Localizing Forest Management Using GIS and Remote Sensing: The Research Agenda of the Industrial Research Chair in Geomatics Applied to Forestry

Kirn Lowell, Geoffrey Edwards, and Chris Gold
Industrial Chair in Geomatics Applied to Forestry
Centre de recherche en géomatique
Pavillon Casault, Université Laval Sainte-Foy,
Québec G1K7P4 Telephone: (418) 656-2530
Fax:(418)656-7411

Abstract
Historically, quantitative Forest Management tools have become increasingly scientific and localized. With the sophistication of existing digital spatial technologies (DSTs) typified by geographic information systems and digital remotely-sensed data, it may be possible to add a spatial component to present-day Forest Management tools. Such a step would further the goal of conducting forest management using locally-reliable data and analysis. The current barriers to this involve limitations of existing data, statistical techniques, and utilization of information about spatial errors.

Résumé
Historiquement, les outils numériques pour l'aménagement des forêts ils ont devenu plus scientifiques et plus localisés. Avec la rafinment de technologies numériques et spatiales qui existent — c'est à dire, les systèmes d'information géographique et les télédétection — peut-être c'est possible ajouter un element spatial a des outils actuels de aménagement forestier. Une démarche comme ça servirait le but de conduire l’aménagement forestier en utilisant les données et les analyses qui sont sûr pour une localisation donnée. Les barrières actuelles impliquent des limitations des données existantes, des techniques statistiques, et l’utilisation de connaissances des erreurs spatials.
1. INTRODUCTION AND PERSPECTIVE

A. FORESTRY RESEARCH PRIOR TO THE 1980S

Until the late-1800s in North America, forest "management" was generally synonymous with "exploitation." Human populations had not yet swelled to a point at which forest resources were considered to be a limited resource which must be utilized wisely to ensure the existence of adequate wood supplies in the future. However, by the early 1900s it had become clear that North America's wood supply was, indeed, limited. Thus if future supplies of wood products were to be available, foresters needed to adopt a philosophy of "perpetual management" rather than "limitless exploitation."

This philosophical change was (arguably) the start of the modern discipline that is today called Forest Management. In a sense, this new philosophy was an adaptation to forestry of ideas already embedded in commercial agriculture. That is, instead of merely being willing to harvest a crop (wood) as grown by Nature, the need for human intervention in the forest was recognized in order to improve the quantity and quality of available wood. At the same time, humans began to understand the need for improved stewardship of forest land and the responsibility inherent in the utilization of natural resources. As a result, embryonic "seat-of-the-pants" Forest Management began to yield to a desire to make the subject more "scientific."

Thus in the 1930s a variety of quantitative forest management aids began to appear. These included tools such as site index curves for measuring site productivity, yield tables for estimating the future wood volume of existing forest stands, and volume tables for determining the amount of merchantable volume in standing trees. Initial techniques were relatively crude and methodologies for the creation of these tools were not standardized. Consequently, the years following the 1930s saw improvement and standardization in the ways these were created, and a subsequent increase in the reliability of each.

By the 1960s, many problems inherent in the development of these early quantitative forest management tools had been resolved. Furthermore, researchers began to recognize that certain limitations were related to the quantity and quality of data available, rather than to existing analytical techniques. Thus, although some research continued on these and similar subjects, a large number of researchers began to shift their emphases. Instead of trying to build better management tools, considerable effort was devoted to trying to obtain better information about the forest resource.

At this point, considerable effort began to be spent on refining forest inventory. Researchers sought to determine things such as the "optimal" sample size, the "best" inventory design, and the "ideal" field procedures. A critical part of this work was the utilization of aerial photography to improve forest inventory estimates. In fact, a number of researchers focussed specifically on how to obtain the maximum amount of information from aerial photographs while minimizing the amount of relatively expensive field work required. Arguably, it was at this point that remotely-sensed data (in the form of aerial photographs) began to be commonly viewed as an indispensable part of forestry data collection efforts.

Around the start of the 1970s, a significant revolution occurred as, for the first time, powerful computers became readily available to forestry researchers. One arena in particular whose development accelerated greatly because of this was growth and yield modelling. In one sense, growth and yield modelling was a catalyst for the integration of existing management tools, as well as a gateway to the development of increasingly sophisticated mathematical models. Another subject which benefited from the widespread availability of computers was Operations Research. At this time, optimization techniques such as linear programming began to be applied to Forest Management in harvest operations, forest fire control, and others.
B. PERSPECTIVE ON PRE-1980S RESEARCH

Each of the "eras" of forestry research mentioned in the previous section has advanced Forest Management as a science in an important way. Early graphical and tabular management tools initiated in the 1930s provided the ability to assess the amount of wood in an existing forest stand. Forest inventory techniques refined during the 1960s enabled foresters to estimate reliably the amount of wood present on a relatively large area. Growth and yield models integrated early management tools with forest inventory data to allow the characteristics of an existing forest to be projected into the future under a variety of management scenarios. Linear programming techniques provided a way to determine and evaluate the optimum way in which an existing forest can be harvested, regenerated, and treated silviculturally.

With this perspective, it is evident that a number of common themes run through forestry research from the early 1900s to the 1970s. First, management tools have become more quantitative and scientific. Second, as these have become more sophisticated, each has been able to utilize more ecophysiological factors to estimate quantities such as site quality or forest volume. Third, each "generation" of research has relied heavily on previous successes (and failures) to produce state-of-the-art management systems and tools. And, finally, though not an explicit goal, the evolution of forestry research has led to an ability to conduct increasingly localized forest management.

This latter point is the subject of this paper. By "localized forest management" we mean planning the utilization of the forest over large areas using tools and data which are reliable for a relative small area. For example, rather than knowing the mean volume for Type A over a 1000 ha forest, it is desired to prepare forest plans using the volume for a specific piece of Type A (an individual stand polygon) in the southeastern section of the forest. Similarly, rather than knowing, for example, the wood volume that will be produced 20 years from now if a silvicultural treatment is applied to half of a forest at the present time, it is preferable to know how much volume will result and where it will be located if a specific set of stands are treated. The inherent difficulty in being able to do this is that previous data collection procedures and analytical management methodologies have focussed on questions of "How much...?" and "What will happen on this forest if...?" That is, existing technology - hardware and analytical techniques - and data collection costs have necessarily precluded the use of location from most modern forest management tools. We believe that the next questions that the field of forestry research has the potential to answer and must address are questions of "Where should...?" and "What will happen here if...?"

C. FORESTRY RESEARCH: 1980S TO PRESENT

By the mid-1980s, with the advent of desktop microcomputers, another technological revolution had progressed considerably. At the same time, microcomputer software and hardware had become increasingly sophisticated while its price continued to decrease. This has led to the development and application to forestry of computer-intensive subjects such as artificial intelligence and expert systems. The focus of this paper, however, is another discipline whose development accelerated with the introduction of microcomputers: digital remotely-sensed data and geographic information systems (RS/GIS).

The digital analysis of remotely sensed data for forest management is a relatively new technology. Research in the 1980's focussed largely on the use of satellite remote sensing for carrying out large scale surveys for activities such a long range planning and the production of synthesis maps showing broad classes of species at low resolution [Beaubien, 1983; Leckie, 1990]. The late 1980's brought new imaging sensors on airborne platforms with high spatial resolution and larger dynamic range than aerial photos [Till et al., 1983]. New digital sensors are planned for the 1990's which provide spatial coverage close to that achieved with aerial photography [Neville and Till 1989]. Images acquired with these and other modern sensors have the potential of replacing...
the older, analog technology and hence increasing the efficiency of forest map production, especially through direct interfacing with GIS technology.

GIS is likewise relatively new to forest management. This is most obvious when one considers that the largest source of data for GIS is found in existing paper maps. For the most part, practitioners are still "converting" paper maps to their equivalent electronic form and have not yet moved on to the creation of electronic maps directly (for example, directly from image analysis).

Both GIS and the analysis of remotely sensed images should allow one to answer the question "Where?" in Forest Management. (Collectively, we refer to the data, hardware, and software used for the collection, storage, retrieval, and analysis of numerical, geographically-referenced data as "digital spatial technologies" or DSTs.) However, a myriad of problems exist before this can become reality. Three specific problems will be discussed here which we believe to be the most immediate barriers to the adaptation of existing forest management tools to a spatial framework using DSTs: availability of appropriate data, knowledge of suitable statistical techniques, and inability to quantify and utilize spatial errors. These problems are presently targetted by the research program of the Industrial Chair in Geomatics Applied to Forestry.

2. DATA

Research involving forest inventory has led to a relatively standardized methodology in North America. Usually, the first step is a photo-interpretation-based stratification of an area into distinct forest types. The goal of this stratification is to divide a relatively large population into relatively uniform sub-populations (strates in Quebec) in order to increase the precision of subsequent population estimates of volume. In statistical terms, the goal is to minimize within-group variance and to maximize between-group variance. After stratification, a sample size for on-site data collection is determined and individual samples are distributed among sub-populations (strates) based on the variance, importance, and/or size of each strate. The cost of data collection can also affect the number of samples in each strate. After determining the distribution among strates, each sample plot is located on a map and established on the ground. A field crew then collects all data desired.

In this process, if it can be assumed that each sub-population is "perfectly uniform," then the location of a plot within a particular sub-population is unimportant. However, it is well-accepted that, for example, a linear arrangement of plots which follows existing topography is not desirable. It is also considered unwise to place all samples for a particular sub-population within a single polygon if, for example, 20 polygons of that sub-population exist in the forest population. These examples suggest that, despite stratifying a forest into relatively uniform groups, the specific location of samples does remain important.

Therefore, the first question that must be answered relative to using existing forest inventory data to localize forest management is "Are the data collected for a strate in one location suitable for the same strate in a different location?" Current forest inventory procedure assumes that they are. Or, to be fair, this assumption must be made if conventional forest inventory data are to be used to make forest-wide estimates. If this is true, then existing forest inventory data can be adapted to DSTs relatively quickly because forest estimates obtained at a given location within a particular strate will be applicable to other locations of the same strate.

Research Priority 1: The spatial change in forest parameters within a given strate must be examined and fully characterised (Figure 1).

For those strates where forest parameters are found to vary significantly with location, new ways to collect locally-reliable forest inventory information are needed. One possible way to do this is
through the use of digital remotely-sensed data. New sensors are continually collecting digital data with increasing spectral and spatial resolution, improved hardware is allowing more complex analysis, and better computer-based analytical techniques are constantly being developed. For example, whereas early Landsat satellite images provided for the identification of broad vegetative classes only, work is ongoing to determine if certain bands of radar can produce reliable estimates of forest biomass [Sader, 1987]. If this work is successful, potentially it will be possible to produce estimates of forest volume “automatically” with a minimum of costly ground-based data collection. Also, analysis of high resolution digital airborne data may lead to the ability to identify the species, age, height and local density of individual trees [Gougeon and Moore, 1989] or the percent stocking of regenerating cutovers [Edwards et al., 1991]. Data on individual trees could then be generalised to obtain continuous maps rather than category maps showing species location. This kind of map is likely to be much easier to produce from digital imagery than the maps produced by human photo-interpreters. If new management techniques for exploiting this information can be developed, in the long term such locally-reliable results may also be more useful.

Research Priority 2: The ability of remotely-sensed data to produce locally reliable forest estimates must be examined as a supplement to, and possibly even a replacement of, existing forest inventory procedures.

The single largest obstacle to understanding and using remotely sensed data at this level of localisation is validation. Most of the information which could be used for validation does not exist at the level of individual trees. Even ground inventory data does not give information about individual trees. Furthermore, even if information was available on isolated trees, they cannot readily and unambiguously be identified in remotely sensed imagery. Geopositioning can help, but experience to data indicates that even with GPS position to less than a meter resolution, locating sample plots is still difficult. Hence new methods of validating very precise local information must be developed, methods which are economical as well as being reliable.

Research Priority 3: New ways of validating local information using both existing information sources and new but not too costly sources must be explored.

A second possibility exists for obtaining locally-reliable forest inventory data, a method which might also be used to help validate the remotely sensed data. It may be possible to develop procedures to collect these data as part of conventional normal forestry operations with little additional effort. For example, when a forest is harvested, the first step is generally to clear the roads which will access the area. In Quebec the wood cut from the roads is harvested, scaled, and taken to a factory for processing. If the wood were scaled by strate, and the area of the road from which it was cut were known, locally reliable forest inventory data could be obtained (Figure 2). The only additional cost of such a scheme would be piling the wood by strate and measuring the road length of the strate from which it came. (Road width could be assumed to be fairly constant, and/or spot-checks of road width could be conducted.) For such an approach, a long-term effort would clearly be required in order to establish a spatial "data dictionary" for a large area using a GIS. However, once established, a minimum of effort would be required to maintain and update such a system. Prior to DSTs, it would not have been possible to think of creating such a management tool. Even now, it remains unclear whether existing DSTs in general, and GISs in particular are suitably developed for the storage and manipulation of such data.

Research Priority 4: Operational procedures for the collection of locally-reliable forest inventory data must be developed.

Finally, the density of forest inventory data collected must be examined. As noted, data are collected currently to be "forest-reliable” and, in general, are well-suited for managing "average conditions” within each strate. However, some strates will have enormous within-strate variability
while others are likely to be uniform over a large region. This varying uniformity of strates and individual stand polygons has tremendous implications for estimating the area of each strate within a forest, as well as the amount of wood volume within each. To make these data locally-reliable, it may even be necessary to increase the number of samples while (possibly) reducing the size of each in order to maintain present costs.

**Research Priority 5:** To determine the necessary sample density to obtain locally-reliable forest inventory data, the within-strate and between-strate change in various forest parameters must be examined for a variety of conditions.

### 3. STATISTICAL TECHNIQUES

Paralleling trends in other research disciplines, the development of analytical tools in forestry has relied heavily upon the use of parametric statistics. Principal among these are regression analysis and the analysis of variance. However, the use of parametric techniques requires that a number of a priori assumptions concerning the data be satisfied. The most common of these are that each observation is independent of the others (which is almost always violated in the construction of growth models), variance along a regression line is homogeneous (almost always violated when biological data are analyzed), and prediction errors are normally and randomly distributed (rarely verified). In the adaptation of these statistical techniques to the analysis of spatial data, these assumptions become even more dubious. The principal reason for this is that, when distributed spatially, almost everything is interrelated, non-uniform, and non-random.

Another class of problems exists because of the general nature of map-based data. Most existing analytical techniques in forestry research utilize continuous data. For example, growth models project quantities such as basal area or stand diameter into the future. However, spatial data are never continuous when stored as a map. Thus existing analytical procedures for these geographic data are generally based on discrete variables — e.g., 5 m elevation classes. Furthermore, most forestry research is based on known, unchanging observation units -- individual trees, for example, or 0.1 hectare sample plots. However, the base unit on a map is generally a polygon. In a natural forest system, the spatial characteristics may change over time (Figure 3). This causes problems statistically, and also in how these polygons can be stored and modelled using DSTs. To adapt existing forest management tools to a spatial framework, analytical techniques which utilize discrete data must be explored to enable the modelling of the time-based change in the attributes, shapes, and sizes of map-based polygons. Furthermore, spatial models must be able to account for the spatial dependency among geographic entities - something that is rarely included in existing forest management tools.

**Research Priority 6:** Existing and alternate data structures and operations for GISs must be explored for the use of locally reliable forest inventory data.

Statistical techniques are not the only problem, however. As noted, maps are designed to store discrete data. As a result, with existing DSTs, geographic data are stored either as a series of grid cells each of which is one and one class only, or as a series of sharp lines which separate distinct regions. (This will also be discussed later in "4. ERRORS.") Conversely, "real world" forestry data are continuous. For example, trees grow continually in minute increments. Ecological succession occurs continually but gradually. A probabilistic approach may offer a viable bridge between these two. As envisioned, a predictive system for DSTs would not merely assign an entity to one or another class. Instead, at each time period of interest, the probability of a given entity being a particular class would be assessed. Thus the result of a spatial predictive model, as envisioned, would be a surface representing the probability that a given map entity — raster grid cell or entire polygon — would be assigned to a given class at some point in the future. A potential problem with this is that such probability surfaces may be difficult for humans to interpret.
However, as each entity could be assigned to a class in a fashion representative of its probability and a thematic map produced for human comprehension. The difference between this and the usual practice is that the thematic map used for human interpretation would not be used as "data" in subsequent processing - only the underlying probability surfaces would be used for further analysis.

**Research Priority 7**: Statistical modelling techniques for discrete map-based data must be developed. This includes examination of existing aspatial statistical techniques, the use of probabilistic techniques for predictive maps, and the examination of spatial relationships among map-based entities.

4. **ERRORS**

In most scientific research, an assumption is made that a parameter of interest can be measured "without error" (subject only to the limitations of the equipment used for measuring). Relative to this, a number of DST researchers have concentrated on the analysis and effects of errors of digitizing, map registration, and/or satellite image referencing. It is argued here that, in forestry, these errors are generally relatively unimportant (except for obtaining positions for validation purposes) and are akin to errors of "equipment limitations." This is not to say that these are unimportant in all disciplines. Indeed, in those mapping subjects which utilize political and/or man-made boundaries — we call these "razor sharp boundaries" here — these errors are the major categories with which one must be concerned. That is, in such situations each boundary is a line which has no width and two identifiable endpoints which do indeed separate two distinct entities. Furthermore, that line can be physically located on the ground if necessary.

This is one of the most subtle yet most important differences between maps of natural phenomena, and maps involving human geography. It is especially important to note that the design of most DSTs — especially GISs — is based on concepts inherent in human geography. Notably, the idea of razor-sharp boundaries is not applicable to natural forest stands. In forests, not only is the "true" location of any boundary in doubt, but the very existence of a boundary may be questionable. Any forester who has walked through the woods holding a map which was developed through a reliable photo-interpretation process knows the difficulty of actually identifying forest-type lines on the ground. Furthermore, natural forest types rarely change abruptly as one moves through the forest. Rather, one will notice that one has gradually moved into a different forest without being able to identify where the change occurred. Even the "distinct" break between a clearcut and a mature forest stand may be difficult to locate precisely on the ground. This suggests an idea for natural forest stands which we call "fuzzy boundaries" (Figure 4).

We call boundaries fuzzy when one (forest) type gradually becomes another and the exact line between the two types cannot be identified on the ground without "considerable" error. In this sense "error" is equivalent to "uncertainty" in that there is a level of confidence associated with the existence and location of each boundary, and with the polygon type identified on either side of a boundary. This contrasts with maps of certain other phenomena for which boundaries are definite and only the polygon type is uncertain. Thus we believe that the concept of "fuzzy boundaries" in forest stand maps necessitates the examination of two separate but related errors.

The first is boundary error which exists if a boundary has been identified which does not really exist, or if an existing boundary has not been identified. The situation in which a boundary which definitely exists but is located in the wrong place is also in this category. The second type is category error which occurs when boundaries have been correctly identified, but the polygon on either side of the line has been mislabelled. Both types of errors have long been recognized in the classification of remote sensing data, although only category error is usually measured. A
boundary error might involve the failure of an automated classification procedure to distinguish between two fields, whereas a category error might involve the mislabelling of a corn-field as a wheat-field. In a forestry setting, both errors are related and the probability of observing one or the other depends largely on the characteristics of adjacent forest stands. This will be discussed shortly.

To understand the current inability to incorporate these types of errors into DSTs, one must examine the historical background of maps. The existing form of maps has been guided largely by the technical limitations inherent in pen and paper, and by limitations of human ability to interpret "pictures." Thus present-day thematic maps are composed of a series of polygons which are drawn as having razor-sharp boundaries. Generally, these maps are satisfactory in that they are easy to produce and easy to understand. Furthermore, if polygons are well-defined - provinces or townships — the units represented on a map are "humanly-meaningful" entities. However, with such maps it is difficult to represent a geographic surface of a continuous variable. For example, "topographic maps" which show a continuous variable — elevation — may be viewed as little more than a series of long, narrow polygons which place elevation into discrete classes. Furthermore, for some variables (such as population density) the map scale can radically change the content of the map.

The point of this discussion is not that existing maps are inherently flawed. With pen and paper and human comprehension, it is understandably necessary to represent the world as a series of discrete polygons which have razor-sharp boundaries. However, the point here is that such a map concept is poorly suited for the spatial adaptation of existing forest management tools in order to localize forest management. Simply, the existing form of physical maps has inherent limitations for modelling forest conditions over space and time.

We believe that the technological sophistication of DSTs allow a redefinition of map concepts to better facilitate spatio-temporal modelling. We further believe that the most immediate need in forestry is the ability to incorporate uncertainty (or "error") into forest stand maps for storage, analysis, and modelling. The first step in this effort is the determination of how to measure boundary and category error. This can start by examining selected forest parameters in specific controlled settings. For example, it is intuitively reasonable to expect that the boundary between two highly variable forest stands having similar mean densities will be fuzzier than between two extremely uniform stands with widely different mean densities. Similarly, it should be possible to infer the likelihood of a category error for various positions within a stand polygon or a given shape. In a circular stand, for example, the smallest chance of category error would seem to exist at the center. Whether or not this is true must be evaluated.

**Research Priority 8:** Forest parameters must be examined and appropriate measures must be developed to quantify boundary and category errors.

5. PROJECTIONS FOR THE FUTURE

We believe that existing DSTs must be altered considerably for application to forest management. An "ideal" system will have an error map associated with each thematic map stored, some of which will result from more sophisticated remotely-sensed data and analysis algorithms. In forestry, operations such as "Overlay" which superimposes one supposedly precise map on another should largely disappear and give way to operations which work across fuzzy polygon boundaries. This will only occur if emphasis in the analysis of forest stand maps shifts from the individual polygon to the boundary between two adjacent polygons. (We call this unit a "twain.") This shift can be facilitated by the development of spatially-based probabilistic modelling techniques. These techniques will allow foresters to ask questions of "What will happen if...?" just as current aspatial management tools do. However, the answer to these queries will be a map surface which shows
the probability that a given outcome will occur at a particular location rather than provide a
deterministic result (Figure 5). To convert these surfaces to something more useful for presentation
and human interpretation, the development of these modelling techniques will have to be
accompanied by a robust set of class assignment techniques to produce simplified thematic maps of
a form similar to that which exists presently.

6. CONCLUSIONS

Three major barriers exist which hinder the use of DSTs for locally-reliable forest management:
available data, suitable statistical techniques, and the use of spatial errors. Data limitations possibly
may be overcome by using smaller forest inventory plots but more of them, collecting data as part
of forest operations, constructing and maintaining long-term spatial data dictionaries and by
increasing the use of high resolution digital remotely sensed data. Non-traditional statistical
techniques must be examined which can accommodate discrete (categorical map class) variables,
and which are robust enough to tolerate some violation of underlying assumptions. Perhaps non-
parametric statistical techniques must be used. The initial step in the use of errors is to be able to
quantify different types of uncertainty inherent in forestry maps. After this, new analytical
operations which utilize error information as useful knowledge can be developed.

7. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contribution and support to this work of the Association
des Industries Forestières du Québec and the National Science and Engineering Research Council
of Canada.

REFERENCES:

Beaubien, J. [1983]: Une méthode de rehaussement d'images Landsat pour la classification du
couvert végétal. Proceedings of the 8th Canadian Symposium of Remote Sensing, Montreal,
Québec, 559-566.

Resolution Airborne Imagery Via Texture Measurements. Proceedings GIS Forestry '92,
Vancouver, B.C. (in press).

Gougeon, F.A., and T. Moore [1989]: Classification individuelle des arbres à partir d'images à
haute résolution spatiale. Télédétection et Gestion des Ressources: Sixième Congrès de
l'Association Québécoise de Télédétection, Bonn and Gagnon (eds.), 185-196.

Leckie, Donald G. [1990]: Advances in remote sensing technologies for forest surveys and

Proceedings of the International Geoscience and Remote Sensing Symposium '89, Vancouver,
B.C., 2390-2393.

Sader, S.A. [1987]: Forest biomass, canopy structure, and species composition relationships with
multipolarization L-band synthetic aperture radar data. Photogrammetric Engineering and Remote

Figure 1. Populations where characteristics do not change with location (left) and where variance changes (right).

Figure 2. Isolated sample points for global estimates (left) Potential road network sample for local estimates (right).
Figure 3. Change over time in attributes and polygon form.

Figure 4. Map whose “real world” boundaries are sharp (left) and areas whose “real world” boundaries are not sharp. Note that the map representation is the same for both.

Figure 5. Deterministic (left) and probabilistic (right) maps of “suitable locations.”