Alternative Future Scenarios:
Development of a Modeling Information System

Project funded by:

United States Department of Defense
Strategic Environmental Research and Development Program

Contract #DACA72-01-P-0045

Submitted by

Dr. Mary E. Cablk
Desert Research Institute
Division of Earth and Ecosystem Sciences
2215 Raggio Parkway
Reno, NV 89512
(775) 673-7371
mcablk@dri.edu

Dr. William Langford
1140 Bel Aire Dr.
Santa Barbara, CA 93105
(805) 705-8485
langfob@yahoo.com

Dr. Anna Panorska
University of Nevada, Reno
Department of Mathematics, MS 084
Reno, NV 89512
(775)784-6773
ania@unr.edu

April 22, 2003

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
22 APR 2003		**00-00-2003 to 00-00-2003**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Alternative Future Scenarios: Development of a Modeling Information System	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Desert Research Institute,Division of Earth and Ecosystem Sciences,2215 Raggio Parkway,Reno,NV,89512**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**76**	

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Acronyms

AFS	**Alternative Future Scenario**
AFSM	**Alternative Future Scenario Modeling**
AFSN	**Alternative Future Scenario Null**
ANOVA	**Analysis of Variance**
CUF	**California Urban Futures**
CURBA	**California Urban Biodiversity Analysis**
DEM	**Digital Elevation Model**
DIAS	**Dynamic Information Architechture System**
DoD	**Department of Defense**
EPA	**Environmental Protection Agency**
EROS	**Earth Resources Observation System**
GRASS	**Geographic Resources Analysis Support System**
GAM	**Generalized Additive Model**
GIS	**Geographic Information System**
GLM	**Generalized Linear Model**
GUI	**Graphical User Interface**
GWRP	**Global Warming Research Program**
IRLS	**Iterative Reweighted Least Squares**
ITAM	**Integrated Training and Area Management**
LUCAS	**Land Use Change Analysis System**
MSS	**Multi-Spectral Scanner**
NAD	**North American Datum**
NALC	**North American Land Characterization**
OO-IDLAMS	**Object-Oriented Integrated Dynamic Landscape Analysis and Modeling System**
ROC	**Receiver Operator Characteristic**
SERDP	**Strategic Environmental Research and Development Program**
SLEUTH	**Slope, Land cover, Exclusions, Urban areas, Transportation, Hydrologic**
UTM	**Universal Transverse Mercator**
XML	**eXtensible Markup Language**

EXECUTIVE SUMMARY

Military installations face challenges that may impact mission readiness and daily operations. Of these challenges, civilian urban development on lands adjacent to installations is among the most pressing, primarily due to the number and variety of unintended consequences associated with urban expansion. The consequences include safety risks; noise; impacts to plants, animals, and cultural resources; dust emissions and other air and water pollution; and installation-specific issues. The Department of Defense (DoD) has investigated the use of alternative future scenario modeling (AFSM) to predict and remediate potential impacts of civilian development on military bases. The application of AFSM to military installations throughout the US has produced many predictions, or futures, that suggest how landscapes that surround selected installations may change during the next few decades. While the success of AFSM in increasing DoD's understanding of the vulnerabilities associated with regional land use changes is well documented (Cablk, et al, 1999; Gonzalez et al., 2000; Gunter et al., 2000; Steinitz et al., 1996), transfer of the AFSM process to the military has yet to occur. This is primarily due to the complexity of this process, which requires expertise in remote sensing, demography, geographic information systems (GIS), computer programming, and statistics.

This project had several related objectives. First, we sought to identify common variables thought to either be correlated with or actually drive patterns of development. As part of this objective, a comprehensive statistical analysis was conducted for the Coachella Valley, CA to determine if some basic rules regarding development could be established. The results of this analysis were compared to results from other alternative futures research. An extensive review of existing models that can be used to create one to many components of alternative future scenarios was completed, including an assessment of OO-IDLAMS and LUCAS. Neither of these models is applicable beyond their current geographic application and OO-IDLAMS is not an alternative future scenario modeling tool, it is an ecological modeling tool. Based on the results of these objectives, it was determined that more information is needed at an installation level, across all installations, before a prototype dynamic information system can be presented for consideration by DoD. In fact, a comprehensive understanding of needs across installations is currently lacking and is thought to be where alternative futures modeling can be supplemented with tremendous gains in the field.

There are many modeling tools in existence to date and we believe that components of these models, programs, and tools can be combined effectively for an alternative modeling assessment tool. It is unclear whether or not the actual construction of alternative futures can be accomplished entirely without the contribution of certain experts. However, it is known that military officials and laypeople can create non-empirically based alternatives with existing technology tools.

INTRODUCTION

Military installations face significant pressure from urban growth. This trend is expected to continue. In fact, pressure from civilian development is expected to continue at an increasing rate. Individual installations must have spatially explicit scenarios that depict and describe the actual growth (physical development) immediately adjacent to and in the general vicinity of each base to determine what actions would be appropriate to mitigate or circumvent restrictions to the military mission at the installation level. Currently alternative future scenario modeling is conducted as one or more custom models for one or a group of geographically clustered installations. Alternative future scenarios are also evaluated for a specific set of foreseen or occurring issues relevant to the selected installation(s). For example, in the Mojave Desert of California, dust, noise and impacts to biodiversity are key issues facing each branch of the armed forces. In the southeast, non-point source pollution and biodiversity are critical. Noise and civilian safety issues are in the spotlight wherever human population is dense. These situation and installation-dependent analyses are expensive because they are reactionary. The quality of analyses on a case-by-case basis may vary greatly. The military needs to consider whether there are other options that can improve outcomes of modeling efforts, decrease expense, and relieve pressure of urban development in a proactive manner.

The Department of Defense (DoD) requires reusable models to project scenarios of urban growth pressure around military bases to avoid the cost and/or variation in quality from base-specific solutions. At the same time there is a need to categorize the types of problems (issues) and the concomitant local constraints, as well as define evaluation criteria for these models in order to choose appropriate solutions. Solutions may be at an installation level, a geographic level, or DoD-wide. Models and solutions must be flexible.

Modeling frameworks like DIAS may provide a partial solution, but do not address the entire problem. Different bases have different needs, but these needs remain unknown beyond the installation officials. The specificity of installation needs may or may not be shared, and this information is unknown as well. This raises questions regarding *how much reuse of models is possible*? Is it necessary to begin anew every time an issue is raised, as in the case of Camp Pendleton? If the technology advances between analytical time frames to an extent that warrants beginning from relative scratch, then perhaps the answer is yes. How would one know if this were or were not the case? To answer this question we need to define the requirements of specific installations and determine if there are generalizable types of situations that admit reusable solutions. There are currently many models to choose from, but the field is still young and changing rapidly. Evaluation criteria must be defined for the models based on the classes of problems found in the base requirements and on the resource and institutional constraints present in each situation.

We can define the evaluation criteria now and we can identify many models now, but we need to define the requirements and the assumptions about SERDP's goals and policies for the problems before it further considering specific models in detail. The long-term nature of the problems installations face and the complexity of the existing and available models required to address those problems may mean that SERDP needs one or more people on staff to provide a continual

base of expertise in the evaluation, selection, and use of the model or models selected.

Given the need for a model to help with understanding and solving a particular problem, there are many types from which to choose, and many criteria to evaluate in making a choice. In the following sections we will first give some brief background on the different classes of models that are available and then give a detailed list of criteria used to evaluate specific models. Afterwards, we will briefly discuss a couple of specific models since there are several excellent reviews of existing models that do not need to be duplicated here. Moreover, the actual choice of a model should be entirely dependant on the specific questions that require investigation in combination with the constraints and desires of the particular users of the model. Those specific questions, constraints and desires are still largely undefined for DoD, and were thus found to be likewise for this project, and will be enumerated in a separate section at the end of this report. Until those questions are answered, detailing the specifics of a large number of models is not worth doing since the answers to those questions will likely eliminate whole classes of models almost immediately.

Much of the information in this report is drawn from the reviews mentioned above and the reader seeking more detailed information should consult those reviews (Agarwal, 2000; EPA, 2000; Parker, 2002). Our research also turned up other reviews that may be worth consulting (Baker, 1989; Sklar, 1991) but we did not examine those for this report since there were more recent reviews available. However, we did find that there was not a lot of overlap in the models covered in each review, so the older reviews may be worth examining as well.

Research Objectives

This research was a high-risk pilot project, designed to be the first of potentially three phases which would ultimately provide DoD with a fully operational tool for simulating alternative future scenarios. Within this three-phased approach, it was recognized that expertise, computational services, and facilities vary at each installation, and that the system design must take these differences into account. For this first phase, the objectives were as follows:

1. Identify a suite of "standard" variables necessary for simulating/building alternative future scenarios
2. Evaluate different mathematical and statistical models appropriate for building alternative futures
3. Build a system that will facilitate direct input by military officials and their contractors by providing feedback to the modeling process based on local knowledge of the installation and regional specifications
4. Provide DoD with a prototype dynamic information system that models alternative future scenarios

Assumptions

There must be some bounding box, some set of criteria, recognized and defined before any decisions on models can be undertaken. What are the issues at hand and what are their rankings, respectively, in terms of importance? Do we assume that what is important today will be at the

forefront of concern in *x* years from now? What is *x*, that time frame that will dictate where focused efforts lie across the landscape and how priorities for assessment will be set?

Urban development is a universal challenge

We must make some assumptions at a general level. First, we assume that all military installations face impairment of mission in some way, shape or form from urban development. This may be from direct impacts, as in building of structures that result in forced change of flight patterns, or indirect impacts, such as compounded non-point source pollution, or threats to biodiversity. An ecological system surrounding an installation may be well-adapted or capable of "handling" the load imposed by that installation. Under a scenario of urbanization within that once low density system, the ecology may be disrupted to a level at which fingers point to the military as the cause of the resulting adverse conditions. No installation will escape impacts from civilian development. Each installation will be faced with its own challenges in this respect. These challenges may be categorized. Not all categories will be filled at the same time and the membership in each category of threats-from-urbanization will be dynamic over time.

Challenges vary with mission and geography

Second, we assume that issues or challenges each installation faces or will come to face, varies by installation mission and by geography. Some installations are relatively inactive in terms of environmental impacts or in terms of conducting business in such a way as to produce negative externalities noticeable by the surrounding community. Hawthorne Army Depot and similar munitions depot are examples. Likewise, other installations may have a very large landholding and conduct extensive training and testing continuously but these may go fairly unnoticed due to their geographic location. The Naval Station at China Lake and the Naval Air Station Fallon are examples (Gunnery range west of Salt Lake). This assumption must include recognition of transience or a certain level of dynamics in that missions may change over time, and geography at a national level changes in terms of population density, distribution, and popularity. Ironically, locations where the military once sought refuge from public profile, now are among the fastest growing and most popular areas for civilian settlement, especially the desert southwest.

The military will always be identified as the source of the problem

Monitoring may be a standard component of an installation's natural resources division or division of public works, and this varies with branch of the armed forces. For example, the Army's Integrated Training and Area Management (ITAM) program is designed to monitor and mitigate, among others, Army training lands. As development expands towards an installation, regardless of branch, the ecological system is affected. The ecology of the area is more heavily burdened and may reach a threshold beyond which the existing system becomes dysfunctional. The non-military community, relatively new to the area, will always first blame the military for whatever the problem is. There are a few reasons for this. An obvious reason is that many communities are inherently suspect of government, particularly one perceived to be as secretive as the military. A second reason is that people do not look to themselves for causes of negative impacts, period. Air pollution proven to be caused from local vehicles is vehemently blamed on tourists and non-residents, even when provided with evidence (proof) to the contrary. Therefore, regardless of whether or not the military installation contributes no, little, or a significant amount to the problem, the military will always be suspect and burdened with proving innocence. Even

with monitoring data, programs such as ITAM are relatively new, are not universal across all installations, and do not necessarily collect the relevant data.

DoD cannot afford to ignore the hierarchical effects of development pressures

The risk to national security from closure and/or impairment of the function of military installations is both real and significant. A model for risk analysis was presented to SERDP at the 2001 SERDP symposium by Cablk and the reason for conducting alternative future scenario modeling remains rooted in this principal. At some point, national security faces risk of collapse and that risk can be comprised of many different combinations of factors. Those factors include curtailed training activities, unrealistic battle conditions, inability to mimic expected conditions, low morale, insufficient resources (human or equipment), unbalanced resources inappropriate for global situations, inability to keep up with a stochastic political environment, as well as the compounding of these across the hierarchy within the national security structure. Urbanization contributes to each and all of these risk factors and for this reason cannot and should not be discounted, nor should the effects of which be assumed uneven or unequal in distribution across installations or throughout the national security hierarchy.

TYPES OF MODELS/MODEL ENVIRONMENT

In their discussion of various model types, Parker et al. (2002) observe that the complexity of the systems being modeled means that they are characterized by many interdependencies, various forms of heterogeneity, and hierarchical nested structures. The result of this complexity is that there are many possible stable states in the system and the path that a system takes and the state that it reaches are highly dependant on the initial state and the sequence of choices made along the way. Significantly different outcomes can result from small variations in the input. This, along with the number and variety of different problems to be addressed, makes the development and testing of a model extremely difficult. Consequently there is an equally wide variety of tools and approaches used to build models.

Models are often classified by the techniques used to implement them, but in selecting a model we need to consider a number of other ways of classifying models since the implementation technique alone is not sufficient to guide our choice. In fact, in many ways it is the least important attribute of a model since we are most concerned with *whether* it meets our needs more than *how* it meets them. Here are a number of different categorizations of interest:

- by purpose
- by developer
- by generality
- by scale
- by discipline
- by uncertainty handling
- by technique

Purpose

In considering the attributes of various models, Couclelis (2000) makes the important distinction between models that are developed for research purposes and those that are developed for policy

purposes. Models developed for research are primarily concerned with the novelty of the technique used and with scientific rigor in the choice and evaluation of the methods used. The most important aspects of the work are in their explanatory and predictive power, not their usefulness in producing good public policy (though some models do strive for this as well).

Models developed for use in producing good policy tend to put less emphasis on novelty or even devalue it since credibility may rest on the perceived past performance of a model. Two important considerations in models used for making policy is whether the method used by the model is transparent to the users of the model so that they feel that they understand how and why it arrives at a particular outcome and whether users can manipulate the behavior of the model themselves. Most of all, these models must include variables that can be manipulated by policy makers. For example, variables like local topography cannot be manipulated through policy while things like zoning can.

Developer

Closely related to purpose is the nature of the model developer. Research models are most often developed by academics and therefore generally lack the support and maintenance provided by commercial products. Academic products often (though not always) have more current, more sophisticated algorithms at their heart, but they often lack the sophisticated user interfaces that commercial products may have.

Generality

Most models begin their lives addressing the specifics of a single problem at a single site. Over time they may be generalized to multiple sites and/or multiple problems. They may also be originally designed with the intent of being generalized, but have varying degrees of difficulty in actually achieving that generality.

Scale

Models can operate at a wide range of temporal and spatial scales. They can operate over any time step and look-ahead duration that the developer chooses, for example daily to yearly to decades. These step sizes and look-ahead durations are particularly important when different types of models must be mixed and each has a different time scale of interest. The same is true for spatial scale and resolution. Here the model may be concerned with information at the level of a single pixel or a single lot or a city or a region, etc.

Moreover, the model may use input at one or more scales (temporal and/or spatial), operate at another, and output at yet another. The choice of scale is important to the accuracy of the model and to our perception of its accuracy. For example, if one or more components of the model operate at a coarse scale but we measure the models predictions at a fine scale, it may not have difficulty meeting our expectations.

Discipline

The most significant aspect of categorizing models by discipline is whether the model can accommodate more than one discipline, whether its in the input and/or output or in its inner workings. For example, land use, economics, and environmental considerations are all

intimately interrelated but have often been modeled separately. The EPA review (2000) discusses four different types of planning models with their own histories and purposes: land use, transportation, economics, and environmental impact. In turn, each of these has many sub-disciplines like groundwater modeling, habitat modeling, etc. Recently, more model developers are attempting to integrate multiple disciplines in a single model, but this is a significantly more difficult task both in terms of the knowledge and data required and in terms of the complexity of the software.

Uncertainty handling

The complexity of the modeling problems of interest means that the outcomes of the models exhibit a great deal of uncertainty. The conflicting assumptions and objectives of users along with the variation in quality and availability of data contribute further to uncertainty. This uncertainty is often ignored or not expressed, but it is important to the choice of model and to how the model is used. Lindley (2001) discusses the current trend toward modeling environments that present multiple alternative scenarios rather than single forecasts as their output. This acknowledges the uncertainty and it allows users to explore a range of possible outcomes with the knowledge that the system is unlikely to correctly predict the true outcome in a single try or with a single set of assumptions.

Another aspect of uncertainty is controlling for the types of errors the models make. For example, a model may be parameterized to be more or less conservative in making certain types of errors such as whether a parcel will become developed or not. This may be important in cases where there is a serious problem if the site does become developed and the model fails to predict it whereas there may be less concern over the model predicting development where it actually doesn't occur. Models can address these issues through using cost-sensitive algorithms where the user weights the importance of different kinds of errors and through output of probabilities of outcomes such as development.

Technique

Parker et al. (2002) gives a good summary of a number of different modeling techniques including mathematical equation-based, statistical, systems theoretic, cellular automata, agent-based, and hybrid systems. The equation-based models look for a static or equilibrium-based solution while the statistical models use regression to estimate a model from known historical parameters.

Models based on general systems theory assume that systems can be recursively decomposed into smaller subsystems and the flows among the components. This can be done either to describe the internal structure of the system in a static way or to be run dynamically to forecast future outcomes. These models rely on modeling the structure of the system itself rather than on historic data (Couclelis, 2000; Agarwal, 2000).

Cellular automata models have become very common in the last ten years. These models operate at the level of a single unit (e.g., pixel or parcel) and use transition probabilities to describe the change of that unit to a different type (e.g., rural to urban). The transition probabilities are usually derived from historical data (though some modelers have found better performance when

the historical data is not used to set the probabilities (Jenerette, 2001). They are based on the assumption that the system can be described through these local interactions. Since new urban development often spreads by adding at the boundary, these neighborhood processes are often effective in describing the general land use dynamics. These models have the advantage of being particularly easy to describe in software, however they are not well-suited to long-range predictions where seeds of new urban growth occur in areas away from the current edge of urban growth.

Agent-based modeling is an active area of current modeling research. Rather than basing on the model primitive units and their neighborhood interactions, agent-based models are built from various entities whose interactions with each other are described by the model. These agents may represent homeowners, policy-makers, industries, etc. Each has a description of the kinds of things that it can do and who and what it interacts with (e.g., other agents and the physical environment). The agents then interact with each other and the environment to produce the model outcome(s). One of the main reasons for interest in this area is that it allows modelers to represent human decision making processes in the model, which is difficult to do in models based on neighborhood interactions or mathematical techniques like regression.

Hybrids of all of these different techniques are another approach to the modeling problem. Since all of these modeling techniques have their own advantages and disadvantages for different situations and since constraints on the users of the system may force the use of existing models, combinations of the techniques may be either or more effective or simply forced on the modeler.

One method for hybridizing models is through the use of modeling frameworks like DIAS (Christiansen, 2000). The advent of object-oriented programming has made it simpler to build wrapper structures that can encapsulate the behavior of a given model and provide a protocol for communication with that model. Frameworks and modeling languages can also provide direct support for basic modeling activities like discrete event simulation as well as display of and interaction with model output.

EVALUATION CRITERIA FOR MODELS

Once we have established the questions to be answered using the model, there are numerous criteria to use in evaluating the many different models that exist. In the early 70s, Lee (1973) gave a short list of criteria that is often cited by modeling researchers:

- transparency
- robustness
- reasonable data needs
- appropriate spatio-temporal resolution
- inclusion of enough key policy variables to explore policy questions

This was the list of what should be considered when evaluating a model for suitability of use. Since then, this list has greatly expanded. Our needs appear to be more sophisticated, and while this may or may not be true, certainly our modeling capabilities are more sophisticated. We will list these criteria under the broad headings of:

- Capability
- Resources
- Integration
- Credibility

Some of these criteria are taken from the three reviews mentioned earlier (EPA, 2000; Agarwal, 2000; Parker, 2002), while others are ones that we have added.

Capability

The capability of a model is an important factor, and particularly pointed one. That is, can the model answer the questions of interest in the way that is required? It may not be useful to select a model that gives output in a form that cannot be assimilated, understood, or interpreted as needed to address pertinent questions.

Relevance

- Does the purpose of the model match the questions that the base is trying to answer? For example, does the model predict land use change or environmental changes like habitat or groundwater, etc.

Scenarios

- Does the model generate multiple scenarios or a single forecast?

Spatial resolution and extent

- What is the smallest spatial unit that the model operates over?
- How much space is covered by a run of the model? For example, a city, a region, a watershed, etc.

Temporal resolution and extent

- What is the smallest unit of time that the model operates over?
- What are the time periods that the model can make projections over?

Landscape features

- What constituents of the map/future can and can't the model produce?
  - Roads: Many (if not most) models cannot generate new roads. They have to be entered by the user. Since roads are often used as driving variables for where development will occur, this limits the type of output that can be generated 10 or more years out.
  - Land use distinctions: Can the model distinguish only urban/not urban or can it handle multiple subclasses of urbanization (residential, commercial, etc.)?

User interface

- How difficult is the program to parameterize and run?
  - Is there a GUI interface or does it use a command line interface or batch control file?
- Are there multiple levels of complexity to the interface?
  - Some systems (like those in Lindley, 2001) have a group of pre-set scenarios as well as

full access to all of the system parameters. This allows novice users to interact with the system as well as experts.

- Is there visualization support for outputs (instead of tabular output)?
  - maps
  - statistical plots

Public Accessibility

- Can the model be run interactively in public?
- Are the results in a form that is comprehensible to the public?
- Can the public run the model themselves to try things out (e.g., over the web)?

Speed

- How long does it take to generate outputs?
  - Can it be done at a public meeting, i.e., real time?
  - Is the resolution of the model so fine that it takes a prohibitively long time to do a model run over a large enough area to be of interest?

Versatility

- Can the model project values for multiple types of variables (i.e., land use, economic, environmental, etc.) or does that require running separate models?

Bias

- Is the model able to handle very different environments equally well or is it biased toward one situation over another. For example, does it handle rural environments and urban environments equally well?

Resources

What resources are required to use and maintain the model?

Cost

- What labor costs are there?
  - model developers
  - consultants
  - local staff
  - programmers

Hardware costs

- What level of new hardware is required?
  - Disk storage
  - Processors
  - Memory
  - Networking

Software costs

- Does the model require users to program some of it themselves?
- What is the cost of associated software for the model?
  - model itself
  - other required software (compilers, GIS, databases, statistical software, etc.)

Technical Expertise

- Does the user have the technical expertise required to install, use, extend, calibrate, and interpret the results of the model?
- What types of expertise are required?
  - planning
  - data (GIS, remote sensing, etc.)
  - programming
  - military
  - policy

Data Requirements

- Does the user have, or can they obtain, the data necessary to run the model?
  - At the required spatial and temporal scale
    For example, is aerial or satellite imagery available at the resolution required by the model?
  - At the level of spatial and temporal coverage required
    For example, is there coverage for all years needed for input?
  - Is data collection or compilation required?
    For example, collecting data on building permits issued for the near future
- How accurate is the user's input data?
  - For example, are the base maps accurate?
- Is data publicly available or is it proprietary?
- Is the kind of input data necessary available?
  - Difference questions (dust, noise, etc.) need different kinds of input information to answer them.
  - Different models may answer the same question using different kinds of input data.
- Is historical data available for validating the model?

Model Support

- Models are complex and generally require support for use and/or for installation, maintenance, integration, and validation. This may mean support from any or all of the following sources:
  - vendor
  - documentation
  - books and journals
  - consultant
  - academia
  - SERDP
  - base personnel

    - user groups
- What is the user base of model?
    - Have enough people used the model to provide a base of expertise to draw on, e.g., in papers about the model or in finding users to talk to?
    - Is the model standard enough to have developed add-ons or modules from sources outside the original developers?

Integration

Assuming that a model is capable of answering the questions of interest, can it be integrated under the constraints of the existing infrastructure?

Compatibility with existing local systems (both base and community)

- software
    - operating system (Windows, Unix, Macintosh)
    - support software (data bases, GIS, etc.)
- hardware
    - computer (PC, Unix workstation, Macintosh)
    - networking
    - display
- expertise
    - Are local personnel already familiar with software and hardware required for the model (e.g., the GIS used by the model)?
- data
    - What format, resolution, and types of data do the base and the community track currently?

Linkage

- Can the model be linked to other models?

Transferability

- Does the model apply to more than one site?
- If it applies to more than one site, how much effort is required to adapt to a new site? For example, does it require new programming?

Extensiblity

- Can new functionality be added to the model by end users?

Credibility

Are the model's outputs accurate and believable?

Accuracy

- How reliable are the model's outputs when measured against known historical data?
    - Specific spatial outputs (e.g., the land use class of a specific parcel)
    - General characteristics of the model's outputs (e.g., landscape metrics like fractal

dimension)
- How accurate is the model across multiple projects vs. its performance in a single location?

Robustness
- How sensitive is the model's performance errors in input data?

Uncertainty
- Does the model quantify or even express its uncertainty in any way?
- Can the model give probability estimates for things like a parcel's land use classification?
- Can the model bias its actions toward avoiding some kinds of errors more than others (e.g., Type I and Type II errors)?
- Does the model account for uncertainty inside its own workings?
- Can the model output multiple scenarios?

Transparency
- Do users and the public have a way of explicitly knowing what the assumptions, mechanisms, and parameter settings of the model are?

Theoretical grounding
- Is the model well grounded in both the computational techniques that it uses and in the domain knowledge it uses (e.g., planning and environmental knowledge)?

Previous use
- How much has the model been used in other real-world situations?

SPECIFIC MODELS

In reviewing the land use change literature it becomes immediately clear that there are hundreds of papers and models. In their review, Agarwal et al. (2000) discuss narrowing their search from 250 relevant citations to 136 possibilities and then choosing 19 models to review. Even then, they only review four models that are also in the EPA review of 22 models (EPA, 2000): CUF-California Urban Futures, CURBA-California Urban Biodiversity Analysis, LUCAS-Land Use Change Analysis System, and SLEUTH (formerly known as the Clarke Cellular Automata model). Moreover, neither of these reviews considers any agent-based models, nor do any of them even mention the DIAS system that we were asked to consider.

This extreme fragmentation of the literature suggests that two things. First, there is likely to be lots of duplication in all of these models since there is so much literature to examine before someone decides to build yet another model. Second, it suggests that at this point, we are likely to have missed a number of models that may be of interest.

Since we have yet to precisely identify the requirements for any model that we may choose, in this section we will briefly discuss the two models that we were asked to review, but not go into any detail or review any other models. There is no point in duplicating the thorough reviews that we have already cited and more importantly, with so many models to consider, there is no

point in examining any of them in detail until we know what our requirements are.

There is one important thing to note in the following discussion: both LUCAS and OO-IDLAMS simulate land *cover* change (in the sense of vegetation types), not land *use* change (in the sense of residential, commercial, etc.). This makes them suitable for ecological modeling, but not for modeling the growth of urban areas.

LUCAS – Land Use Change Analysis System

LUCAS is included in two of the detailed reviews mentioned earlier (Agarwal, 2000; EPA, 2000), so we will only briefly characterize it here and discuss some of its strengths and weaknesses.

LUCAS is a land use change model developed to examine the effects of land use on landscape structure in watersheds (Berry, 1996). The model consists of three modules: socioeconomic model module, landscape change module, and impacts module which characterizes ecological effects of land cover change, for example on habitat. In particular, it has been developed for and applied to the Little Tennessee River basin in North Carolina and the Olympic peninsula in Washington state. LUCAS is intended to be useful beyond these two locations, but the distribution web site (Tennessee, 1999) says that the "LUCAS software provided will only simulate land cover change on a small region in western North Carolina. It will not provide any relevant information for other geographical locations."

The model appears to use watersheds as its fundamental frame of reference and models the land use in terms of land cover (vegetation type) and ownership (private, public). It uses historical land data to parameterize transition probabilities for individual pixels or patches from one state to another and then runs stochastic simulations using these probabilities. It allows the user to specify that more than one run be made from a given set of probabilities so that multiple outcomes can be compared.

Advantages and Disadvantages

The model allows multiple scenarios, incorporates socioeconomic factors in its model structure, models the effects of landscape change on the environment, and a lot of work has gone into making the interface user friendly. It also makes use of the non-commercial GRASS GIS which means that users do not have to purchase a commercial GIS to use the package. It provides landscape metrics that characterize the landscape structure in terms of patch characteristics rather than just the total number pixels of a particular type. It can also compute transition probabilities and model at the patch level rather than just the pixel level.

From the documentation available on the web site, it appears that installing and using the software elsewhere requires a knowledge of Unix, C++, and the GRASS GIS as well as expertise in developing parameters for the Socioeconomic module of the model and ecological parameters for the Impacts module. It is not a commercial product and is meant to be used in conjunction with researchers and would take considerable expertise to install, parameterize, and program for a particular location. Moreover, it depends on the use of GRASS, which is also not a commercial product and has its own bugs and lacks in documentation.

The fact that the model operates at the watershed level means that it is not clear whether it is suitable for modeling outside of a watershed context (e.g., adjacent to a large metropolitan area or in the Mojave desert) nor is it clear how difficult it would be to adapt it to a non-watershed context. More importantly, it deals with land use in the form of vegetative cover and ownership type rather than the various classes of urbanization such as residential and commercial. If the installation intending to use the model is in a rural watershed environment where this is not an issue and the questions revolve around environmental impacts, then this may not be a problem. However, in situations where the goal is to project particular types of urban development instead of vegetation cover, LUCAS may not be suitable. Again, there may be some way that it can be made to work in that environment, but it is not clear how or what amount of effort would be required.

One last observation is that none of the publications about LUCAS mentioned anything about whether and/or how uncertainty is characterized in LUCAS or what kind of validation results are available for LUCAS in either of the environments where it has been tested.

DIAS / OO-IDLAMS

IDLAMS (Integrated Dynamic Landscape Analysis and Modeling System) is a model-integration framework for simulating ecosystem dynamics in natural resource management, particularly in a military environment (Li, 1998). It consists of four major modules: vegetation dynamics, wildlife habitat suitability, erosion, and scenario evaluation. It was designed for easy use by land managers and was written in C. Like LUCAS, it was also built on top of the GRASS GIS.

To allow for greater flexibility and adaptation, pieces of the model were rebuilt in an object-oriented version called OO-IDLAMS based on the general purpose modeling framework DIAS (Dynamic Information Architecture System) (Sydelko, 2000). Rather than continuing to rely on GRASS, which is not object-oriented, OO-IDLAMS uses an object-oriented GIS as well as the DIAS framework for interaction.

DIAS simulations consist of objects and classes of objects that represent real-world entities along with models that simulate the dynamics of the interactions of the objects. In particular, DIAS allows for the incorporation of multiple external models with different interfaces and data demands by putting wrapper software around each of them. This wrapper software specifies the protocol for interaction and exchange of data among the objects and models. By turning the various pieces of the system into black boxes whose internals are unseen by other parts of the system, how the pieces are implemented no longer matters. All that matters is that the pieces all provide the services that they are supposed to provide and share data in the prescribed manner.

OO-IDLAMS is just one particular model that could be developed inside DIAS. It simply demonstrates how various parts of the system could be built using DIAS. DIAS is meant to be a general tool for all kinds of simulations. In fact, OO-IDLAMS does not implement the full functionality of the original IDLAMS package that was developed for Ft. Riley, Kansas.

Advantages and Disadvantages

The original IDLAMS package has modules that are of use to ecological simulations on military bases. However, like LUCAS, these kinds of simulations are not the only kind that need to be done. There is no provision in IDLAMS for urban growth modeling since it is meant for ecological modeling. Moreover, the documentation only describes its use in the Ft. Riley situation and gives no indication what, if any, of the Ft. Riley knowledge is useful in any other situation. The manual for the program simply says that the Ft. Riley vegetation dynamics model can be used and the C code modified for new situations. The OO-IDLAMS version has re-expressed the vegetation dynamics model inside DIAS, but again, there is no explanation of what, if any, of that knowledge is transferrable to a new site.

Since DIAS has been developed as a generic object-oriented modeling framework, it provides the structure for many of the services needed in writing a model, whether it is for vegetation dynamics or for urban growth. Therefore, it may be very useful in developing new land use change models, however, it looks like those kinds of models would have to be built from scratch inside this framework since the existing OO-IDLAMS modules are not oriented towards urban growth and are specific to Ft. Riley. If there are no satisfactory models in the literature or commercially available, then DIAS should be considered as a possible framework for building a new land use change model.

PROBLEMS AND ISSUES

Social Issues

Some of the issues involved in solving land use change problems are not necessarily technical ones. Social issues are one example of factors that are affected by the choice of technical solution and they primarily relate to how the purpose of the model is defined. If the purpose of the model is simply to project what kind of development is likely to happen, then a very different model can be used than in the case where part of the intent of the model is to provide a vehicle for public participation and communication. How a product, model, or program will be perceived is an important consideration. Those that are more extensive in scope, effort, and cost, might certainly be candidates for active participation throughout the development process. Without participation in this manner, implementation can and usually does, fail. Few people are amenable to having a product, program, or model forced upon them without their input, particularly when the developers are not considered "locals" to the geographic area. Social issues will be an important consideration as the reviewer reads this document.

In general, future scenario models are not run in a vacuum. They are run in the context of local policy and interactions with the local community and local policy makers. The fact that it is so difficult to predict the future means that such models also operate in an environment of great uncertainty, which all would attest to at many levels. Therefore, these factors suggest certain things about the choice of model, regardless of the technical pedigree of the model. This is an important point to consider. It does not matter how solid, appropriate, or accurate a model is. If there is no buy-in by those who would use, and theoretically benefit from its use, implementing or using that model is futile. Results and recommendations based on results will not be accepted, will not be implemented.

To illustrate this point, consider that there exists significant uncertainty in model outcomes, which suggests that a model that will produce multiple alternative futures rather than a single forecast may be both more useful and more credible. If multiple outcomes are allowed, then users can experiment with different assumptions about parameters of the model and directions of public policy, as in the FutureQUEST model for the Northwest UK (Lindley, 2001). Since model outcomes are unlikely to be exactly correct, (King, 1993) suggests that one role of the model is to give useful information about what *not* to do.

A second consideration is that the model may provide a vehicle for increased interaction and cooperation with local community and policy makers. A model provides a focal point for discussion and consideration of alternatives and assumptions. In small communities lacking the financial means and the technical expertise to support their own modeling efforts in planning, military bases may have an opportunity to foster goodwill and promote cooperation through the sharing of model technology and data. Maintaining an ongoing interaction with local communities through a shared model and shared data may also enhance the ability to avert problems before they occur, as each participant in the process is made aware of others' intentions. In this manner discussion can occur over time, throughout the life of a project and beyond, resulting in an outcome of better understanding based on communication. This kind of cooperation would also help reduce one source of model error by way of ensuring input data for the model remains current and correct due to shared data sources. Although it is often assumed that the military will have the most recent and most accurate data, this is not always the case.

The issues involved in solving problems perceived by a community are not necessarily technical in nature, but may be affected by the choice of the technical solution. It is important therefore to be able to show different alternatives and explain the rationale behind assumptions to community members, local planning staff, policy makers as well as military officials and policy makers. This creates an opportunity to promote the necessary cooperation, discussed above. Cooperation with the surrounding community is imperative for successful operation of military installations. The ability to not only provide the community with technical support and expertise, but to articulate and advertise the benefits of such a relationship serves in the best interest of the military as well. In doing so, by creating a relationship with external agencies at various levels throughout the surrounding communities, the ability to share common data and have access to information that will update existing military data benefits the military. Overall, costs are reduced and data, as well as resulting modeling accuracy, increases.

In short, choosing a model requires paying attention to the fact that models do not exist alone. McEvoy (2001) sums this up well as "the aspiration of putting together modeling, monitoring, appraisal and implementation" into one overall framework. It is a challenge, but it is not an insurmountable challenge.

Technical Issues

Knowledge shared by all models

The modeling literature primarily focuses on the modeling techniques themselves, but all of the models embody knowledge of and assumptions about the field that they are modeling (e.g., land

use change, hydrology, etc.). This raises the question of which differences in model outcomes are due to the modeling technique and which are due to the underlying knowledge embedded in the model?

For example, if we run several different models using the same input data (assuming that they can use the same input data), how similar are the results? And if those results are different, to what extent are the differences due to the modeling technique used vs. the knowledge used in the model? Or are the differences simply the result of improperly set parameters? Can the model parameters be tuned to produce similar results?

A second and more important question is to what extent can this knowledge be reused in different modeling techniques? If we can somehow treat the knowledge as data separate from the modeling techniques themselves, that allows us to compare models more equitably and it allows us to develop that knowledge independently of the development of modeling technology.

In trying to answer these questions we might ask whether we can even identify what knowledge is commonly agreed upon by model developers and users. This is not at all clear from reading the literature. For example, while roads and population are common elements of land use change models, Parker et al. (2002) say that "there is still a fair deal of uncertainty and disagreement in the literature on the relative contribution of phenomena such as roads or population to land-use/cover change."

This uncertainty raises several questions for those trying to choose a model for a particular application. Is there *anything* that is commonly accepted among all of them? Are some things accepted but the parameters of their form not agreed upon? Do models break into different classes that agree on the form of the relationship but not necessarily the parameterization?

One of the most straightforward ways for identifying some of the knowledge included in models is to look at lists of the driving variables that they use in generating patterns and preferences. A first step toward codifying the knowledge embedded in the models is to assemble lists of these factors. We found two such lists in the model reviews:

Agarwal et al. (2000) derived a list of some variables that characterize relevant human drivers from two government reports (NRC, 1992; Redman, 2000). Some of these variables were derived from a more global perspective and have less utility in a strictly U.S. setting.

- population size
- population growth
- population density
- returns to land use (costs and prices)
- job growth
- costs of conversion
- rent
- zoning
- tenure
- relative geographical position to infrastructure

    - distance from road
    - distance from town/market
    - distance from village
    - presence of irrigation
- generalized access variable
- village size
- silviculture
- agriculture
- technology level
- affluence
- human attitudes and values
- food security
- age

Lindley (2001) gives the following list of inputs to the FutureQUEST model. Since this model was developed in the context of planning for the northwest UK, some of these variables are also not as suitable for U.S. use. However, this model was heavily oriented towards user-interaction and may warrant further investigation as something that can serve as a template for developing a suitable tool for DoD's purposes.

- General development issues
    - population density
    - land use
    - clustering
    - settlement size
    - proximity to settlements
- Social issues
    - crime risk
    - proximity to heavy industry
    - social deprivation
    - population change
    - rural amenity
    - health
    - derelict land
- Economic issues
    - business service ratio
    - employment change
    - labor skills index
    - average earnings
- Accessibility factors
    - urban hub access
    - airport access
    - road network access
    - rail network access
- Capacity factors

- · derelict land
- · vacant land
- · buildings for redevelopment

- Land use constraints
  - · rural protection
  - · agricultural class
  - · flood/other natural hazard risk
  - · topographic effects
  - · green/brown field land
  - · Policy incentives including NS/EW development axes and EU structural programs

Given the wide variety of installations whose needs must be addressed, it may be important to identify the shared knowledge among models and find ways of treating that knowledge as data instead of as program code. If this is possible, it may make it easier to build and update models with less programming and to maintain consistency and quality in the knowledge used.

One possibility would be to consider encoding as much knowledge as possible using a data description language such as XML (eXtensible Markup Language). XML is a data description language that makes it easier to exchange data among applications by allowing its users to describe the structure of the data in a form that is independent of any particular application. We are aware of one ecological modeling framework that is incorporating XML into its structure, the Object Modeling System (USDA-ARS, 2002). Given the huge number of models in the literature, there may be land use change models that also incorporate something like XML, but we have not found them. However, XML is not the only way that knowledge can be separated from modeling technique and there may be models that do make this separation explicit in some other way.

MODEL VALIDATION METHODS

One important aspect of choosing a model is articulating what the criteria are that are used to evaluate the performance of the model. Evaluating a model's performance is often referred to as model validation although the term "validation" has generated controversy among ecological modelers (Goudie, 1997; Rykiel, 1996; Vanclay, 1997). Many view the word validation as implying that a model can be proven correct even though models can never be proven correct for all situations. Since it only takes one counterexample to demonstrate incorrectness, models can be shown to be incorrect, that is, *in*validated. This leads to semantic arguments about the correct terms to use, such as verification, evaluation, testing, checking, etc. We will use the term validation here because that is the term that leads to the literature on evaluating models, regardless of its absolute correctness.

There is a large body of literature on validating ecological models and a good review of that literature is found in (Rykiel, 1996). He states that model validation has two important parts. First, before the model is evaluated, three things must be specified: the purpose of the model, the evaluation criteria for the model, and the context in which it will be used. Evaluating the model can only be done against this background and consists of evaluating three different components: its operation, its theory, and its data. He suggests that the most common problem with model

validation is a failure to explicitly state the evaluation criteria.

There are many different ways that models are evaluated (Law, 1991; Mayer, 1993; Power,1993) and this is perhaps best illustrated by a list of different techniques that Rykiel has derived from (Sargent, 1984):

- *Face validity* - Does the model output seem reasonable?
- *Turing tests* - Can a knowledgeable user tell the difference between model outputs and real data?
- *Visualization techniques* - Visual examination of goodness of fit of a model to time series, etc.
- *Comparison to other models* - Does the model behave like other models?
- *Internal validity* - Does a test set produce consistently similar output in a stochastic model?
- *Event validity* - Qualitative evaluation of the model's ability to reproduce relationships among variables rather than their specific values.
- *Historical data validation* - If existing data is split into a training set for calibrating the model and a test set for evaluating the output of the model, how well does the model reproduce the test data?
- *Extreme-condition tests* - How well does the model behave outside of the normal range of input parameters?
- *Traces* - How well does the model follow the expected behavior of specific variables throughout the course of a run?
- *Sensitivity analysis* - Is the model sensitive/insensitive to changes in the same parameters that the system is sensitive/insensitive to?
- *Multistage validation* - Apply various validation methods at different points in the model's development: design, implementation, and operation.
- *Predictive validation* - How well does the model predict system behavior that occurs after the model is built?
- *Statistical validation* - Does the model output have the same statistical characteristics as values observed in the system being modeled? Are the errors in output variables within the limits specified in the evaluation criteria?

Many of these tests incorporate some kind of statistical analysis of the model output. Statistical techniques used in these validation methods are often the common statistical tests of significance, but Bayesian methods are also becoming more common (McCarthy, 2001) since they allow the expression of relative confidence in models rather than simple acceptance or rejection.

Another useful technique in the context of maps of urban vs. non-urban development is the ROC curve (receiver operator characteristic) (Brooker, 2002; Provost, 1996). These curves plot the number of true positives on the y-axis vs. the number of false positives on the x-axis to indicate the tradeoff in accuracy of finding all occurrences of a given class (e.g., urban) vs. the number of errors in falsely identifying some other class as the desired type (e.g., calling non-urban pixels urban). This is useful because it is often the case that as the classifier is tuned to find more and more of the urban pixels, it will also classify more and more non-urban pixels as urban. Plotting the ROC curve shows this tradeoff and allows model users to use cost-sensitive techniques based on the relative importance of finding all of the urban pixels (or whatever type is of interest). Moreover, a single measure of the tradeoff can be seen by measuring the area under the ROC curve.

Land use change models have an added component of difficulty in model validation space. While there are many papers on model validation, very few of them discuss the particulars of validating spatial models. Spatial models are even more difficult to validate because the notion of the similarity of two spatial objects is more difficult to quantify than the simple differences between two variables such as the number of correctly labeled pixels. This makes it difficult to determine whether the map generated by a land use change model is similar to the map of what really happened.

Spatial accuracy assessment is also a problem in remote sensing and there has been significant work on assessing the accuracy of classified images (Congalton, 1991) but with little regard for the spatial arrangement of the classifications. However, some work is being done on accuracy assessment for the polygons generated in classified images (Beauchemin, 1997) and it may be of use in spatial model validation as well.

One of the few papers that explicitly discusses spatial model validation is (Turner, 1989). In this paper, they go beyond simple counts of agreement in classified pixels to try to characterize the overall spatial characteristics of the map using measures from landscape ecology and information theory (e.g., fractal dimension, contagion, spatial predictability, etc.). They also discuss a multi-resolution measure of goodness-of-fit due to Costanza (1989). This is important to consider since the notion of correctness of a model output may be different at different scales.

In summary, evaluating the performance of the models considered is an important task in choosing a model to use, but there is no simple consensus on the best way to do it, particularly for evaluating spatial correctness of model output. This is an area that will require more attention in future work on this project.

REQUIREMENTS ANALYSIS - QUESTIONS FOR DOD

Before it makes sense to try to select a particular model, there is certain information that the model chooser needs that is currently lacking, that is, the basic requirements for the system. At this point it is not clear how many different kinds of questions need to be answered by models or what are the contexts in which they will be used.

Because of this uncertainty, it is also unclear whether a single modeling framework would serve

DoD's needs or whether a number of different models are required. What is needed is to figure out how many different kinds of situations the model(s) have to address, for example, if all bases had the same data and the same data format and the same political environment and the same questions to answer, then only one modeling solution would be necessary. This is clearly not the case, but without further research we have no idea how many different situations need to be addressed and how different they are. For example, are there numerous situations but they're all gradations of the same kind of problem and can be handled by locally reconfiguring parameters or components within a single modeling framework or are the problems qualitatively so different that no single framework can accommodate all of them and a family of models is required?

The most important thing that is missing right now is the set of questions that need to be answered. Without a thorough understanding of what the questions are, choosing a model is left to comparing lists of features that may or may not have any impact on the utility of the model for answering particular questions. The next thing after the question is to know the context in which it is to be answered. Third is the quality of answer required and the cost of achieving it. Given all of these things specified across the military, we can then break the bases, questions, and contexts into classes of problems that have similar attributes in terms of the kind of modeling solution required.

Below is a list of the pieces of information required for each installation:

- *problems/questions influenced by urban growth* (dust, noise, etc.)
- *size* (physical and personnel)
- *location* (Southwest, etc.)
- *physical environment* (desert, urban, etc.)
- *expertise/support* (on base and off)
    - · planning
    - · economics
    - · computer
    - · etc.
- *existing computer*
    - · hardware
    - · software
- *data*

Do they already have the data they need to answer the question(s) of interest or do they need to collect it? Furthermore, will there be new data anticipated to be acquired, existing data retrieved, or a combination of both types. If data need to be acquired considerations include:

- · format
- · quality
- · quantity
- · span (in terms of time and space)
- · any standardization?

- *neighboring community type* (small rural, large city, etc.)
- *level of need*
    - · perceived degree of threat from urban growth

  - · cost of failing
    - for example, tiny base vs. big one
    - ratio of the two, big city base they know is surrounded so it doesn't matter vs. rural supersonic
  - · current costs
    - doing nothing
    - doing something
- *level of local cooperation/interaction/coordination*
  - · hardware, software, and knowledge
- *purpose of model*
  - · public interaction
  - · internal use only
  - · experts only
  - · etc.
- *existing models in use*
  - · at the installation
  - · locally in the community
- *installation plans*
  - · activities
  - · expansion/contraction
- *current setup vs. future expectations for all of the above*
- *how useful are model outputs given their cost and uncertainty*
- *degree of accuracy required*
- *time scales*
  - · length of projection (10 years, 50 years, ...)
  - · time step (yearly, etc.)

When these questions have been answered, then the process of making a good model choice begins.

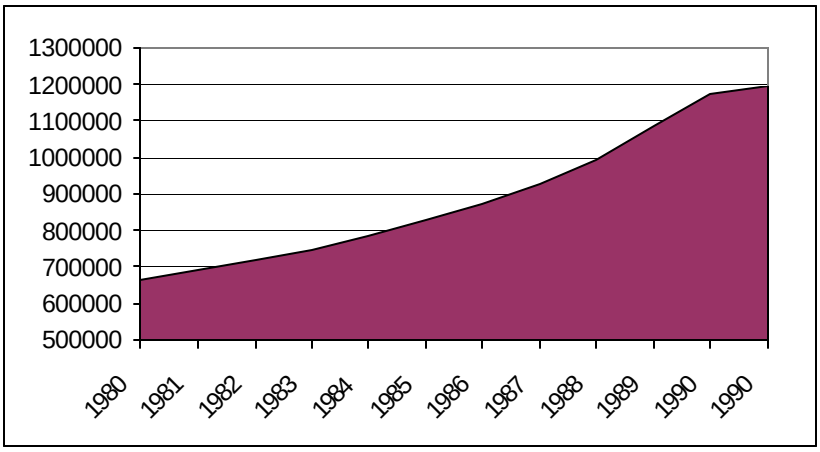
PILOT STUDY: STATISTICAL ANALYSIS AS SOLE BASIS FOR MODELING

Study area – Coachella Valley

Demographics – Riverside County

California is the most populous state in the US, containing 12 percent of the US population in 2000. California's population is expected to increase, and the state is expected to have 15 percent of the Nation's population by the year 2020. Riverside County is of particular concern to Joshua Tree National Park because of the rapid rate of development and projected population increase. Nearly five percent of California's population resides in Riverside County. Figure 1, below, shows the increase in population for Riverside County between 1980 and 1990. Between 1980 and 1990 the population in Riverside County grew from 663,199 in 1980 to 1,170,413 in 1990. In the following decade the population increased from 1,170,413 people in 1990 to 1,545,387 people in 2000. This represents an approximate 32% increase in the number of people living there in the last decade and follows the trend of the previous decade.

Figure 1. Population growth in Riverside County for 1980-1990.



Within Riverside County, the Coachella Valley (Figure 2) is a destination resort comprised of nine incorporated cities. Figure 2 shows the breakdown of population by each of these nine cities, which total 255,790 people or 17% of Riverside County's entire population.

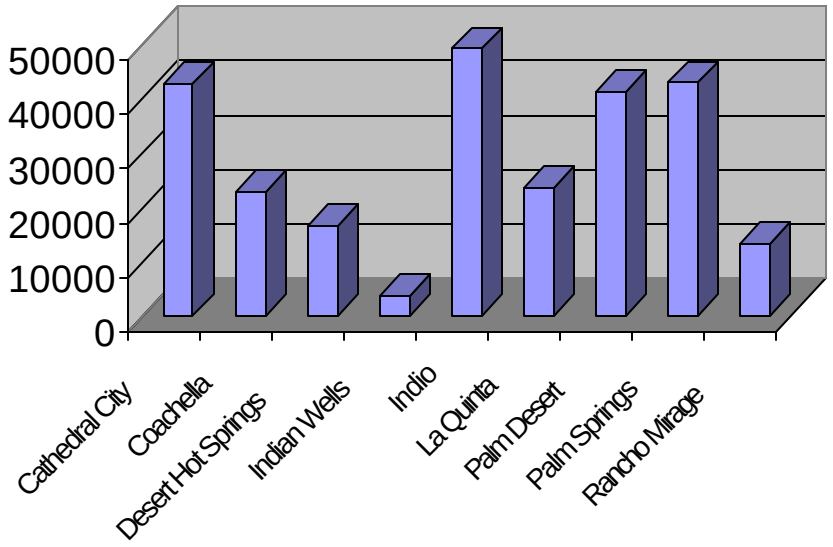


Figure 2. Population by city in the Coachella Valley.

Perhaps the best known city within the Coachella Valley is the resort destination of Palm Springs. Palm Springs is located near the San Andreas Fault, which passes through the middle of the Coachella Valley. North of the San Gorgonia Pass, which brings traffic into the Coachella Valley from Los Angeles and other metropolitan areas, are the San Bernardino Mountains, a popular destination for hiking, camping, fishing and skiing during winter. The San Bernardino

range, which reaches heights of 10,000 feet, is the easterly extension of the San Gabriel Mountains. Because the pass between these two mountain ranges channels and intensifies the prevailing westerly winds at the head of the Coachella Valley, it has been populated with hundreds of large wind generators.

Data

The analysis presented here was based on work in the Mojave Desert funded by the Department of Defense Strategic Environmental Research and Development Program (SERDP) (Cablk et al., 1999; Gonzalez, 2000). Although the same variables used in this study for the Coachella Valley as were developed in the Mojave, the specific probability models were developed based on Coachella Valley data. For this reason, other variables may be cited for incorporation into this analysis at a future date, but were not acquired for the purposes of this study.

Landsat Multispectral Scanner

Landsat Multispectral Scanner (MSS) data were acquired from the North American Land Characterization (NALC) Program. The NALC program was funded by the US Environmental Protection Agency (EPA) Office of Research and Development's Global Warming Research Program (GWRP) and the USGS Earth Resources Observation Systems (EROS) Data Center. The objectives of the NALC project are to develop standardized remotely sensed data sets (e.g., NALC triplicates) for change detection analyses. NALC satellite data are referred to as *triplicates*, three sets of satellite data acquired in the early 1970s, mid-1980s and early to mid-1990s, respectively. Original MSS data have a nominal spatial resolution of 79m but the NALC data are resampled to a nominal spatial resolution of 60m. The MSS instrument has detectors sensitive in four discrete regions of the electromagnetic spectrum from visible green to near infrared. These bands are optimal for detecting vegetation and other biotic landscape features as well as abiotic features such as bare soil, water, or impervious surfaces. While the spatial resolution is somewhat coarse relative to other commercially available satellite data (down to 1m panchromatic) the spatial, spectral, and temporal resolutions of MSS data were appropriate for our large study area and for detecting urbanization in the desert (*from* http://eosims.cr.usgs.gov:5725/CAMPAIGN_DOCS/nalc_proj_camp.html).

Three scenes of MSS data provide complete coverage of the California Mojave Desert. The following table identifies each scene (path and row) and the corresponding image acquisition dates.

Table 1. MSS data used to develop change analysis of urban extent.

1970		
Path	*Row*	*Date*
39	36	7 Aug 72 & 13 Sep 72
39	37	23 May 73 & 9 Jun 73
40	36	29 Jun 73 & 23 Jun 74
1980		
Path	*Row*	*Date*
39	36	10 Sep 86
39	37	6 Jun 86

40	36	10 Jun 85
1990		
Path	*Row*	*Date*
39	36	10 Sep 93
39	37	30 Jun 93
40	36	21 Jun 93

The set of three scenes (not three dates) was pieced together into one large, seamless file that covered the study area. This mosaic was first masked to the project study area to exclude regions outside the scope of the project and then masked again to include only privately owned lands, as federally managed public lands, state, and certain other lands are not available for development. This process generated three seamless images for the entire study area for the mid-1980s and early 1990s, which included pixel data for private lands only, or those lands available for future development. These two data sets were interpreted for extent of urban or other anthropogenic development using spectral and spatial pattern recognition techniques. The 1986 and 1993 scenes were digitized on-screen with a resulting binary classification of two classes: urban and non-urban. The 1970's data were not interpreted due to poor image quality. The minimum mapping unit was one pixel, or 60m. The resulting urban layers were used in the development of models to predict future development to the year 2020. Past and current (i.e. 1990s) development was derived from NALC satellite imagery.

Future development was modeled to the year 2020 based on past patterns of development between 1986 and 1993, population projections from the US Census bureau, distance to existing urban areas, political boundaries, natural landscape features and infrastructure. Population growth between 1986 and 1993 was used over the greater time period of 1973 and 1993 to better capture population growth trends of recent times. Population in Riverside County nearly doubled between 1980 and 1990, but saw only a 32% increase in the last decade.

The resulting model was spatially explicit and resulted in a raster output that retains the original geographical coordinates of the input data. Population predictions, which have no spatial component, were applied uniformly throughout the landscape. Population growth was an estimated density increase of 12.3 people/ha, calculated by dividing the number of people living in urban areas of the Coachella Valley in the early 1990s (224,357 people) by the number of developed hectares in 1993 (18270.7 ha). The number of developed hectares in the Coachella Valley was extrapolated based on proportional area of incorporated cities inside the valley relative to the greater Riverside County. Of the entire county, the cities in the Coachella Valley comprise 16.78% of what is developed. Therefore the number of people living in the Coachella Valley in 1986 and 1993 were calculated based on corresponding percentages. For example, in 1993 the total population of Riverside County was 1,321,304. To calculate the proportional population in the Coachella Valley, this figure was multiplied by 0.1678. The population was assumed to increase in the Coachella Valley at the current rate. The population of Coachella Valley cities was 198,800 in 1990 while the total Riverside population for 1990 was 1,170,413. This means the Coachella Valley population is 16.98% of the whole county by area. Using this percentage, the following was estimates:

Total area developed in 1986 = 89574834.234 m^2 = 8957.5 ha
Population in 1986 = 871209

Proportional population = 871209 * 0.1698 = 147931.3
People/ha in 1986 = 16.5148

Total area developed in 1993 = 182706814 m^2 = 18270.7 ha
Population in 1993 = 1321304
Proportional Population = 1321304 * 0. 1698 = 224357.4
People/ha in 1993 = 12.27963

These calculations assume all development within urban boundaries (>50k people).
Based on these calculations, population densities for the future were estimated as follows:

18270.7 ha in 1993 minus 8957.5 ha in 1986 = 9313.2 new ha of development
Population change between 1986 and 1993 = 76426.1 new people
76426.1 / 9313.2 = 8.2 new people per ha were added to the CV between 1986 and 1993.

The projected population for Riverside County in 2020 is 2,773,431 based on the US Census figures. The proportional projected population for the Coachella Valley is 470,929. Based on this information, the basis for two scenarios, based on past rates of population growth and actual per capita population growth were calculated:

1. Trend rate = 8.2 people/ha:
Projected growth = 2020 population minus 1993 population
470,929 – 224,357 = 246,572 people
Settlement density of 8.2/ha gives an additional 30,070 ha of new development.

2. Growth = 12.28 people/ha:
Projected growth = 2020 population minus 1993 population
470,929 - 224,357 = 246,572 people
Settlement density of 12.28 people/ha gives an additional 20,079 ha of new development.

Although census results for 2000 are available, 1993 population values must be used because the population is tied to the developed lands. Development is available for 1993 only, not for 2000. Predicted development was modeled based on the methods of Landis et al. (1998). A detailed description of the modeling process used in the Mojave Desert study, on which this analysis is based, is detailed in Gonzalez (2000).

Spatial data used to model projected development are listed in the following table. All data, including NALC data listed above, have a 60 m cell size in UTM projection, zone 11 (NAD83). Acquired data layers are those data that were acquired either from direct interpretation of remotely sensed imagery or were obtained from another source. Derived data are products that were created from one or more combinations of acquired data. For example, “Change in development (93-86)” was derived by subtracting development in 1986 from development in 1993.

Table 2. Variables used in statistical analysis of AFSM statistical analysis.

Acquired Data	Derived Data
1986 development	Change in development (93-86)
1993 development	Distance to 1986 development
City boundaries	Percent development
Roads	Slope
Digital Elevation Model data (DEM)	Distance to roads
Private Land	

The baseline data were masked to include only private lands in the Coachella Valley using the same method as for the NALC data, discussed above. Pixel size for this data layer was 60 m^2. The resulting development layer for the year 2020 was resampled to 60m^2 using a nearest neighbor algorithm to allow comparison of results with NALC derived urban data. Projected development was then modeled based on the existing probabilities and patterns of development that occurred between 1986 and 1993.

Alternative futures methodology - creation of variables

Differences between the Mojave Desert methodology exist as scale (60m here rather than 100m grid cells) reduced masking because of integration of products using ENVI software rather than ESRI software, and roads were masked from all files. The rationale for masking out roads was that development does not physically occur on roads, whether the road is existing or new. Development can occur only adjacent to a road. The intent was to eliminate populating roads or counting roads as having population living on them. Also, this method does not produce "missing" values, as occurred with the original methodology. Finally, all roads were treated as equal, rather than categorizing roads to primary and secondary. In the California desert region people are just as likely to build on a dirt or otherwise unimproved road than on a paved road. This is not the case everywhere in the United States.

The modeling process involved acquiring existing data, deriving secondary data products, eliminating areas excluded from development, developing a statistical model to quantify the relationship between these variables and patterns of development, and finally, projecting potential future development based on past development history. Following the Mojave protocol, the following variables were used to develop the predictive model: percent of existing development (pctdev), slope (slope), distance to existing development (devdist), within or outside of city boundaries (citybnds), and distance to roads (roaddist). All spatial layers first had undevelopable lands, specifically public lands, masked out. Those lands in the public sector were not included in the analysis. The variable 'pctdev' was derived using a 20x20 moving window on 1986 and 1993 development data, respectively. Slope was calculated based on the DEM. The variable 'devdist' was calculated using euclidean distance on each of the development data layers, 1986 and 1993, respectively. Pixels were either within or outside of city boundaries and

as such 'citybnds' was binary. The 'roaddist' variable was calculated using euclidean distance, in the same fashion as 'devdist' was created but using roads as the input feature.

The difference between 1986 development and 1993 development (newdev) was the response variable, or *Y*, for the statistical analysis. All of the data layers described above and the variable newdev were exported into Splus statistical package for model development. The model was developed using logistic regression and a generalized linear model (GLM) was developed. The model was developed with 90% of the data, randomly selected, and validated with the remaining 10%. Because drop in deviance is not necessarily the best way to evaluate the goodness-of-fit in logistic regression, Analysis of Variance (ANOVA) using a Chi-squared test was also conducted.

Once the final model was derived, it was applied to the 1993 development layer using the corresponding coefficients and data layers to create the probability of development surface. This surface was populated with 8.2 people/ha and 12.28 people/ha, respectively, as calculated above. A confusion matrix was generated to evaluate the accuracy of the resulting probability surface with probability greater than 30.

The Alternative Future Scenario (AFS) data set contained 432,427 observations. The variables included an indicator of new development between 1985 and 1993 (ND, ND=0 no new development, ND=1 new development), in/out city limits indicator (C, C=1 in city limits, C=0 outside of city limits), percent of existing development (%ED), slope (S), distance to existing development (DD), and distance to roads (DR).

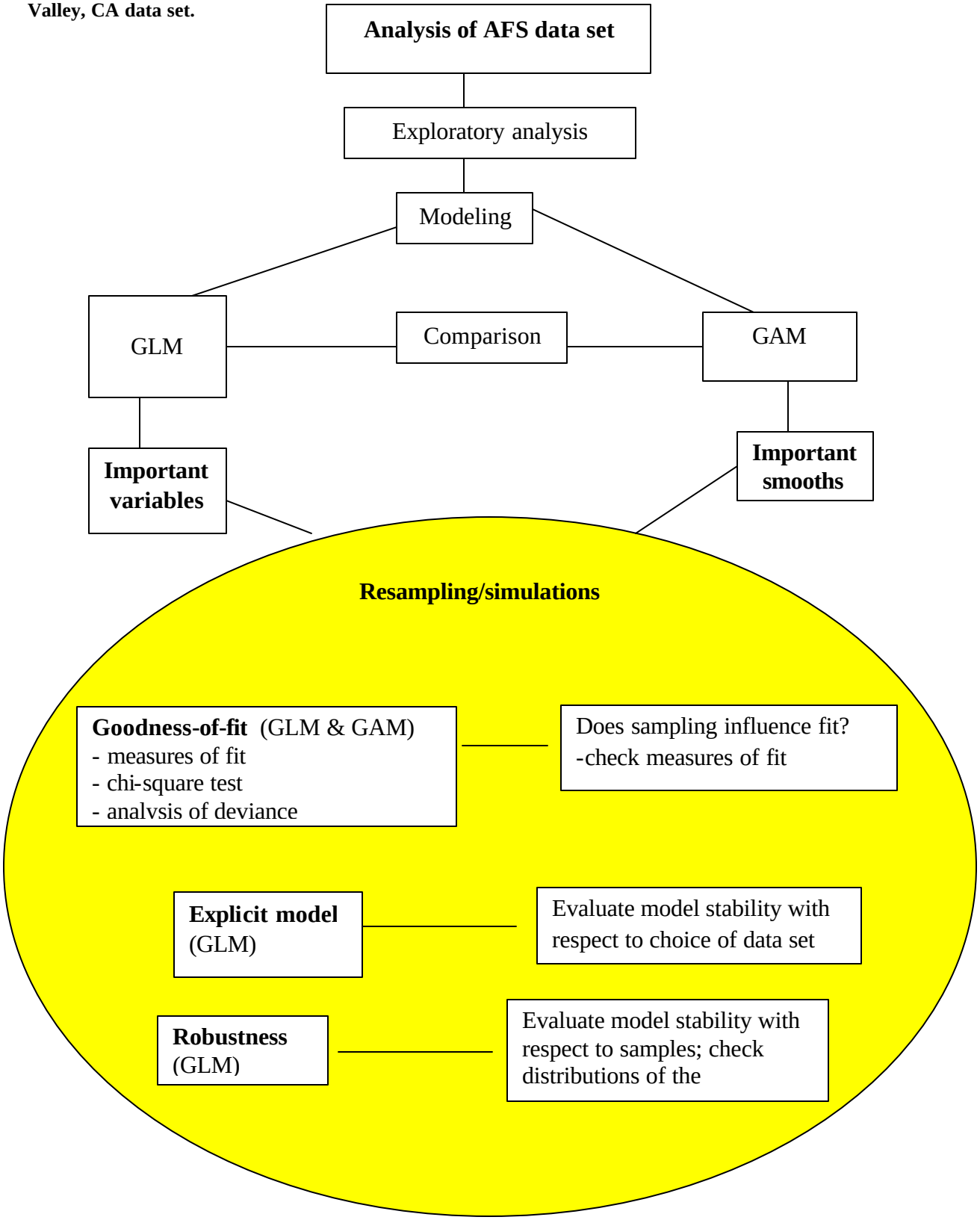
The Alternative Future Scenario Null (AFSN) data set contained 541,490 observations. The variables included an indicator of new development between 1985 and 1993 (ND, ND=0 no new development, ND=1 new development), indicator of development in 1985 (D85, D85=1 developed in 1985, D85=0 undeveloped in 1985), indicator of development in 1993 (D93, D93=1 developed in 1993, D93=1 undeveloped in 1993), in/out city limits indicator (C, C=1 in city limits, C=0 outside of city limits), percent of existing development using 10x10m, 20x20m, and 30x30m windows (%ED1, %ED2, %ED3), slope (S), and distance to roads (DR).

Analysis

The analysis and modeling of change, defined as new urban development was the goal of this analysis and was conducted on the AFS data set. Additionally, we explored the AFSN, or entire data set, regarding similarity of the properties of areas with existing development (in 1985) to areas with new development (between 1985 and 1993). The analysis on AFSN was done to determine if the new development, or change, follows similar patterns. In other words, it is important to evaluate if new development (change, a dynamic measure) depends on similar variables as development measured at one point in time. The reason for this comparison was a practical one: what if only one time shot of the area under consideration were available and a prediction future development was needed? Could a single date snapshot of development serve as proxy to change detection?

Analysis of the AFS data set is charted as follows in Figure 3.

Figure 3. Schematic showing the analysis of the Alternative Future Scenario data analysis for the Coachella Valley, CA data set.



Statistical Methods for Evaluating Data Sets

The statistical methods described in this section were used throughout all of the analysis on both data sets. They are briefly described and referenced in this section.

Exploratory data analysis

Exploratory data analysis entailed computing descriptive/summary statistics for quantitative variables: minimum, 1st quartile, mean, median, 3rd quartile, maximum, standard deviation; for qualitative variables: total numbers/proportions of 0s and 1s, as well as graphical descriptions of the distributions of quantitative data (box-plots), and testing for significant differences between parameters of developed and undeveloped areas. The tests were standard t-tests (Lehman, 1959; Lehman, 1983) for difference of means between two independent populations (developed and undeveloped areas). To test for differences in proportions we used a standard chi-square test equivalent to the familiar test for difference in proportions based on the normal distribution (Snedecor and Cochran, 1980). To check if single percentages are significantly different from given/postulated percentages, we performed one sample z-test for proportions (Snedecor and Cochran, 1980). All tests were performed on the 5% significance level.

In addition to quantitative description of the data we included graphical summaries: box-plots. A box-plot describes the distribution of a quantitative variable. The line in the middle is on the median level, the box extends between 1st and 3rd quartiles (Q_1 and Q_3, respectively), the ends of the whiskers mark $1.5(Q_3-Q_1)$, and any observation beyond the range of the whiskers is marked as an individual line. The box-plots show the symmetry (or lack of) and spread of the data around the median. They are common graphical aides to the understanding of the distribution of the data.

Another mathematical notion used in the exploratory (and other) analysis was that of conditional probability. The conditional probability of an event A given that event B happened (that is A conditioned on B) is defined as P(A|B)=P(A and B)/P(B). For example, the conditional probability of new development occurring (ND=1) given that we look inside city limits (C=1) is P(ND =1 and C=1)/P(C=1). We approximate or estimate conditional probabilities as relative frequencies. For example, P(ND =1 and C=1)/P(C=1)˜(# cells with ND=1 and C=1)/(# of cells with C=1).

Modeling

We used the following primary statistical modeling tools: generalized linear models (GLM - logistic regression) and generalized additive models (GAM - logistic regression). The reason for using these models was of practical nature. Both are available in most professional statistical packages and are optimized (parameterized) automatically according to well-defined statistical principles without any interaction with the user.

Generalized linear model (GLM) used for this analysis was logistic regression (McCullagh and Nelder, 1989). A logistic regression model provides an estimate of a function of the response variable, here either development or no development, as a linear function of the predictor variables. In logistic regression, the response variable is binomial, meaning that there are only

two possible values: development (1) no development (0) with probability of success (development) *p*. The logistic regression equation connects a function *L* of p (the link function) with the linear function of the predictor variables. The link function is the logit function: L(p)=ln(p/(1-p)). In mathematical terms:

$$L(p) = \ln\frac{p}{1-p} = b_o + \sum_{i=1}^{n} b_i x_i, \quad i=1,\ldots,k$$

where ßi are real coefficients (parameters of the model), and x_i are the *k* explanatory variables. Once the model is parametrized, *p* is computed from the values of L(p) for each observation using an inverse logit function

$$p = \frac{e^{L}}{1+e^{L}}.$$

The inverse logit function is closely connected with the logistic distribution and that is why this regression technique is called *logistic* regression. The model is parameterized or fit to the data with maximum likelihood estimates of ßi. These are computed using an iterative optimization procedure called Iterative Reweighted Least Squares (IRLS). For details on the IRLS and generalized linear models please see McCullagh and Nelder (1989). All computations were done using the Splus procedure *glm*.

Generalized additive model (GAM) was also a logistic regression model, but in an additive model setting. An additive model is similar to GLM, but more general (Hastie and Tibshirani, 1990). Instead of fitting a linear function to L(p), we fit an additive functional, that is a sum of smooth nonparametric functions of the quantitative explanatory variables. In mathematical terms, we have

$$L(p) = \ln\frac{p}{1-p} = a + \sum_{i=1}^{n} s_i(x_i), \quad i=1,\ldots,k$$

where a is a real number, and s_i is a smooth nonparametric functions (*smooths*) of the *k* explanatory variables. Only quantitative explanatory variables were smoothed. We used popular cubic B-splines for functions s_i. The model was parameterized using the local scoring algorithm, which iteratively fits weighted additive models by backfitting. The backfitting algorithm is a Gauss-Seidel method for fitting additive models, by iteratively smoothing partial residuals. All procedures used to fit the models to the data are available in Splus and were done using the procedure *gam*. For complete details on the GAM and methods for their fitting to the data please see Hastie and Tibshirani (1990). One can compute the inverse logits from a fitted GAM model in exactly the same way as from the GLM model.

The outputs of both GLMs and GAMs can be thought of as probability surfaces over the area under consideration (all cells). That is, the models assign an estimated probability of development to each cell in the two-dimensional grid.

An advantage of a GLM over GAM is that GLM provides an explicit formula for the model function, while GAM exists only in the virtual space of a computer, because smoothing is done in a non-parametric way. That is there are no explicit functional forms for the smooths of the explanatory variables. The advantage of GAM over GLM is its flexibility.

Variables influencing development

To assess the impact/influence of the individual land/urban characteristics on the probability of development, we employed analysis of deviance and Cp statistic criterion. For GLM models analysis of deviance we used two tests (McCullagh and Nelder, 1983): partial t-test for importance or significance of single variables after adjustment for all the other variables and a sequential chi-square test for significance of sequential addition of individual variables to the null model. The null model has only a constant term. For GAM models analysis of deviance employed an approximate partial chi-square test of importance of smooths of each explanatory variable (Hastie and Tibshirani, 1990). The Cp statistics, (McCullagh and Nelder, 1983; Akaike, 1973) was used on both GLM and GAM to get a rough handle on the relative importance of each variable or its smooth to the response. The smaller the Cp statistics, the more "influential" the corresponding explanatory variable.

Goodness-of-fit techniques

In any statistical modeling problem, we have a choice of measures of goodness of fit. These measures always depend on the questions we want to answer by modeling and on the cost of erroneous predictions with the model. In this project we felt that it was most important that the models classify the observations correctly, i.e. that the agreement between the cells *observed* as developed/not developed and *classified by the model* as developed/not developed was reasonable. To quantify the goodness of fit we used traditional as well as new (developed for the purpose of this project), more intuitive measures of fit.

The traditional measure of fit is a chi-square test for independence between the observations and model predictions. The chi-square test (Hosmer and Lemeshow, 1989) provides means for assessing classification accuracy by testing independence between the observed and classified/model predicted observations. We classified observations as D (developed) or ND (not developed) based on the probability of development estimated by each model. This required choosing a cutoff for probability of development that would classify a cell as D or ND. We chose this to be 0.4, which was the global maximum of the conditional probability of observed given predicted development function. A global maximum of a function is the x-coordinate of the highest point on the graph of that function. We chose to look at this conditional probability because we felt that correct prediction of development for areas that were actually developed is very important, that is, an error in this prediction may be quite costly.

Another look at the fit can be provided by analysis of deviance of a model. Deviance or residual deviance[a] of a GLM or GAM model is defined as

***Residual deviance* = -2 x ln(*F*),**

where F is the likelihood function. That is the residual deviance is the familiar *"–2*loglikelihood function"*. The likelihood function (in logistic regression) gives the probability of observing the observed data under a logistic regression model. That is, F is the probability of observing the data we actually observed computed using the probability surface estimated from the model. In mathematical terms

$$F = \prod_{i=1}^{n} p_i^{ND_i} (1 - p_i)^{1 - ND_i},$$

where every $p_i = P(ND_i = 1)$ is the probability of development in cell *i* estimated by the model, and *n* is the total number of observations. The common interpretation of residual deviance is: having two models that give two sets of probabilities of development for every cell (two probability surfaces), the one with larger likelihood F is better. Thus, the model with smaller residual deviance is better. We would like to note that although the definition of residual deviance is very different from the definition of the error (residual) sum of squares in the familiar linear regression, their use for assessing fit of a regression model is similar. For linear regression we seek a model with minimal error sum of squares, for logistic regression we seek a model with minimal residual deviance. One can test if difference between two models is significant using a chi-square test based on the difference between the residual deviances of the models (McCullagh and Nelder, 1983).

Another, perhaps more informative use of the residual deviance is computation of *n-th root* of the likelihood function for a given model, that is $\sqrt[n]{F}$. Mathematically, $\sqrt[n]{F}$ is the *geometric* average probability of observing what was actually observed in a cell. Averaging is with respect to the number of cells. Geometric average is a more suitable estimate of an *average term* in a product (F is a product) than arithmetic average. Having the residual deviance computed for a model, it is easy to compute the $\sqrt[n]{F}$. Namely,

$$\sqrt[n]{F} = \exp(-\text{Residual deviance}/2n),$$

where n is the number of observations (cells) in the data set. Again, a good model would give a large *average* probability (per cell) of observing what was actually observed.

The new, intuitive measures of fit were conditional probabilities of match between the observations and model predictions. We felt that a good model should predict development with a reasonable accuracy. However, we had to decide on measures of accuracy. An intuitive approach is to look at the four combinations of model predictions coupled with the observations.

[a] We will use term residual deviance to be consistent with the familiar notion of residual sum of squares for linear models.

We have two outcomes from the model: development (1) or no development (0). The observations are also partitioned into developed (1) and not developed cells (0). When we couple the model predictions with the observations we can look at the probabilities of the model correctly predicting what was observed. More precisely, we look at the following conditional probabilities:

1. P(o1|p1), the probability of observing development (o1) given (|) that the model predicted development (p1);
2. P(o0|p0), the probability of observing no development (o0) given that the model predicted no development (p0);
3. P(p1|o1), the probability of predicting development (p1) given that development occurred (o1);
4. P(p0|o0), the probability of the model predicting no development (p0), given that no development occurred (o0).

We will refer to these probabilities as measures of fit. The choice of a particular measure of fit should be left to the user. We estimated all of them for all the models.

In order to estimate the measures of fit, we classified the probabilities of development returned by the models into two categories: development or no development. That required choosing a cutoff point for the probabilities of development. All probabilities smaller than the cutoff were classified as prediction of no development. All of those equal to or larger than the cutoff were classified as predicted development. For every model, we explored a range of cutoffs. As expected, a given cutoff does not maximize all the measures of fit at the same time.

The conditional probabilities of fit were estimated as relative frequencies. For example, the conditional probability of observed development given predicted development was computed as

P(o1|p1)= P(o1 and p1)/P(p1) =

˜ (the number of cells with observed and predicted development)/(the number of cells with predicted development).

Sampling

Both data sets (AFS and AFSN) were heavily weighted towards undeveloped areas. That is, the majority of the observations were areas (cells) with no new development (AFS) or with no development in 1985 and 1993 (AFSN). In an attempt to balance the data sets used for modeling, we combined samples of observations from the undeveloped population/cells with the full set of the developed observations to create new data sets. These were later used for modeling and the models obtained on the samples were compared with the models obtained from the full data set. All sampling was simple random sampling.

Another important reason for sampling was to explore the effect of choosing different data sets for the analysis in the first place. If two people were to choose the *spatial window* of cells (area) for analysis, they would likely choose different areas, that is, different data sets. Performing analyses on samples from our data simulated that experience.

The number of cells with new development in the AFS data set was 13,528. The number of developed cells in 1985 in AFSN data set was 23287. The number of cells with new development between 1985 and 1993 in AFSN was 13528. The samples of the following sizes from the population with no new development were used: 14,000 (about the same size as the size of the population with new development), 30,000, 60,000, 100,000, 150,000, 200,000, and 250,000. The sample sizes were chosen as approximately multiples of the size of the population with new development. The reason for this choice of the sample sizes was to get a reasonably complete picture of what is the influence of the sample size on the models: the sample sizes covered a spectrum from about the same as the size of the population with new development to almost the size of the population with no new development.

Model Robustness

Robustness of the models fit on the samples was quantified using resampling methods (Efron and Tibshirani (1993)). For any given sample size, we generated 100 (or 50 for time efficiency when dealing with larger sample sizes) independent samples from the undeveloped population. Each of these samples was combined with all the developed observations to become 100 (or 50) data sets. GLMs were parameterized on these data sets. The distribution of the models' coefficients was examined for spread and symmetry. A tight (small variance) and fairly symmetric distribution indicates a modeling process robust to the choice of the sample.

STATISTICAL ANALYSIS - RESULTS

The AFS data set contained 432,427 observations (rows). The response variable was an indicator of new development between 1985 and 1993 (ND, ND=0 no new development, ND=1 new development). The explanatory variables we focused on were those readily accessible using a GIS: in/out city limits indicator (C, C=1 in city limits, C=0 outside of city limits), percent of existing development (%ED), slope (S), distance to development (DD) and distance to roads (DR).

Exploratory data analysis

Exploratory data analysis showed statistically significant differences between the average values of the explanatory variables for the areas with (ND=1) and without (ND=0) new development. All tests were standard two sample t-tests for difference of means described in the STATISITCAL METHODS section. All p-values were 0 (that means below 10^{-6}). Table 3 contains values of the means of all quantitative exploratory variables for areas with and without new development.

As for the difference between the percentage of new development within (C=1) and outside (C=0) city limits (C is a qualitative variable), about 6% of the cells within city limits and about 2% of cells outside of the city limits underwent new development. The difference is statistically significant (p-value=0) according to the chi-square test described in the STATISTICAL METHODS section. Looking at new development and its location within/outside city limits from another point of view, namely what percentage of cells that underwent new development are located within/outside city limits we get the following distribution. About 63% of newly

developed cells were within and about 37% of them were outside city limits. The 63% is significantly different (p-value=0) from 50% we would expect if there were no relationship between city limits and new development.

Further, note that the dispersion/standard deviation (st.dev) of the distributions of distance to existing development, distance to roads, and slope is smaller for the areas with new development than for the areas without new development, see Table 3. Finally, the means of all quantitative exploratory variables are larger than their medians which suggests that their distributions are skewed to the right.

The differences in variability of explanatory variables for areas with/without new development as well as skewness of the distributions of the quantitative explanatory variables can be seen on the box-plots in the APPENDIX I.

Table 3. Mean, median and SD of the distributions of the quantitative explanatory variables for areas with and without new development.

New development indicator	Average/median (st.dev) percent existing development (%ED)	Average/median (st.dev) distance to existing development (DD)	Average/med (st.dev) slope (S)	Average/median (st.dev) distance to roads (DR)
ND=1	10.4/3 (14.4)	942/536 (1289)	1.5/1 (2.3)	92/60 (60)
ND=0	0.9/0 (4.8)	9482/6010 (9595)	6.8/3 (9)	1571/360 (3170)

We now turn to the analysis of the dependence of new development on the percent of existing development and distance to the existing development.

The (conditional) probability of new development for the areas with zero percent existing development is only about 0.015 (6070/396283). For the areas with nonzero percent existing development, the probability of new development increases almost linearly (maximum value about 0.4) as a function of %ED when %ED remains below about 45-50%. Once %ED exceeds 50% the amount of variability/scatter in the probability of new development increases dramatically and no clear relationship can be seen. Figure 4 presents the conditional probability of new development as a function of %ED for areas with %ED>0.

The likelihood of new development also changes with distance to existing development. The largest probability of new development (about 0.1) have the areas closest to the existing development, with about 6% of new development occurring within 60 m, 23% within 200m and 70% within 1 km from the existing development. The functional relationship between probability of new development and distance to existing development is presented graphically in Figure 5.

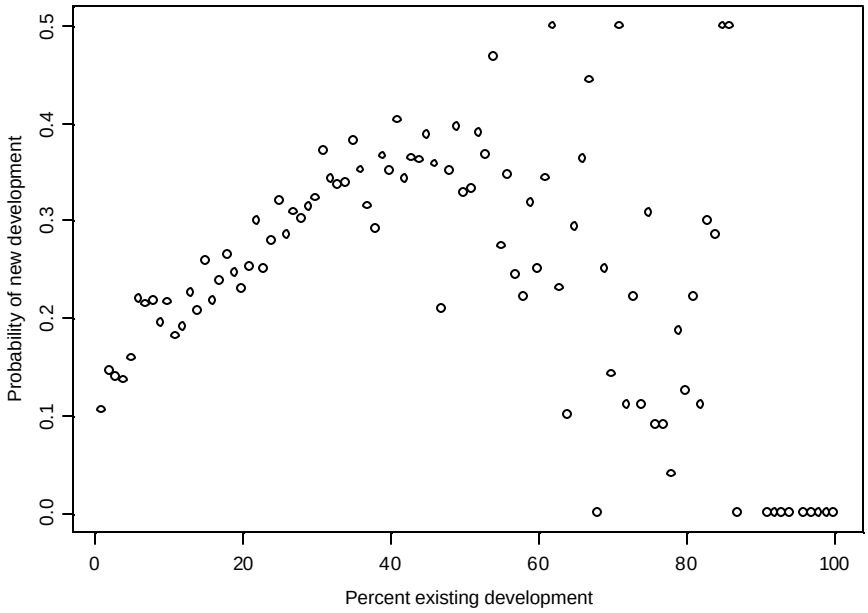


Figure 4. Probability of new development as a functon of the percent of existing development.

Variables influencing development

Both GLM and GAM were fit to the entire AFS data set. The following are the results of the analyses described in the STATISTICAL METHODS section.

GLM

Partial t-tests were conducted for all variables in the model. A partial t-test for any variable tests for the significance of adding that variable to the model containing all other variables. The null hypothesis states that the coefficient for that variable is zero and it is tested against an alternative that is is different from zero. We present the results of this analysis: estimated value of each coefficient, its estimated standard error and the corresponding t-statistics. Below the results for the partial t-tests are null and residual deviances (with their degrees of freedom) for this model (described in the STATISTICAL METHODS section).

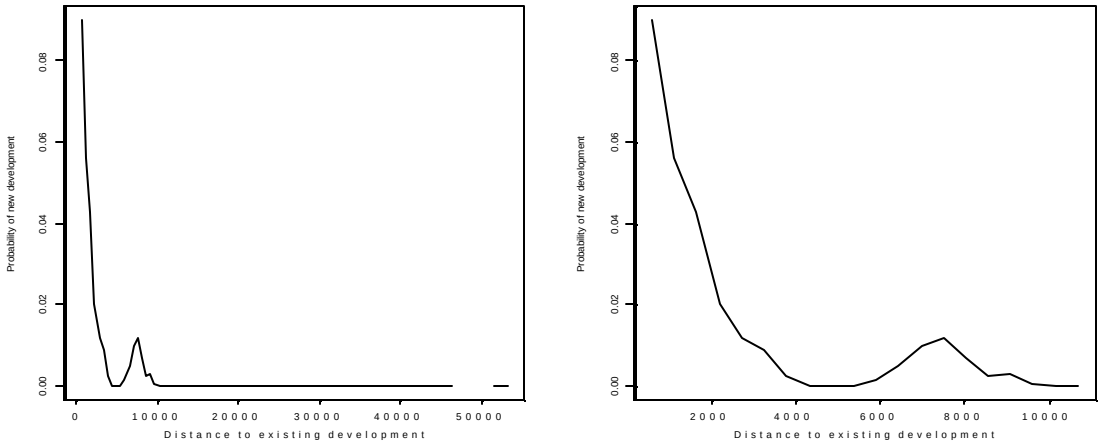


Figure 5. ***Left panel:*** **Probability of new development as a function of distance to existing development.** ***Right panel:*** **Probability of new development as a function of distance to existing development, DD less than 10km.**

Results of partial t-tests:

```
            Coefficients   Std. Error   t value
  intercept -0.5393109987 2.343254e-02 -23.01548
      S     -0.1164207453 4.808505e-03 -24.21142
     DR     -0.0082962002 1.529117e-04 -54.25483
     DD     -0.0005304171 9.649497e-06 -54.96836
      C      0.3082298894 1.049022e-02  29.38260
    %ED      0.0176472252 7.849244e-04  22.48271

    Null Deviance: 120367.9 on 432426 degrees of freedom
Residual Deviance: 76638.99 on 432421 degrees of freedom
```

All variables proved significant to modeling development because all t-statistics are large implying that p-values are essentially zero.

Next, we have results of the sequential chi-square tests for significance of all variables added sequentially (first-top to last-bottom) to the model. The first column lists all the explanatory variables in the order they were added to the model. The second column provides residual deviance of the sequentially upgraded models. The last column contains the p-values of the sequential chi-square tests. The last line of the table above corresponds to the model with all explanatory variables. Null model includes an intercept, but no explanatory variables.

Results of sequential chi-square tests:

```
              Res.Dev       p-value

  Null model  120367.9
      S       110716.8        0
```

```
     DR          96429.5       0
     DD          78228.7       0
      C          77130.7       0
    %ED          76639.0       0
```

The sequential chi-square tests are all significant (all p-values are zero), implying that all variables added to the model (in that sequence) are providing new significant information.

Finally, we computed Cp statistic for all models that included only one explanatory variable. We provide the Cp statistic values for the null all one-variable models below. Null model includes an intercept, but no explanatory variables.

Results of the Cp statistics analysis:

```
                 Cp
  Null model  432429.0
         S    427778.4
        DR    429504.4
        DD    421979.2
         C    413098.8
       %ED    395458.6
```

The Cp statistics for all explanatory variables are very close to one another. Percent existing development has the smallest and distance to roads has the larges Cp statistic suggesting that perhaps percent development has more influence on the probability of new development than distance to roads. We would caution against more affirmative statements made on the basis of Cp statistics, which is designed as an indicator, not an absolute measure of importance.

GAM

To assess significance of the smooths of all explanatory variables, we ran approximate partial chi-square tests on the additive model built on the full data set. This partial chi-square test corresponds to the partial t-test done for linear models. Below, we report the p-values for the test of significance of smooths of all quantitative variables and the null and residual deviances of the additive fit.

Results of (approximate) partial chi-square tests:

```
           p-value
intercept
   s(slope   0
s(roaddist)  0
 s(distdev)  0
 s(percdev)  0

    Null Deviance: 120367.9 on 432426 degrees of freedom
Residual Deviance: 75900.31 on 432409.9 degrees of freedom
```

All smooths of the explanatory variables are significant for the model. The significance of the indicator of city limits was established earlier, along the GLM analysis.

GAM and GLM modeling

The objective of GLM modeling effort was to construct/parameterize models on the sample data for different sample sizes and on the full data set. This analysis was done to see if the models can be fitted on the samples which in general takes less machine time.

We fitted models to samples of increasing size. Since the number of observations with new development was minimal compared the size of the data set, we used all of them in all samples. We sampled the observations with no development and used the combined data sets for modeling. The number of observations with new development was about fourteen thousand (13,528), so we paired that data set with 14k, 30k, 60k, 100k, 150k, 200k, and 250k observations with no development to form seven samples. These were then used for estimating the models. After the models were estimated we applied them to the entire data set and computed measures of fit.

The objective of GAM modeling was to see if the improvement of a GAM over GLM was statistically significant. Since we cannot write GAMs explicitly, we report only the result of the test of significance for difference between GAM and GLM fitted on the full data set.

GLM modeling

The following table contains mean values of the coefficients for all explanatory variables in the GLMs fit on samples and on the full data set. The coefficients are for the linear models estimated on samples and full data set. We computed mean coefficients using resampling. For the samples of size 14k to 100k, we created 100 data sets by combining all observations with new development with 100 samples from the undeveloped population. A GLM was fit on every data set with given size. The coefficients of each of the 100 models were averaged and are reported in Table 2 as mean coefficients. For samples of size 150k to 250k, we used samples of size 50 for time efficiency[b]. The columns correspond to the sample sizes, rows to the explanatory variables. To get estimated mean probabilities of development, we need to invert the logit transform[c]. For example, the linear model fit on the full data set is (see last column of Table 4):

$$\text{Ln}(p/(1-p))=-0.5393-0.0005*\text{DD}-0.0083*\text{DR}+0.0176*\%\text{ED}-0.1164*\text{S}+0.3082*\text{C}.$$

Substituting values for explanatory variables into the above equation gives a value of the logit Ln(p/(1-p)) for each cell/observation. Having a value of the logit, say L, we can find the estimated mean probability of development by taking the inverse logit $e^L/(1+e^L)$.

[b] Fitting 50 simulated (sampled) data sets with no new development samples of size 150k took over 3 hrs on SunBlade 100.

[c] The results of the inverse logit transform are usually part of the standard output of a professional statistical package. If not, they can be easily computed from the fitted linear model for the logits.

Table 4. Mean model coefficients for GLMs fit on samples and full data.

sample size	14000	30000	60000	100000	150000	200000	250000	Full data
Intercept	2.5448	1.8632	1.2495	0.7886	0.4216	0.1545	-0.0534	-0.5393
Distance to existing development- DD	-0.0004	-0.0004	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005
Distance to roads – DR	-0.0080	-0.0080	-0.0081	-0.0082	-0.0082	-0.0083	-0.0083	-0.0083
Percent existing development - %ED	0.0346	0.0300	0.0261	0.0233	0.0214	0.0202	0.0193	0.0176
Slope – S	-0.1172	-0.1162	-0.1163	-0.1166	-0.1165	-0.1165	-0.1168	-0.1164
City – C	0.3766	0.3564	0.3414	0.3303	0.3226	0.3172	0.3149	0.3082

All coefficients, except the intercept (corresponding to the null or constant model) are relatively stable with respect to the choice of the sample size. Stability in this context means that they do not change much with change in the sample size. Only the change in the intercept reflects the influence of the sample size on the models. The coefficients for all the explanatory variables are remarkably stable.

In order to further assess the change in the coefficients for models estimated on different samples, we looked at the mean (averaged over all simulations, like the coefficients) standard errors of all the coefficients. These are listed in Table 5.

Table 5. Mean standard errors of model coefficients for GLM models fit on sampled an full data sets.

sample size	14000	30000	60000	100000	150000	200000	250000	full
Intercept	0.04192	0.03401	0.02996	0.02744	0.02589	0.02514	0.02447	0.02343
Distance to existing development- DD	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001
Distance to roads - DR	0.00022	0.00019	0.00018	0.00017	0.00016	0.00016	0.00016	0.00015
Percent existing development - %ED	0.00253	0.00176	0.00133	0.00114	0.00100	0.00093	0.00088	0.00078
Slope -S	0.00685	0.00614	0.00561	0.00528	0.00515	0.00502	0.00496	0.00481
City - C	0.02113	0.01649	0.01380	0.01243	0.01165	0.01124	0.01098	0.01049

The standard errors of all the estimated coefficients are very stable across the sample sizes. This indicates that the GLMs fit on different samples (or full data set) are remarkably similar not only in terms of their coefficients, but also in terms of the dispersion of the coefficients.

GAM modeling

The nature of GAMs is that we cannot write them in a closed form (see Statistical Methods for Evaluating Data Sets). Since a GAM is more general than a GLM, the observed smaller residual deviance for the GAM was expected. However, we can formally compare the performance of a GAM to performance of a GLM using a chi-square test described in the Goodness-of-fit techniques section. We performed the test for GAM and GLM fitted on the full data set and its p-

value was zero (chi-square statistic ˜738.6/11.1=66.5 on 11 degrees of freedom). Although the test showed that the models were statistically significantly different, we do not believe that the difference between GLM and GAM is of great practical importance.

Goodness-of-fit analysis

We computed all measures of fit (conditional probabilities of matches between observed and predicted development) as well as performed chi-square test of independence and analysis of deviance described in the STATISTICAL METHODS section.

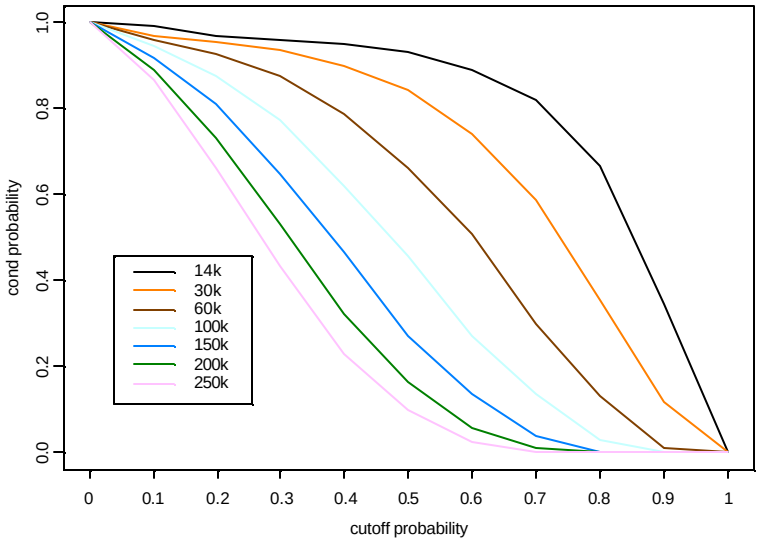
Measures of fit

We start with results on the conditional probabilities of fit for both GLMs and GAMs. Both types of models were fit on different samples and then the estimated models were applied to the full data set. This resulted in estimated probabilities of development for each cell/observation. In order to compute conditional probabilities of fit, we had to chose a cut-off value for the model predicted probability of development. Any observation with the estimated probability of development below the cutoff value was classified as having no new development, otherwise it was classified as newly developed. Then, the conditional probabilities of match between observed and predicted development were computed. This process was repeated for every sample size and for a range of cut-off values between 0.1 and 0.9 (equally spaced). The results are presented in a graphical form.

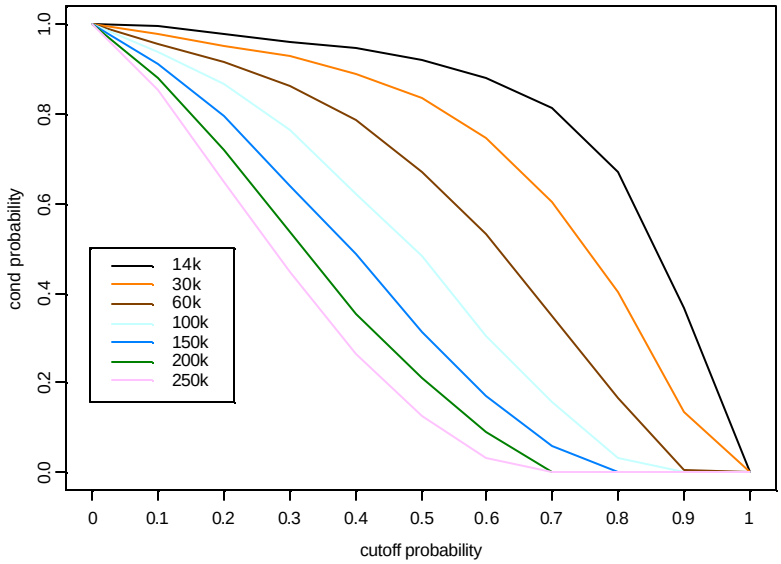
The following graphs present the estimated conditional probabilities of fit (match) as functions of the cutoff value for different samples (different curves). The patterns and the values of the probabilities are very similar for GLM and GAMs. We present results for both for the completeness of the exposition. Since the properties of the measures of fit are so similar for GAMs and GLMs, the comments we include with each set of graphs are accurate for the results of both models.

Probability of development. **P(p1|o1).** The conditional probability of predicted development given observed development decreases with cutoff point and with sample size. If this probability is chosen as the main measure of fit, then the "best" model will be one estimated on a small sample. The cutoff point for the classification of a cell as "developed" should still depend on the other measures of fit. It can be chosen, so that the other measures of fit are reasonable to the user.

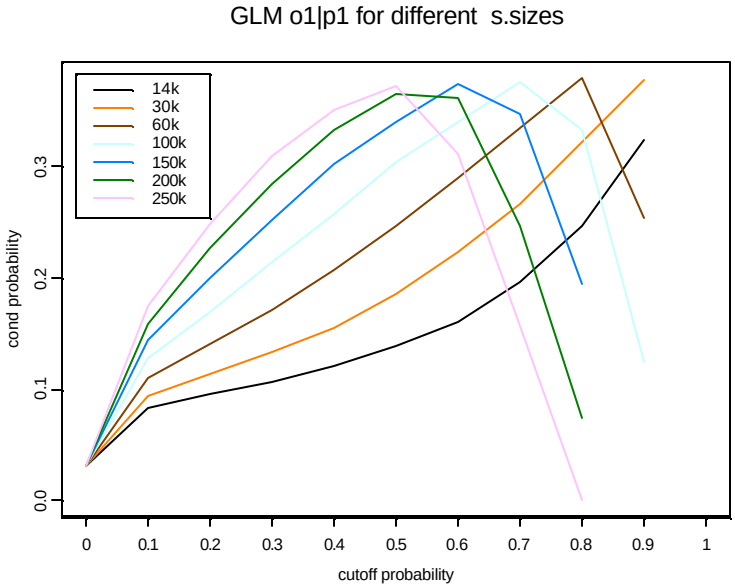
GLM p1|o1 for different s.sizes

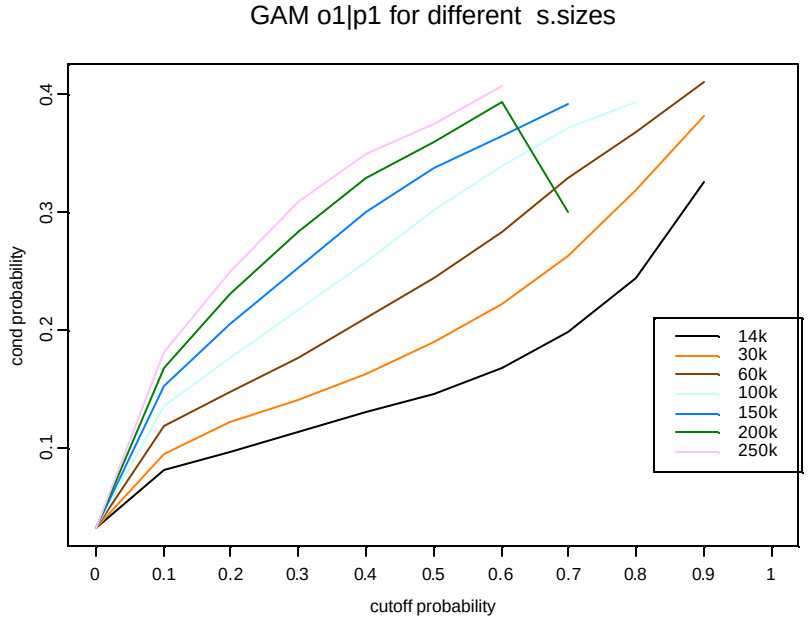


GAM p1|o1 for different s.sizes

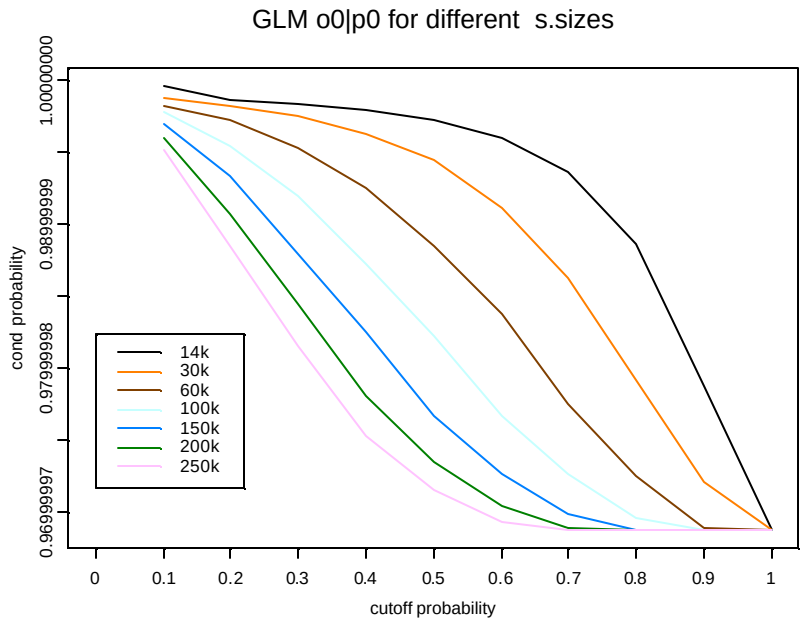


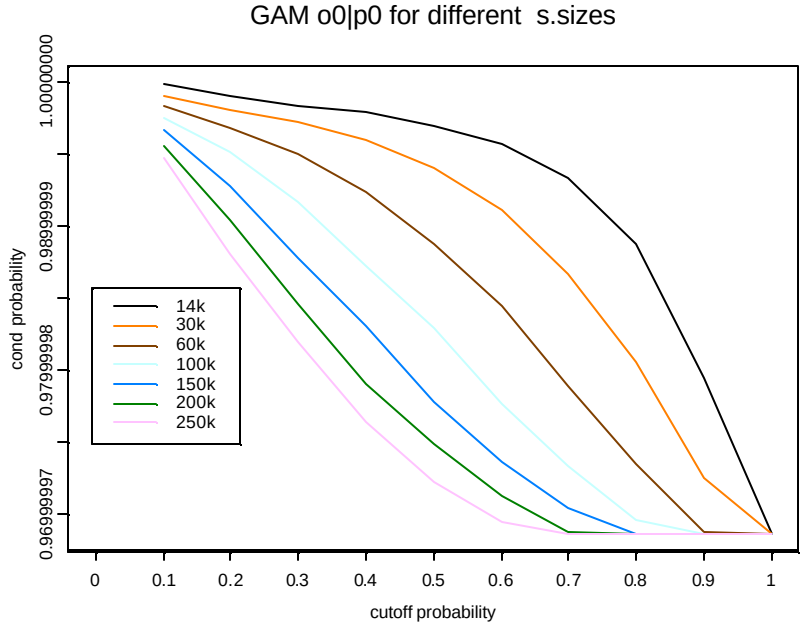
P(o1|p1). The conditional probability of observed development given predicted development is not a monotone function of the cutoff value. It first increases and then sharply decreases with the cutoff. The maximum value of about 0.4 is obtained for most of the sample sizes, but for different cutoffs. The probabilities are first increasing, and then not monotonic as functions of the sample size. If this measure is to be the main measure of fit, then the model estimated on the full data set is perhaps the "best".



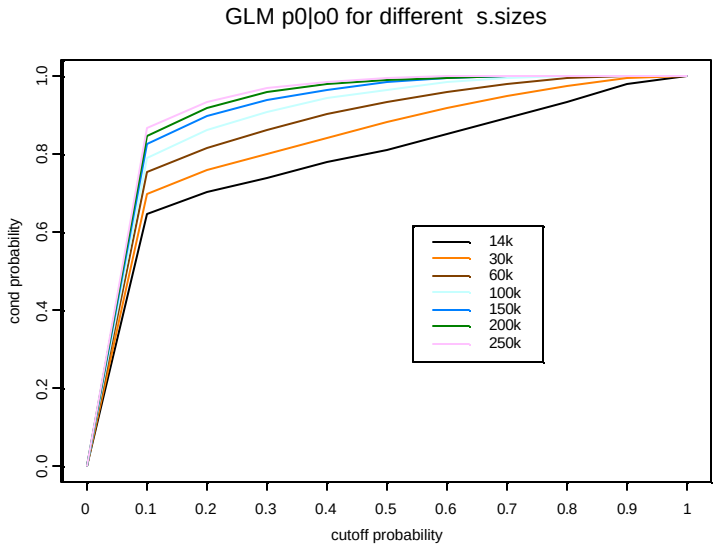


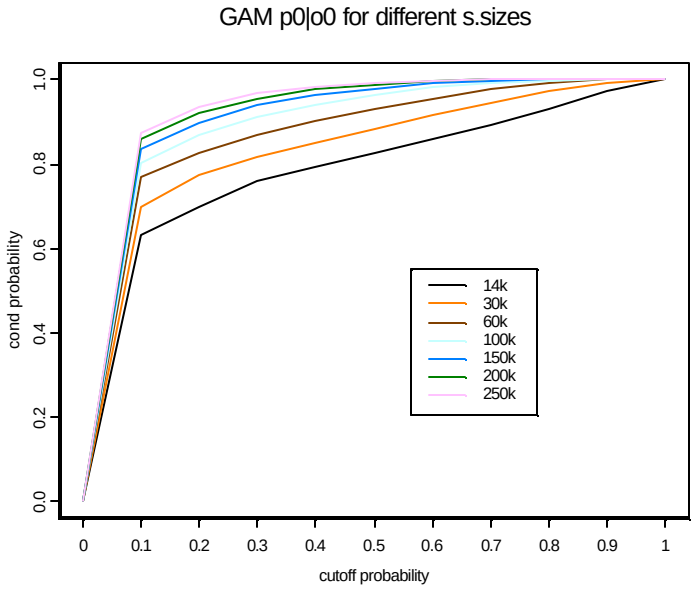
Probability of no development. **P(o0|p0).** Probability of observed no development given predicted no development is a decreasing function of both the cutoff value and sample size. In general, this probability is close to 1, so we do not need to take it into consideration when choosing the best model. No matter which model, sample size or cutoff value we chose, this measure of fit will be excellent.



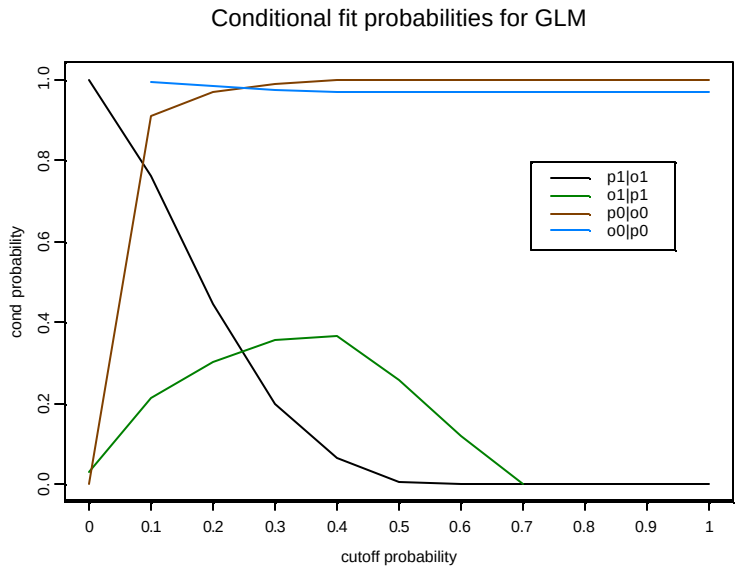


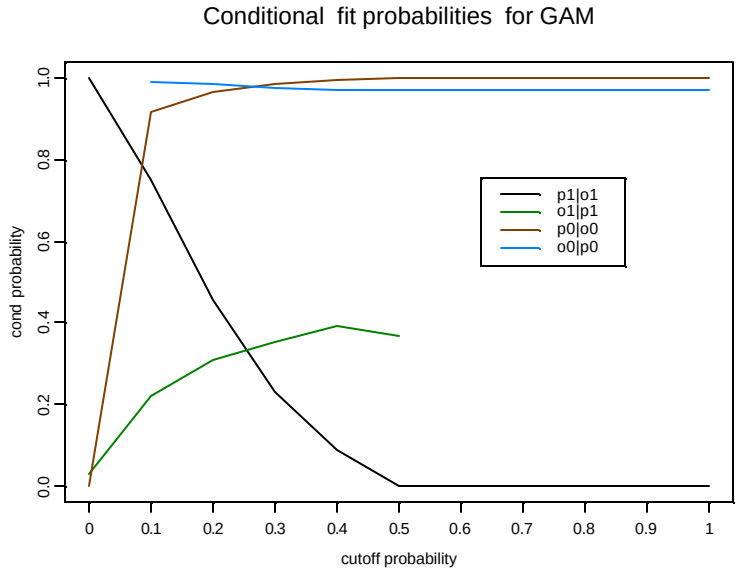
P(p0|o0). The conditional probability of predicted no development given observed no development is an increasing function of both the cutoff value and the sample size. For larger samples, the models perform well, that is this measure of fit is above 0.8 for all cutoffs above 0.4.





For any given sample size, the models fit to the entire data set behave similarly. Three measures of fit P(o0|p0), P(p0|o0) and P(o1|p1) can be adjusted by choosing an appropriate cutoff value. The cutoff value needs to be chosen based on the P(01|p1) as its maximum. The only measure of fit that will not be optimized using this approach is the P(p1|o1), but any model that optimizes with respect to this probability will not be optimal with respect to the other three. To illustrate that point we provide a graph of measures of fit as functions of the cutoff for a model build on the entire data set below. All the models based on different samples described above exhibit the same type behavior, so the plot below is representative of all models we discussed.





Overall we were positively impressed with the fit of the models. For datasets of this size and with such a large amount of variability, the models performed reasonably well.

Chi-square goodness-of-fit analysis

We also tested fit using more conventional method of a chi-square test for independence between the observed and model predicted development. We used GLM and GAM built on the full data sets for this analysis. In both cases, the tests showed dependence between observed and predicted development (all p-values=0). The tests required that we decide which cells the model predicts as new development, that is what is the cutoff probability for classifying a cell as new development. We used the maximum of the conditional probability of observed given predicted development function as the cutoff value. It was 0.4 in both cases.

Analysis of deviance

Analysis of deviance was done for GLMs in much the same way as the analysis of model coefficients in the GLM modeling section. We computed mean null and residual deviance using resampling. For the samples of size 14k to 100k, we created 100 data sets by combining all observations with new development with 100 samples from the undeveloped population. A GLM was fit on every data set with given size. The null and residual deviances for each of the 100 models were averaged and are reported in the last two rows of Table 4. For samples of size 150k to 250k, we used samples of size 50 for time efficiency[d]. Then, geometric means of the mean null and residual deviances were computed as described in STATISTICAL METHODS section and are reported in the first two rows of Table 6. These are estimates of the *geometric* average probabilities of observing the data that was actually observed for models parameterized on the sampled data sets.

Table 6. Results of the analysis of deviance.

sample size	14000	30000	60000	100000	150000	200000	250000	Full data
Geometric mean null deviance	0.500	0.538	0.620	0.694	0.751	0.790	0.817	0.870
Geometric mean residual deviance	0.736	0.744	0.782	0.819	0.849	0.870	0.885	0.915
Mean null deviance	38153.8	53951.2	70201.8	82931.8	93334.3	100827.9	106689.4	120367.9
Mean residual deviance	16891	25759.4	36114	45264.6	53356.7	59534.9	64533.6	76639

Since the sample sizes were increasing, both mean null and residual deviances were increasing reflecting more variability in the modeled data sets. The *geometric* average probabilities of observing the sampled data for models parametrized on different samples increase with the sample size.

Robustness of the models

The GLMs trained on different samples were tested for stability with respect to sampling. For each model and each sample size, we performed 100 (or 50 as described in the previous sections)

[d] Fitting 50 simulated (sampled) data sets with no new development samples of size 150k took over 3 hrs on SunBlade 100.

simulations. The result of one simulation was 100 (or 50) observations of each model coefficient. Their means were discussed in section GLM Modeling. Here we discuss their distributions.

For all sample sizes, the distributions of all coefficients were remarkably stable. That is they were fairly symmetric and had small variance. For illustration of these results, we included summary statistics of the distributions of all estimated coefficients for 100 samples of size 14,000 in Table 7 below. Histograms of the distribution for all coefficients for sample size 14,000 are presented in the Appendix. The results for the sample size of 14,000 are representative of the results for all sample sizes.

Table 7. Summary statistics of the simulated distributions of coefficients for all explanatory variables in the GLM model. 100 simulations of sample size 14,000.

	Intercept	DD	DR	%ED	S	C
Minimum	2.472	-4.32E-04	-8.20E-03	0.028	-0.129	0.336
1st Quartile	2.524	-4.20E-04	-8.02E-03	0.033	-0.120	0.368
Mean	2.545	-4.16E-04	-7.93E-03	0.035	-0.117	0.377
Median	2.545	-4.16E-04	-7.93E-03	0.035	-0.117	0.377
3rd Quartile	2.561	-4.12E-04	-7.85E-03	0.037	-0.115	0.386
Maximum	2.624	-3.98E-04	-7.70E-03	0.043	-0.107	0.417
Standard deviation	0.029	6.37E-06	1.20E-04	0.003	0.004	0.015

Since GAMs fit nonparametric smooths to the data and we do not have explicit "coefficients" for their fit, they do not yield themselves to the resampling analysis we did for GLMs.

Analysis of AFSN Data - Results

Analysis of AFSN data set was secondary to the analysis of the AFS data set. It also had a much smaller scope. We analyzed AFSN to answer the following questions:

1. Are the quantitative properties of the areas developed in 85 and new development areas different or similar?

2. Are the variables important for modeling development in 85 also important for modeling new development?

Having these goals in mind, we performed limited exploratory data analysis (1st question) and GLM and GAM modeling of development existing in 1985 and new development to see if the same variables show significant for modeling of both data sets (2nd question).

Exploratory analysis

The AFSN data set contained 541,490 observations (rows). The variables included an indicator of new development between 1985 and 1993 (ND, ND=0 no new development, ND=1 new development), indicator of development in 1985 (D85, D85=1 developed in 1985, D85=0 undeveloped in 1985), indicator of development in 1993 (D93, D93=1 developed in 1993,

D93=1 undeveloped in 1993), in/out city limits indicator (C, C=1 in city limits, C=0 outside of city limits), percent of existing development using 10x10m, 20x20m, and 30x30m windows (%ED1, %ED2, %ED3), slope (S), and distance to roads (DR).

Since we were interested only in the differences between the statistics of the explanatory variables for the areas developed in 1985 and those with new development (developed between 1985 and 1993) we computed all summary statistics for the subsets of data with: D85=1, ND =1 and (for completeness) D93=1. The areas that were developed in 1985 (D85=1) were still developed in 1993 (D93=1), thus the observations of the explanatory variables such as slope, percent development etc., were not independent for these two data sets. However, we assumed that the development in 1985 was independent of development after 1985 (new development). That assumption allowed for testing if the differences between means of the explanatory variables for the two data sets were different.

Most of the areas developed in 1985 were inside city limits (85%). The new development occurred in 63% within city limits. As a result, 77% of areas developed in 1993 were inside city limits.

Table 8. Summary statistics for all explanatory variables for areas: developed in 1985, developed between 1985 and 1993 and those developed in 1993.

Explanatory variables	Min	1stQ	Mean	Median	3rdQ	Max	Standard Deviation
Slope for areas developed in 1985 (D85=1)	0.0000	1.0000	1.3805	1.0000	2.0000	28.0000	1.3259
Slope for areas with new development (ND=1)	0.0000	0.0000	1.4750	1.0000	2.0000	61.0000	2.2572
Slope for areas developed in 1993 (D93=1)	0.0000	1.0000	1.4152	1.0000	2.0000	61.0000	1.7280
Distance to roads for areas developed in 1985	0.0000	0.0000	31.7614	0.0000	60.0000	540.0000	43.3306
Distance to roads for areas with new development	60.000	60.000	91.8361	60.000	120.0000	553.1727	59.5222
Distance to roads for areas developed in 1993	0.0000	0.0000	53.8364	60.000	60.0000	553.1727	57.6912
Percent existing development 20x20 window for areas developed in 1985	0.0150	0.4225	0.5936	0.5950	0.7950	1.0000	0.2373
Percent existing development 20x20 window for areas with new development	0.0000	0.0000	0.1070	0.0350	0.1775	0.8725	0.1459
Percent existing development 20x20 window for areas developed in 1993	0.0000	0.1075	0.4148	0.4125	0.6700	1.0000	0.3138
Percent existing development 10x10 window for areas developed in 1985	0.0200	0.5700	0.7390	0.7800	0.9500	1.0000	0.2300
Percent existing development 10x10 window for areas with new development	0.0000	0.0000	0.0734	0.0000	0.0925	0.8100	0.1360
Percent existing development 10x10 window for areas developed in 1993	0.0000	0.0300	0.4944	0.5400	0.8600	1.0000	0.3785

Explanatory variables	Min	1stQ	Mean	Median	3rdQ	Max	Standard Deviation
Percent existing development 30x30 window for areas developed in 1985	0.0067	0.3233	0.4972	0.4778	0.6778	0.9678	0.2297
Percent existing development 30x30 window for areas with new development	0.0000	0.0000	0.1243	0.0733	0.2044	0.7711	0.1435
Percent existing development 30x30 window for areas developed in 1993	0.0000	0.1211	0.3601	0.3378	0.5600	0.9678	0.2707

All differences between the mean values of all exploratory variables between areas developed in 1985 and new development were significant (all p-values were 0, i.e. below 10^{-12}). All tests were standard two sample t-tests for difference of means.

Modeling

All modeling was done on the full AFSN data set.

GLM Modeling

New development (ND) as response

Partial t-tests were conducted for all variables in the model. A partial t-test for any variable tests for the significance of adding that variable to the model containing all other variables. The null hypothesis states that the coefficient for that variable is zero and it is tested against an alternative that is different from zero. Below, we present the results of this analysis: estimated value of each coefficient, its estimated standard error, the corresponding t-statistic as well as the null and residual deviances (with their degrees of freedom) for this model.

Results of partial t-tests:

```
             Coefficients  Std. Error    t value
  intercept -2.775888424 1.440596e-02 -192.690314
          C  0.671484668 1.035916e-02   64.820378
          S -0.161548379 4.919959e-03  -32.835311
         DR -0.002157735 6.626408e-05  -32.562669
       %ED1 -5.619291404 1.506934e-01  -37.289560
       %ED2  0.876628463 3.166785e-01    2.768197
       %ED3  4.789603746 2.351114e-01   20.371640

    Null Deviance: 126540 on 541489 degrees of freedom
Residual Deviance: 103194.2 on 541483 degrees of freedom
```

All variables except %ED1 proved significant to modeling development because the t-values corresponding to them are large implying that p-values for testing their coefficients being significantly different from zero are essentially zero.

Next, we computed Cp statistic for all models that included only one explanatory variable. We provide the Cp statistics values for the null and all one-variable models below. The null model includes an intercept, but no explanatory variables.

Results of the Cp statistics analysis:

```
                    Cp
Null model     541492.0
       S       537592.1
      DR       539442.0
    %ED1       541005.4
    %ED2       538616.6
    %ED3       535643.4
       C       528162.1
```

The Cp statistics for all explanatory variables are very close to each other. The city limits indicator has the smallest Cp statistic and %ED1 has the largest Cp statistic suggesting (together with the results of partial t-tests above) that perhaps %ED1 has little influence on the probability of new development. It is not safe to make more affirmative statements made on the basis of Cp statistics alone, which is designed as an indicator, not an absolute measure of importance.

Chi-square goodness-of-fit analysis

Goodness-of-fit was tested using a Chi-square test for independence between the observed and model-predicted new development. Results showed dependence between observed and predicted new development (p-value = 0). The test required that we decide which cells the model predicts as new development, that is what the cutoff is for probability of classifying a cell as new development. A threshold value of 0.4 was used, the same as for the AFS data analysis.

Development in 1985 (D85) as response

Partial t-tests were conducted for all variables in the model and their results are reported below in the same way as when new development was the response variable.

Results of partial t-tests:

```
             Coefficients   Std. Error     t value
intercept    -5.523808947 0.0411790558 -134.141224
        C     0.239404928 0.0215449048   11.111905
        S    -0.073650109 0.0091011532   -8.092393
       DR    -0.004747751 0.0002789762  -17.018482
    %ED1     21.508492973 0.2172245355   99.015026
    %ED2    -14.820800741 0.4365570977  -33.949284
    %ED3      4.787883184 0.3435623624   13.935994

    Null Deviance: 192099.7 on 541489 degrees of freedom
Residual Deviance: 26097.06 on 541483 degrees of freedom
```

All variables proved significant for modeling development in 1985 because the t-values corresponding to all of them are large, implying that p-values are essentially zero.

Next, Cp statistics were computed for all models that included only one explanatory variable. We provide the Cp statistic values for the null and each of the single-variable models below. The null model includes an intercept, but no explanatory variables.

Results of the Cp statistics analysis:

	Cp
Null model	541492.0
S	534357.6
C	485724.3
%ED1	112177.4
%ED2	182281.0
%ED3	233914.2
DR	537470.9

The Cp statistics differ quite a bit for each of these models. Percent of existing development variables have the smaller Cp statistics, with %ED1 having the smallest Cp statistic of all explanatory variables. City limits indicator, slope and distance to roads have larger, more than two-times larger, Cp statistics than the % development variables, respectively. This suggests that the most important drivers of development up to 1985 was existing development. That is, the best predictor of existing development to 1985 was proximity to existing development. Development begets more development.

Chi-square goodness-of-fit analysis

Goodness-of-fit was tested using Chi-square for independence between the observed and model-predicted development in 1985. Results showed dependence between observed and predicted development (p-value = 0). Again, the test required that a decision was made regarding which cells the model predicts as developed in 1985, that is what is the cutoff probability for classifying a cell as developed. Again the threshold value was 0.4.

Statistical Analysis Discussion and Summary

AFS Data

In general, new development occurs in zones closest to roads and to the existing development. New development is also more likely to happen within the city limits. The dispersion of most of the explanatory variables is smaller for the areas with new development. That means that urban and environmental properties of the developed/developing areas are defined tighter than those for areas where development does not happen. For the areas with nonzero percent existing development, the probability of new development increases almost linearly as a function of the percent of existing development when %ED remains below about 45-50%. The likelihood of new development also changes with distance to existing development. The largest probability of new development (about 0.1) have the areas closest to the existing development, with about 6% of new development occurring within 60 m, 23% within 200m and 70% within 1 km from the existing development.

All explanatory variables (or their smooths) are significant for both GLM and GAM models and based on the values of the Cp statistics they have influence new development with about the strength.

Resampling analysis showed that coefficients, except the intercept for the GLM models are remarkably stable with respect to the choice of the sample size as well as the sample itself.

Stability in this context means that they do not change much with change in the sample size or within samples of the same size. Additionally, for all sample sizes, the *distributions* of all coefficients were remarkably stable. That is they were fairly symmetric and had small variance.

The GLM show statistically significant difference from GAM in terms of residual deviance, but since the actual difference between their residual deviances is relatively small we do not believe that this difference of great practical importance.

Overall we were positively impressed with the fit of the models. All measures of fit were quite reasonable, chi-square goodness-of-fit showed dependence between observed and predicted development and the *geometric* average probabilities of observing the sampled data for models parametrized on different samples increased with the sample size. Note, that even for the small samples the average probabilities were above 0.6.

Modeling – other notes

Before we settled on GLM and GAM modeling, we explored the possibility of using (binary) classification tree models (Breiman et al., 1984). Tree-based models are useful and might be considered relatively new in the statistical field. They are complex and involve a lot of user interaction to provide well-fit models. Decisions on technical statistical aspects and mathematical functions must be made by experienced users and for this reason tree-based models were not explored as a viable option for a user-friendly model. Both GLMs and GAMs are readily available in all statistical packages, and their estimation/fitting is a well known optimization process that is carried out without any user interaction or decision making. Additionally, preliminary analysis of tree models did not show improvement of fit. Since our objective was to look for models that can be used/applied by a wide user audience, we focused our analysis on the GAMs and GLMs.

Final interpretation

Here are the answers to the questions we investigated:

Q1. Are the quantitative properties of the areas developed in 85 and new development areas different or similar?

All differences between the mean values of all exploratory variables between areas developed in 1985 and new development were statistically significant and the actual averages showed the two were practically quite different. This essentially means that development patterns change in time, and that modeling future development based on existing development is not advised. More than one time-frame is required and change analysis is warranted when future scenario models will be based in some part on historical changes. Historical changes may be recent, as in present to only a few years earlier, or may span larger time frames. However, using larger time periods to estimate past trends in development change may cause accuracy or detail to suffer as averages are taken over longer time frames. In other words, tremendous change may have occurred in the form of sudden development boom. The cause-effect may be diminished if change analysis is conducted on a time frame that encompasses many years of relatively low rates of development.

Q2. Are the variables important for modeling development in time period 1 (here, 1985) also important for modeling new development?

There is a noticeable difference between the variables driving new development and those variables that drive development in 1985 (one time period). The proximity and extent of existing development variables were more important predictors of development in 1985 (one time period) than of new development, or a dynamic variable of change. Therefore, the answer is no, different variables were found to be important for explaining development patterns.

Overall, these results indicate that having data on change in the past is necessary for modeling change in the future.

Populating the cells = distributing the population

Once the probability surface is established, the next question is the algorithm used for populating cells. That algorithm depends on the interpretation of the probabilities. Given a population to distribute and the number of persons per cell fixed and the same for every cell, there are two main choices for the distribution algorithm:

1. Populate the cells with the highest probability first, then those with lower probability etc, until the population is exhausted. If there are several cells with the same probability, "toss a coin" to get the order of populating them. The "coin" would be a random permutation of the cell numbers.

2. Start with cells of highest probability. "Toss a coin" to get order within groups with the same probability. For each cell with probability *p*, toss a coin that comes up H with probability *p*. If the coin comes up H, populate that cell, if not, do not populate it. Move to the next cell, etc. Repeat until the population is fully distributed.

Other questions to consider are: How to move the population with the passage of time? How do the probability surfaces evolve with time?

SUMMARY

Define requirements

Definitions are always the place to start once development begins. In this case, requirements include the spatial scale, the temporal scale, the scope of the project, the number of issues, their complexity and interactive effects, as well as a solid understanding of where the call for action is coming from and why.

Inventory installations and their needs

Knowledge exists with individuals about each and every installation in the United States. The level of knowledge varies with the particular individual and their role of involvement, from a foot soldier in training to the President. No one person has complete knowledge across all levels of information nor can any one person have this information because of the depth and rate of change, on at least a daily basis, of installation mission and management. For any alternative

future scenario modeling tool to be not only effective but timely in the hierarchy of national defense, there must be justification first for selecting a particular installation on which to develop alternative futures and second, the needs of the installation must be clearly identified. For example, if resources are to be allocated to installation *x* although installation *y* has also identified a need, on what basis is that decision made? Can resources be shared between the two installations? How should the decision be made to focus what is traditionally an enormous effort in terms of time and cost on a single installation?

We put forth that there is not complete information compiled to make resource allocation decisions for individual installations. To successfully implement an alternative future scenario modeling tool that can be used at any installation there has to be an understanding of what the role of the installation is within the larger national security mission, what the management and training regimes are and why they are that way, what is the computational infrastructure and how capable are the people that run it. There also needs to be an understanding of the surrounding community, how it is developing and why, what the relationship is between the installation and this community. An understanding of the issues the installation faces is critical and furthermore it is necessary to understand how the issues came to exist or why they are forming. All of this must be known for each installation if a universal tool is to be developed for DoD-wide use.

Literature review

An in-depth and thorough literature review is warranted once requirements are clearly defined. Requirements are based on the installation inventories and would be expected to have a high degree of variability based on a number of different factors, as discussed above. There is a considerable amount of segregation within the modeling literature. Even though there are a number of reviews and many specific non-review papers include their own reviews, a lot of these reviews barely overlap in the systems named if they overlap at all. That is, evaluations are made in a relative vacuum and cross-model comparisons are not typical. This makes literature-based evaluation criteria a greater task.

Involve an expert in the field

But it is necessary to be careful to have someone without a strong bias toward one kind or class of models. The desire is to have an unbiased assessment so that the result doesn't mean an automatic selection of the kind or class of model that the expert originally favors.

Maintenance provisions

There is a need to provide for maintenance and perpetuity of the model rather than simply conducting a one-time analysis, as is the current standard. The reason these models should be maintained is that first and foremost, the problems will perpetuate both in issue and scope. Some shifts in issues or challenges may be subtle or even repeats on a given theme, while others may differ in unexpected and drastic ways. For example, noise restrictions may the focus for an extended time period and suddenly concern regarding harm to marine mammals may take center stage. Having to conduct a completely new analysis because existing models were not maintained would be counterproductive. Many of the models are integrated with some kind of information system that can be maintained (GIS, etc.)

This also serves to maintain a connection with the local community for positive public relations and for improved relations, if necessary. Good working relationships take both effort and time to build and become established. Once in place, coordination and facilitation of solutions are generally much easier to reach. Along these same lines is the shared resources, such as data, which can be used for coordination where typically no information or data are shared, and data are difficult to share due to technical constraints.

Knowledge encoded as data instead of as code

Variables should be the flexible element because values or range of values are going to vary with geography, time, space, etc. For this reason the desire should be to encode knowledge as data rather than knowledge as code. Knowledge will change with time and with advancements in technology. Hard coded knowledge is not easily nor readily updated, but data are. The result will be an increase in flexibility, transparency and knowledge re-use among applications.

Design considerations?

As discussed throughout this report, models do not exist alone and there should be some accounting of that fact. Therefore, the choice of a model should be done in the context of all other parts of the process as well. A great modeling package may fail completely if the other elements are not handled well too. Furthermore, there is a need to view the whole process of solving the installation's problem(s), not just the piece that generates one or more futures. For example, "good input, good output" (GIGO) means that there needs to be good infrastructure for model inputs. A model that is technically very sound, but not transparent to users and therefore, not believable, may be much less useful than one that people will believe. This is especially true given the great degree of uncertainty in these models and the difficulty of validating them.

Can DoD gain local favor and get better projections by working with local communities to provide them with a shared modeling tool?

If installations desire to actively pursue strategies in the wake of future development that involve creativity and problem-solving exercises with the surrounding community and local level government, such as counties, partnership is key. Armed with models that allow for planning and evaluation by the military, DoD will also have a valuable bargaining tool – a model that may be used for planning purposes at a community level but that would otherwise be beyond the financial capabilities of those communities. It is for this reason that DoD must take the initiative to conduct alternative future scenario analyses in the first place; even solutions (models) that are relatively inexpensive by federal agency standards are far beyond what most communities can afford and are willing to pay for. Much of the willingness to pay component also arises from a lack of education about the utility of future projection tools that can be used for evaluation and not simply for basic zoning, taxation, or other current community planning visions. Therefore it is safe to say that small communities in particular cannot afford to buy and/or support the models themselves. For this reason, teaming with the local surrounding community would help develop cooperation and potentially facilitate creative solutions at the installation level. This extension of partnership, good-faith sharing of technology, may also encourage good perception of DoD from the community.

Determine total cost to military

Clearly there are costs involved with any modeling effort. Costs include model research and development, implementation, technology transfer, maintenance, updates, technical support, and other hidden or unexpected costs that inevitably arise. These costs need to be compared to the cost of doing nothing, evaluated in the form of risk assessment and/or real dollar value. Other costs that are typically not included but are real factors for evaluation are the costs of doing nothing that result in ecological, planning, or economic changes.

REFERENCES

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle, in Second International Symposium on Information Theory (eds. B.N Petrov and F. Csaki). Akademia Kiado, Budapest, 267-281.

Agarwal, C., G. L. Green, et al. (2000). A review and assessment of land-use change models: dynamics of space, time, and human choice. Fourth International Conference on Integrating GIS and Environmental Modeling (GIS/EM4). Banff, Canada.

Baker, W. (1989). "A review of models in landscape change." Landscape Ecology **2**(2): 111-133.

Beauchemin, M. and K. Thomson (1997). "The evaluation of segmentation results and the overlapping area matrix." International Journal of Remote Sensing **18**(18): 3895-3899.

Berry, M., R. O. Flamm, et al. (1996). "Lucas: A System for Modeling Land-Use Change." IEEE Computational Science & Engineering **3**(1): 24-35.

Breiman, L., Friedman, J.H., Olshen, R., and Stone, C.J. (1984). *Classification and Regression Trees*: Wadsworth International Group, Belmont, California.

Brooker, S., S. I. Hay, et al. (2002). "Tools from ecology: useful for evaluating infection risk models?" TRENDS in Parasitology **18**(2): 70-74.

Cablk, M.E., J.S. Heaton, R.J. Lilieholm, M. de J. Gonzalez, M. Stevenson, and D. Mouat. (1999). Military ecology: The role of the Defense Department in protecting and preserving our biotic resources and challenges for civilian researchers in this realm. In D. Brunkhorst, ed., Landscape Futures: An international symposium on advances in research for natural resource planning and management across regional landscapes. Armidale, NSW, Australia. 22-25 September. ISBN 1 8639 664 3.

Christiansen, J. (2000). A flexible object-based software framework for modeling complex systems with interacting natural and societal processes. 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs. Banff, Alberta.

Congalton, R. (1991). "A review of assessing the accuracy of classifications of remotely sensed

data." Remote Sensing of Environment **37**: 35-46.

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

Couclelis, H. (2000). Modeling frameworks, paradigms, and approaches. Geographic Information Systems and Environmental Modeling. K. Clarke, B. Parks and M. Crane. New York, Longman & Co.

Efron, B. and Tibshirani, R.J. (1993). An introduction to the bootstrap, London: Chapman and Hall.

EPA (2000). Projecting Land-Use Change: A Summary of Models for Assessing the Effects of Community Growth and Change on Land-Use Patterns. Cincinnati, OH, U.S. Environmental Protection Agency, Office of Research and Development**:** 260.

Goudie, J. W. (1997). Model validation: A search for the Magic Grove or the Magic Model. STAND DENSITY MANAGEMENT: Planning and Implementation Conference, Edmonton, November 6 - 7, 1997.

Gonzalez, M. (2000). Futures Scenarios of Land Use in the California Mojave Desert. Ph. D. Dissertation. Department of Forest Resources, Utah State University. Logan, Utah.

Gunter, J.T., D.G. Hodges, C.M. Swan, J.L. Regens. (2000). Predicting the urbanization of pine and mixed forests in Saint Tammy Parish, Louisiana. Photogrammetric Engineering and Remote Sensing. **66**(12):1469-1476.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*, John Wiley and Sons, Inc., New York.

Jenerette, G. D. and J. Wu (2001). "Analysis and simulation of land-use change in the central Arizona-Phoenix region, USA." Landscape Ecology **16**: 611-626.

King, J. and K. Kraemer (1993). Models, Facts, and the Policy Process: The Political Ecology of Estimated Truth. Environmental Modelilng with GIS. M. Goodchild, B. Parks and L. Steyaert. New York, Oxford University Press**:** 353-360.

Landis, J. J.P. Monzon, M. Reilly, and C. Cogan. (1998). Development & Pilot Application of the California Urban & Biodiversity Analysis (CURBA) Model.

Landis, J., M. Reilly, et al. (2001). Forecasting and Mitigating Future Urban Encroachment Adjacent to California Millitary Installations: A Spatial Approach, California Technology, Trade, and Commerce Agency.

Law, A. and W. Kelton (1991). Simulation Modeling and Analysis. New York, NY, McGraw-Hill.

Lee Jr., D. B. (1973). "Requiem for Large-Scale Models." AIP Journal: 163-177.

Lehman, E. L. (1959). *Testing Statistical Hypothesis*, John Wiley and Sons, Inc., New York.

Lehman, E. L. (1983). *Theory of Point Estimation*, John Wiley and Sons, Inc., New York.

Li, Z., P. Sydelko, et al. (1998). Integrated Dynamic Landscape Analysis and Modeling System (IDLAMS): Programmer's Manual, Argonne National Laboratory.

Lindley, S. J. (2001). "Virtual tools for complex problems: an overview of the AtlasNW regional interactive sustainability atlas for planning for sustainable development." Impact Assessment and Project Appraisal **19**(2): 141-151.

Mayer, D. and D. Butler (1993). "Statistical validation." Ecological Modelling **68**: 21-32.

McCarthy, M. A. and D. B. Lindenmayer (Possingham, Hugh P). "Assessing spatial PVA models of arboreal marsupials using significance tests and Bayesian statistics." Biological Conservation **98**.

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*. Chapman and Hall, London.

McEvoy, D. and J. Ravetz (2001). "Toolkits for regional sustainable development." Impact Assessment and Project Appraisal **19**(2): 90-93.

NRC (1992). Global Environmental Change: Understanding Human Dimensions, National Research Council.

Parker, D. C., S. M. Manson, et al. (2002 (in press)). "Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review." Annals of the Association of American Geographers.

Power, M. (1993). "The predictive validation of ecological and environmental models." Ecological Modelling **68**: 33-50.

Provost, F. and T. Fawcett (1997). Analysis and Visualization of Classifier Performance: Comparison Under Imprecise Class and Cost Distributions. Proceedings oft he Third International Conference on Knowledge Discovery and Data Mining, AAAI Press**:** 43-48.

Redman, C., J. Grove, et al. (2000). Toward a Unified Understanding of Human Ecosystems: Integrating Social Sciences into Long-Term Ecological Research. White Paper of the Social

Science Committee of the LTER Network. Accessible at http://www.lternet.edu/research/pubs/informal/socsciwhtppr.htm.

Rykiel Jr., E. J. (1996). "Testing ecological models: the meaning of validation." Ecological Modelling **90**: 229-244.

Sargent, R. (1984). A tutorial on verification and validation of simulation models. Proceedings of the 1984 Winter Simulation Conference. S. Sheppard, U. Pooch and D. Pegden eds., IEEE 84CH2098-2**:** 115-122.

Sklar, F. and R. Costanza (1991). The development of dynamic spatial models for landscape ecology: A review and prognosis. Quantitative Methods in Landscape Ecology. M. Turner and R. Gardner. New York, Springer-Verlag**:** 239-288.

Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods, 7th ed*. Iowa State University Press, Ames, Iowa.

Sydelko, P. J., J. E. Dolph, et al. (2000). An advanced object-based software framework for complex ecosystem modeling and simulation. 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs. Banff, Alberta, Canada, September 2-8, 2000.

Tennessee, University of. (1999). LUCAS web site. http://www.utk.edu/~lucas.

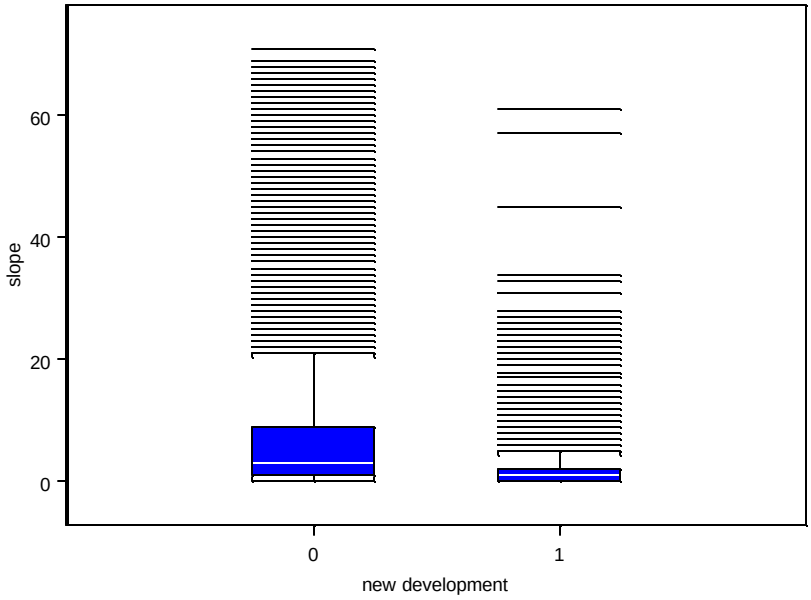
Turner, M., R. Costanza, et al. (1989). "Methods to evaluate the performance of spatial simulation models." Ecological Modelling **48**(1/2): 1-18.

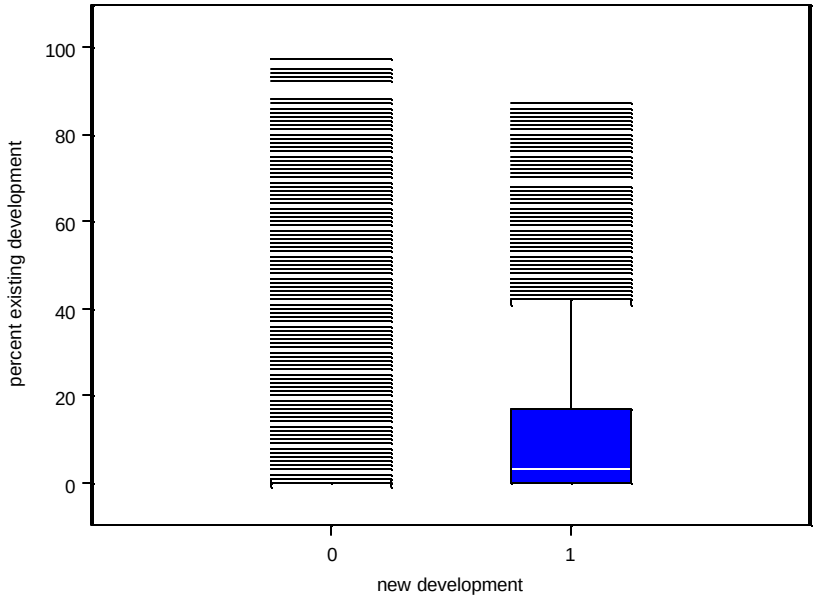
USDA-ARS (2002). The Object Modeling System. http://oms.ars.usda.gov/

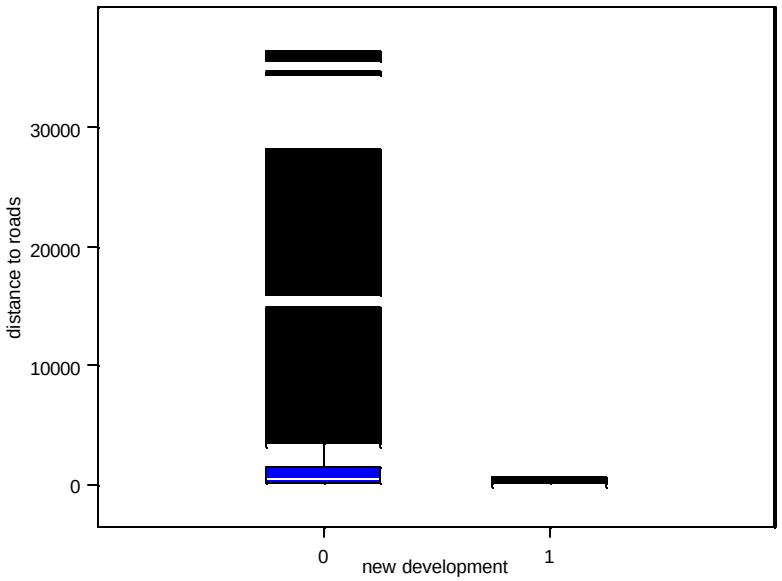
Vanclay, J. and J. Skovsgaard (1997). "Evaluating forest growth models." Ecological Modelling **98**: 1-12.

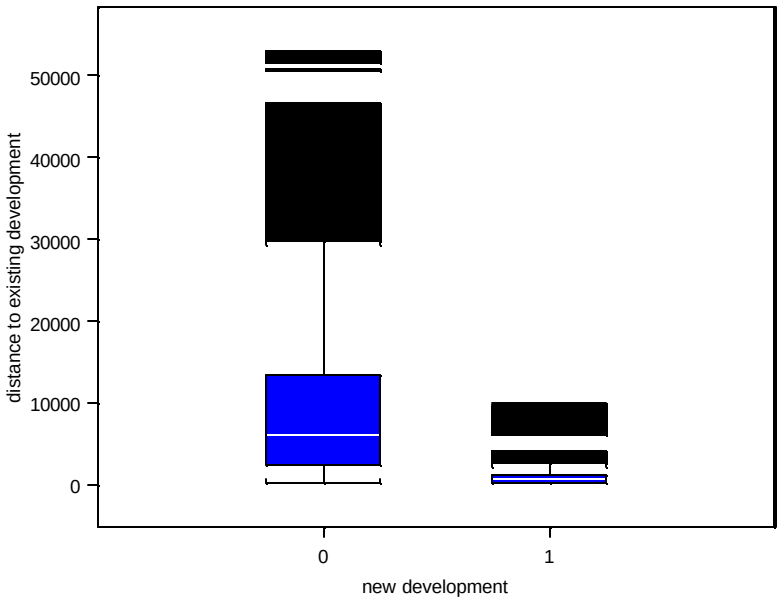
APPENDIX I.

DISTRIBUTIONS of explanatory variables in the developed and not developed areas.



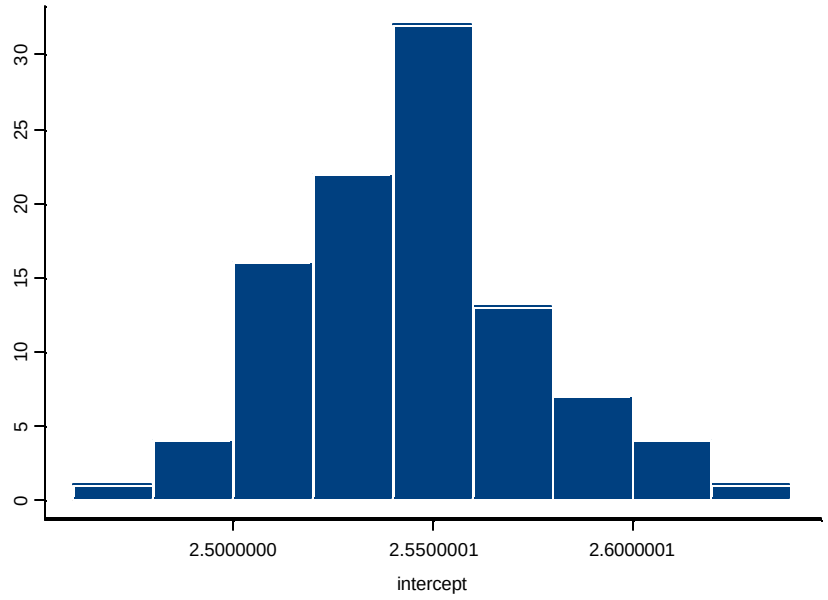


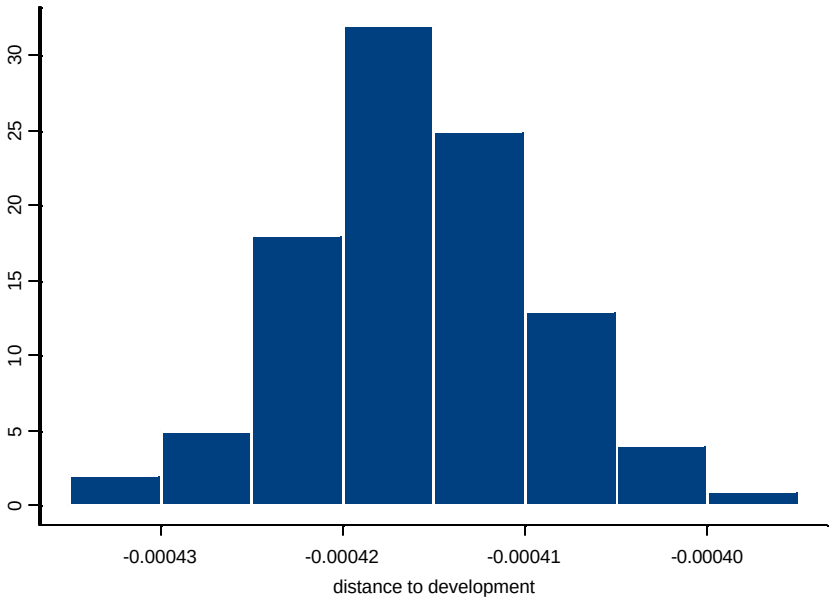

distance to roads
30000
20000
10000
0
0
1
new development

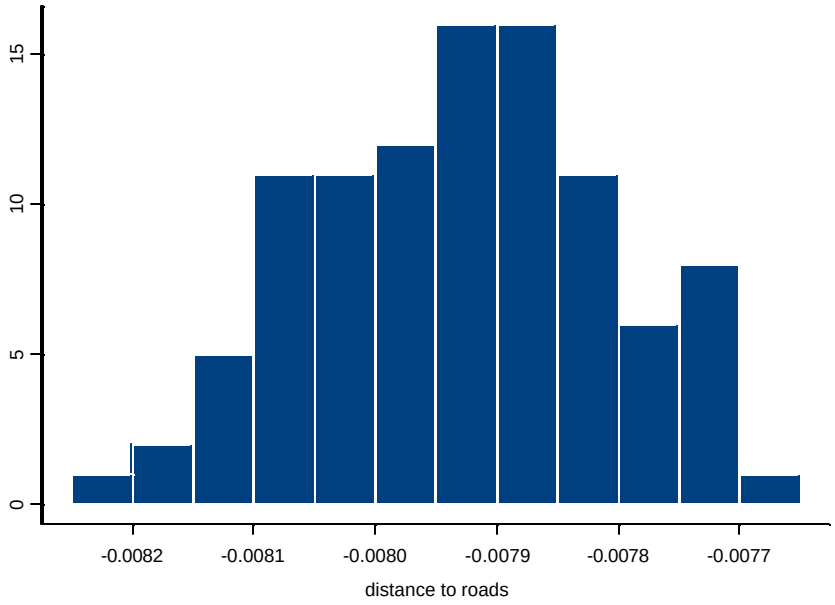

distance to existing development
50000
40000
30000
20000
10000
0
0
1
new development

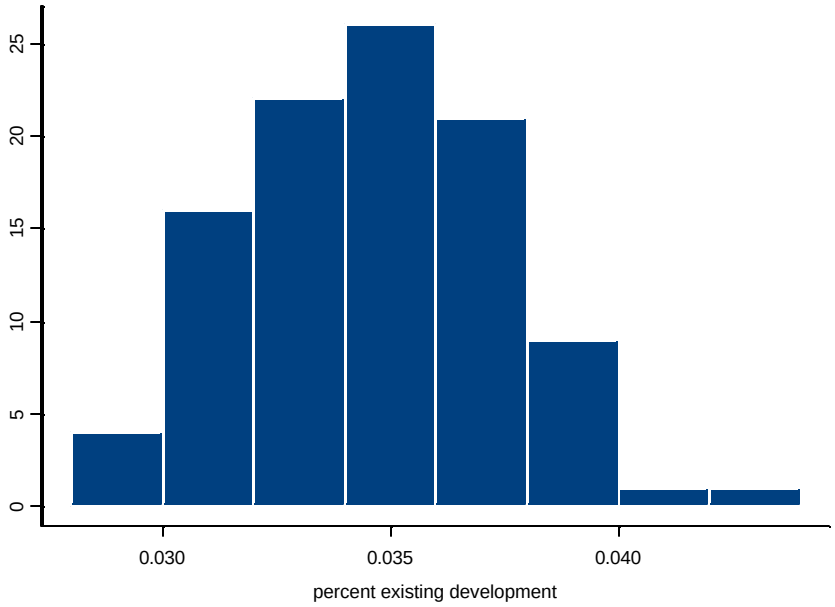
APPENDIX II.

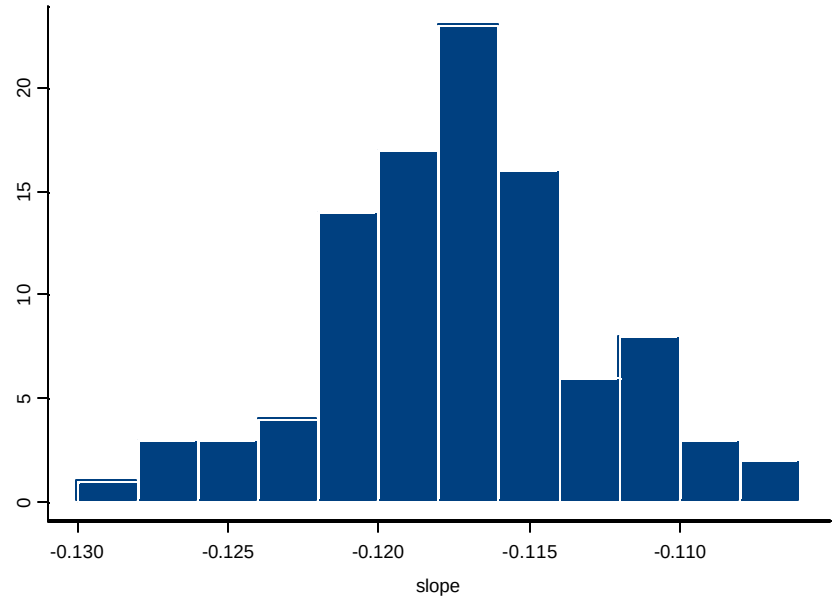
Distributions for the model coefficients for all explanatory variables for GLM model -- 100 simulations, sample size 14,000.

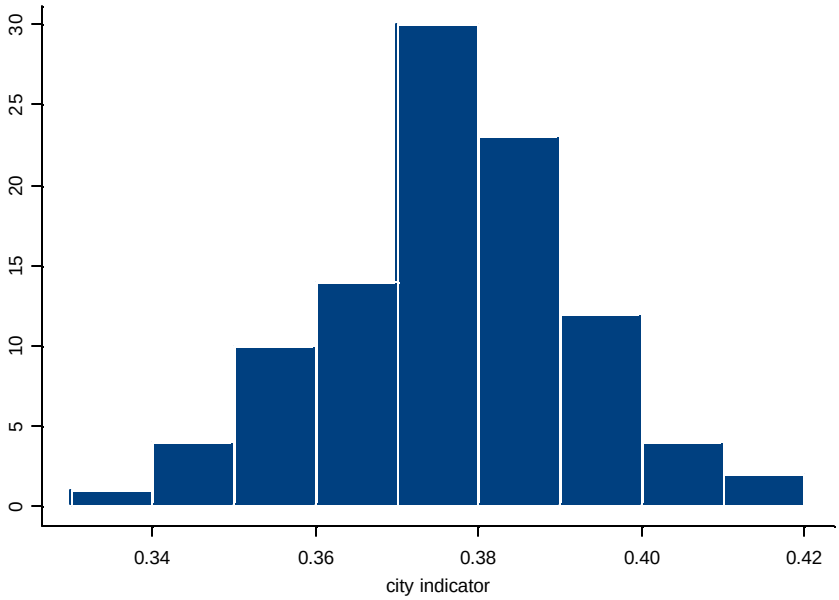





0
5
10
15
-0.0082
-0.0081
-0.0080
-0.0079
-0.0078
-0.0077
distance to roads


0
5
10
15
20
25
0.030
0.035
0.040
percent existing development


0
5
10
15
20
-0.130
-0.125
-0.120
-0.115
-0.110
slope


0
5
10
15
20
25
30
0.34
0.36
0.38
0.40
0.42
city indicator