads), thereby allowing us to develop a model for the potential distribution of future gas exploration activities (PHILLIPS et al. 2006 and 2004).

We relied on spatial analysis/data preparation carried out in GIS and MaxEnt software to develop our initial projections (PHILLIPS et al. 2004). MaxEnt is software based on the maximum entropy approach which produces a model distribution based on a set of environmental layers and geographic coordinates of occurrences or observations (ELITH et al. 2011, PHILLIPS et al. 2006 and 2004). In order to fully understand the model outputs, we ran the projects through multiple iterations, each time adding potentially influential variables not originally considered by JOHNSON (JOHNSON et al. 2010), such as land use, land cover and proximity to infrastructure. Due to global market decline in the price of natural gas, we also ran the model through annual iterations to examine whether the projections would be sensitive to changes in activity due to changing market conditions. Each time we ran the model we reserved 50 % of our sample to test the efficacy of the model.

Using probabilities developed in MaxEnt, we then developed a series of scaled formal environmental simulations, using cost distance modeling, to examine how the future shale gas activities, would potentially influence the development of necessary infrastructure, specifically focused on pipelines and roads. Elsewhere, we have demonstrated the relative importance of pipelines and roads from a design and planning perspective, when compared to well pad locations (ORLAND & MURTHA 2014). As a baseline, we first calculated least cost surfaces for infrastructure development and compared these outputs to models that incorporated federal and local regulations to protect natural and cultural resources. All of these projections were then used to develop a series of spatially relevant design and planning ideas.

Finally, all of these projections were used at a county or regional scale to inform a series of design and planning projects focused on topics such as sense of place, energy futures and water planning. We only briefly describe our modeling methods here emphasizing the application of the information to design and planning. Through this work we have identified two key conclusions or considerations:

1) Outputs from probabilistic models need to be cautiously incorporated into the design and planning process at multiple scales. Specific locations should not be reported (used for design projects) and regional or sub-basin summaries provide more useful design scale information.

2) Like all useful analytical techniques, probabilistic models should be integrated within the design and planning process, iteratively. Simply, we identified a number of ways to improve the projections during the execution of the many design and planning projects. In future attempts we will include time to re-run projections during the design and planning process.

2 Probabilistic Modeling of Decisions

2.1 Maximum Entropy Probabilities of Wellpads

Similar to the Nature Conservancy's first efforts in 2010, we were interested in the form and distribution of future well pad activity as a starting point to understand the spatial and temporal patterns of Pennsylvania's energy landscape related to shale gas (JOHNSON 2010). We used the same software and techniques applied by JOHNSON and colleagues (JOHNSON 2010), *i.e.,* a machine based learning modelling approach known as maximum entropy. Maximum entropy (MaxEnt 3.3.3k) was used to study the relationships between 4085 observations (reported unconventional permits as of 8/2014) and a variety of environmental variables, including: *distance to roads, depth of Marcellus, thickness of Marcellus, slope, distance to pipeline protected natural area* and *water*. Additional variables like current landuse were added to the model, but are not discussed in detail here.

In another paper we will look closely at how selected variables and parameters influence the outcomes. For purposes of this paper, we discuss seven variables detailed above and shown in Table 1. The model results shown in the table illustrate the contribution of each variable, thereby indicating what variables influence the overall probability distribution for potential future occurrences (Figure 1). From our first run, distance to roads along with depth of the Marcellus and thickness of the Marcellus proved to be the most influential variables (positively correlated).

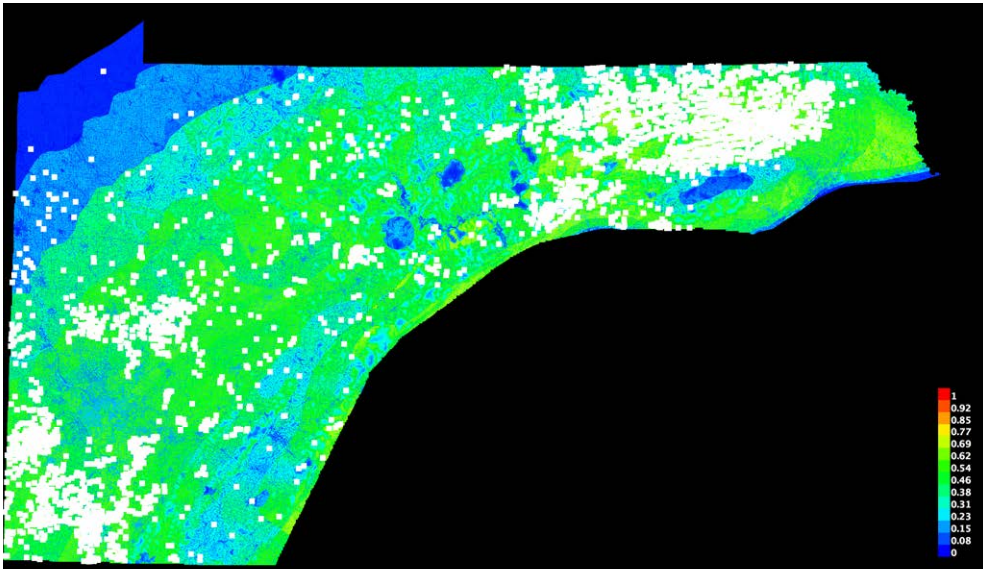

Fig. 1: Probability distribution map of the first run model with white dots representing the 4085 observations used to derive the model and probabilities for future activities

We are developing methods to better test the efficacy of the model and will continue to work on those, but some clear patterns emerged. Natural areas, while not contributing a great percentage for influencing overall future activities is only considered 'non-contributing' because a small percentage of the entire study area is defined as *Natural areas.* These areas are protected from future activities through federal or state regulations and may not appear to be significant to the model as a whole, but are reliable predictors of future activity. What we are learning is that some variables may not contribute a large percentage to the overall model, but are essential for modelling representative surfaces, because of their reliability. The opposite consideration is also important. *Distance to roads* is influential largely due to the ubiquitous distribution of roads in Pennsylvania. Simply, roads are essential to the model, but not reliable predictors on their own. Finally, the two factors most significantly contributing to the probability of future activities both spatially and statistically were the thickness of Marcellus and depth of Marcellus. This most likely reflects a profit optimization model driven by industry and associated with the high costs of fracking (MURTHA & ORLAND2014).

Table 1: Percent contribution of each variable for the annual iterative model and cumulative, 2014 MaxEnt model

Variable	2009 %	2010 %	2011 %	2012 %	2013 %	2014 %
Distance to Roads	34.7	36.3	38.4	35.8	37.4	38.1
Depth of Marcellus	20.8	26.2	30.3	35.1	35.7	35.4
Thickness of Marcellus	18.3	16.4	17.6	17.4	16.2	16.2
Slope	15	12.7	6.7	4.6	4.7	4.7
Distance to Pipeline	8.9	5.9	4.8	4.5	3.1	2.9
Protected Natural Area	1.1	1.6	1.7	1.9	1.9	1.9
Water or Water Body	1.1	0.8	0.6	0.8	1	0.9
Observations	**1166**	**2006**	**2769**	**3352**	**3992**	**4085**

Our first run included all of the permitted well pads as of 8/2014 or 4085 observations. However, in our community work, we noticed a change in pace of activities throughout the region, influenced by a substantial decrease in the price of gas on a global scale between 2009 and 2014. Because of those changes we wanted to test whether the market conditions would influence our first run model. To account for that we assumed that annual permit activity could account for changes in the spatial decision-making by companies, again reflecting a profit optimization model. The results of the iterative model are presented in table 1 and illustrate no fundamental difference between the 2009 model with only 1166 permit observations and the 2014 model with 4085 permit observations. This is a critical area we are now exploring through alternative modelling scenarios. Either permitting doesn't capture the market fluctuations or interim fluctuations in price do not have short term impacts on permitting activities. Additionally, in future work we plan to engineer a smarter probabilistic model that treats permit observations as specific events in time and not only separated by year.

These model outputs only provide us with probabilities of future activities and are not necessarily scaled to the extraction methods and techniques. To translate the probabilities into meaningful landscape information, we relied on estimates developed by the Nature Conservancy (2010), who estimated that 60,000 new wells would be drilled by the year 2030 in Pennsylvania. Because of the complexity involved with horizontal drilling, we also had to estimate how many wells would be drilled per well pad. In 2010 the average number of wells per pad was two, but each well pad can support a number of wells and can also extract gas from 80 to 170 acres (JOHNSON 2010). Here again, because we are far more interested in projecting sub-regional patterns of future activities, we relied on three scaled estimates developed in 2010. The *low impact* model called for 6,000 well pads with10 wells per pad and pads spaced 5,250 ft apart. The *medium impact* scenario called for 10,000 well pads with 6 wells per pad and pads spaced 4,100 ft apart. The high impact scenario called for 15,000 well pads with 4 wells per well pad and pads spaced 3,350 ft apart. For comparative purposes these are the exact same scenarios used in 2010 (JOHNSON 2010, 12).

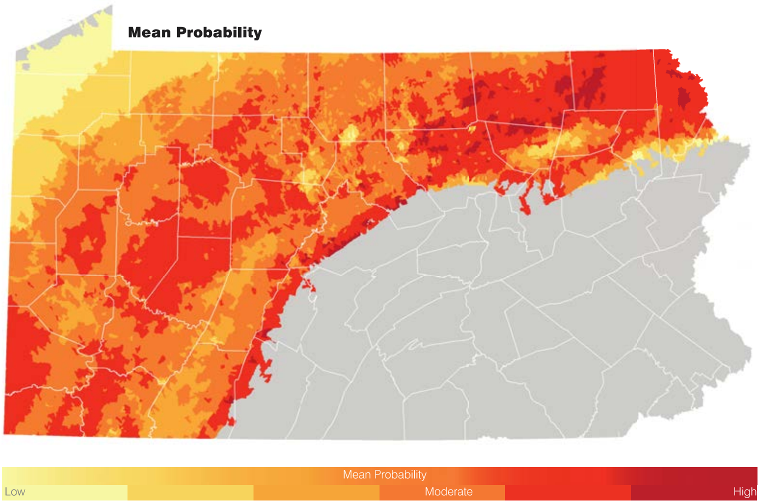

Fig. 2: Mean probabilities for the high impact model were calculated for each of the sub-basins (small watersheds) that intersect with the Marcellus shale deposit

Having established three scenarios we resampled the probability map into three new resolutions, 5250 ft, 4100 ft and 3350 ft. We then selected the number of well pads from the highest probable locations. For example for the low impact model we selected the 6,000 highest probability pixels. We also summarized the probabilities by watershed sub-basin (Figure 2).

From a large landscape scale, the summary model by sub-basin (or small watershed) allowed us to better consider important local and community issues quickly by comparing the potential for future activities to basic attributes of each small sub-basin, such as amount of roads, streams and welands, which are all potentially impacted by these activities. Simply, from a statewide perspective specific locations were not necessarily useful for design and planning, but when relying on summaries we could identify key regions and sub-regions for focused work. So, even just relying on some summary and comparative analyses we were already able to isolate areas of importance and areas that showed signs of potential impact. In fact, we developed several story maps to illustrate these areas (see: http://marcellusbydesign. psu.edu). Importantly, the first run models provide an entry point to formally model broader impacts of these activities, e.g. pipeline and infrastructure development. While there are reliable estimates indicating that a single well pad directly and indirectly impacts 30 acres of habitat. The well pad is just 10% of that total, infrastructure provides another 20 % and roughly 70 % are associated with indirect impacts (JOHNSON 2010). From a landscape perspective, projecting the probable locations of those impacts is critical for design and planning projects.

2.2 Cost Distance Models for Pipelines

Our next step in modelling was to develop spatially explicit projections about pipeline construction so that the gas extracted from well pads could get to market. Much of the recent discussion about pipelines in the United States is focused on large transmission pipelines, but for the Marcellus shale and other shale gas resources thousands of miles of gathering lines need to be constructed from each well pad to a compressor station adjacent to a major transmission line. The process for constructing the gathering lines is far less regulated than large transmission lines, so the purpose of our least cost models were to evaluate several scenarios for developing standards or regulations for protecting natural habitats, like wetlands and core forest.

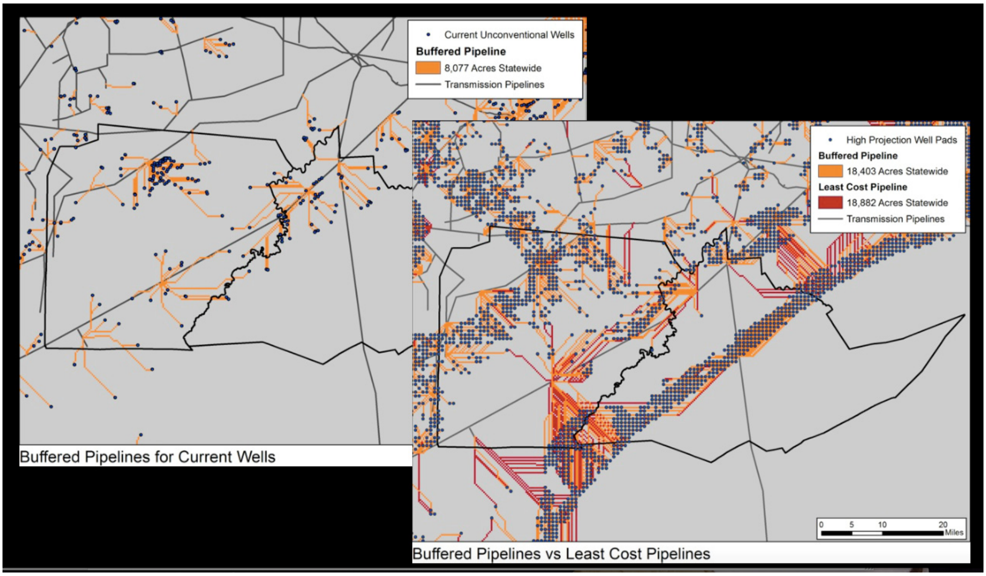

Fig. 3: Comparison of two scenarios illustrating the impact of future drilling and quantifying the difference between a straight least cost model and one model that avoided all wetlands by 200 ft. Overall, there was very little negative impact when buffering wetlands.

The cost distance models we developed were parameterized so that we could isolate the projected cumulative impact given a variety of regulatory scenarios. We started with connecting existing well pads and then connected projected well pads first through a least distance method and then forced the pipelines to avoid important resources, like wetlands (Figure 3). Interestingly, protecting wetlands and other natural resources did not necessarily add substantial impact of or length to pipelines. This was an unexpected, but important outcome of the pipeline models.

3 Discussion

While the initial purpose for developing the probabilistic and formal models described in the paper were to simply allow us to project with reasonable accuracy the future sub-regional patterns for landscape planning and design in the Marcellus region, we quickly recognized how valuable a broader application of probabilistic modelling in landscape design and planning projects could be. Moreover, we identified two key points of discussion:

1) Outputs from probabilistic models need to be cautiously incorporated into the design and planning process at multiple scales.
2) Like all useful analytical techniques, probabilistic models should be integrated within the design and planning process, iteratively.

The outputs from the model provided some spatial awareness and patterning for activities, but need to be thoughtfully engaged as models to inform decision making. When we shared these projections with individuals and students, specific location of both well pads and pipelines became central considerations, instead of our intended outcome. Moreover, models can't be simply run once and responded to by design at one specific scale. Summaries of information were valuable at a regional scale and more specific location and patterns were effective as we shifted from a state to a county or regional scale.

We embraced modelling iteratively and are continuing to test and refine our assumptions. One key needed area for improvement is to revisit the probabilistic modelling during the design and planning projects. By doing so we will not only improve the models, but also better understand how to use these outputs in specific design projects.

4 Conclusion and Outlook

Formal probabilistic models aren't new to design and planning. In fact, they are commonly used by civil engineers, designers and planners for predicting urban storm water dynamics and influencing stormwater design and decision making (ADAMS and PAPA 2000). But beyond stormwater, formal use of probabilistic models in design is fairly underrepresented. Due to the recent emergence of geodesign (STEINITZ 2012) and clear needs for complex systems thinking in sustainable design and planning (ALLEN et al. 2003), we believe that probabilistic modeling should be explored more broadly on a landscape scale. Especially in the context of complex systems thinking, we conclude that there are substantial benefits to applying these techniques. Simply, MaxEnt and the maximum entropy approach offer unique potential for regional landscape design and planning.

References

ADAMS, B. J. & PAPA, F. (2000), Urban Stormwater Management Planning with Analytical Probabilistic Models. 1st edition. Wiley, New York NY.

ALLEN, T. F. H., TAINTER, J. A. & HOEKSTRA, T. W. (2003), Supply-Side Sustainability. Columbia University Press.

ELITH, J., PHILLIPS, S. J., HASTIE, T., DUDÍK, M., EN CHEE, Y. & YATES, C. J. (2011), A Statistical Explanation of MaxEnt for Ecologists. Diversity and Distributions, 17 (1), 43-57. doi:10.1111/j.1472-4642.2010.00725.x.

JOHNSON, N. (2010), Pennsylvania Energy Impacts Assessment: Report 1: Marcellus Shale Natural Gas and Wind. The Nature Conservancy, November 15, 2010. http://www.nature.org/media/pa/pa_energy_assessment_report.pdf.

ODLING-SMEE, J. F., LALAND, K. N. & FELDMAN, M. W. (2003), Niche Construction: The Neglected Process in Evolution (MPB-37). Princeton University Press.

PHILLIPS, S. J., ANDERSON, R. P. & SCHAPIRE, R. E. (2006), Maximum Entropy Modeling of Species Geographic Distributions. Ecological Modelling, 190 (3-4), 231-259. doi:10.1016/j.ecolmodel.2005.03.026.

PHILLIPS, S. J., DUDÍK, M. & SCHAPIRE, R. E. (2004), A Maximum Entropy Approach to Species Distribution Modeling. In: Proceedings of the Twenty-First International Conference on Machine Learning, ICML '04. New York, NY, USA. ACM. doi:10.1145/1015330.1015412.

STEINITZ, C. (2012), A Framework for Geodesign: Changing Geography by Design. ESRI Press, Redlands, CA.