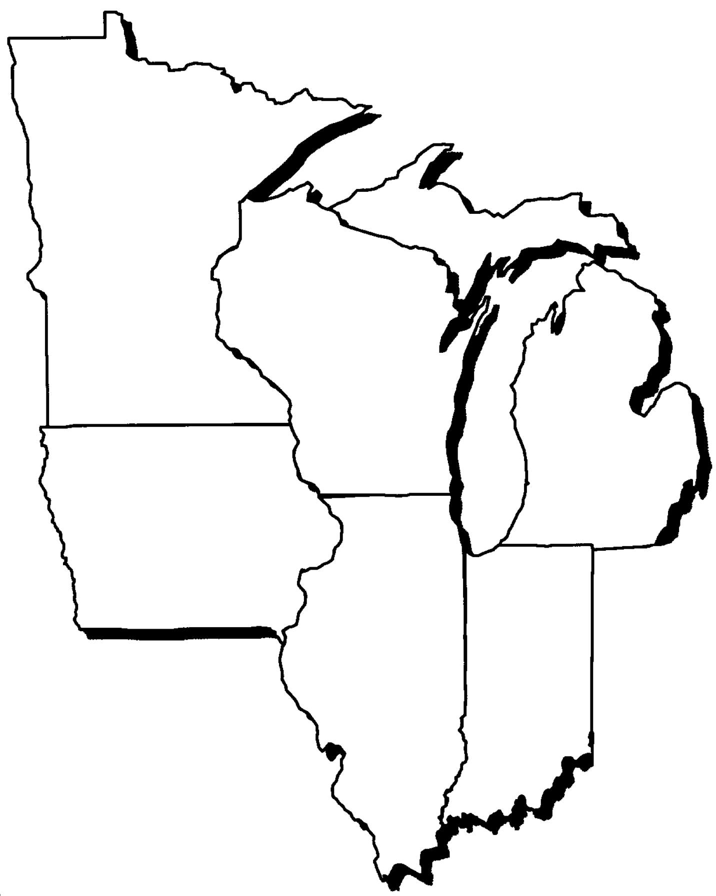
Environmental Management Technical Center
National Biological Service
U.S. Department of the Interior

96-G001

Proceedings Upper Midwest GAP Meeting and Workshop

January 18-19, 1995
Madison, Wisconsin



April 1996

Proceedings
Upper Midwest Gap Analysis Program Meeting and Workshop

January 18-19, 1995
Madison, Wisconsin

Hosted by

National Biological Service
Environmental Management Technical Center
575 Lester Avenue
Onalaska, Wisconsin 54650

April 1996

The Environmental Management Technical Center was established in 1986 as a center for ecological monitoring and analysis of the Upper Mississippi River System.

The National Biological Service . . . gathering, analyzing,
and sharing the biological information necessary to support the wise
stewardship of the Nation's natural resources.

Printed on recycled paper

Contents

Preface

The Gap Analysis Program (GAP) is a National Biological Service project being implemented nationwide with the help of more than 400 cooperators, including the private sector, nonprofit organizations, and government agencies. The purpose of GAP is to identify "gaps" in the network of conservation lands with respect to land cover or habitat types as well as individual vertebrate species and to build partnerships around the development and application of this information (Scott et al. 1993).

The National Biological Service Environmental Management Technical Center facilitates the Upper Midwest GAP, which is currently under way in the States of Illinois, Michigan, Minnesota, and Wisconsin. The project is in various stages of planning or startup in the States of Iowa, Indiana and Ohio.

Gap analysis is conducted by combining the distribution of current vegetation with the predicted distributions of vertebrates as indicators of biodiversity. Maps of species-rich areas, individual species of concern, and overall vegetation types are generated. Using geographic information systems, gaps in biodiversity conservation are identified by overlaying these biodiversity maps with stewardship maps. The resultant information can be analyzed to show where efforts can be most efficiently focused to achieve broad-based conservation of biodiversity. The gap analysis component of the Upper Midwest GAP will be undertaken by the EMTC in close cooperation with its State partners.

These Proceedings document the Upper Midwest GAP Meeting and Workshop hosted by the Environmental Management Technical Center January 18-19, 1995 (see Appendix A for the meeting agenda and Appendix B for a list of participants). During this meeting, status reports were presented by the States of Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio, and Wisconsin. Meeting attendees also participated in group discussions about the protocol being designed for processing Landsat Thematic Mapper imagery for land cover classification and the status and distribution of satellite imagery.

Additional copies of this report may be obtained from the National Technical Information Service, 5285 Port Royal Road, Springfield, Virginia 22161 (1-800/553-6847 or 703/487-4650).

This report may be cited:

National Biological Service. 1996. Proceedings, Upper Midwest Gap Analysis Meeting, January 18-19, 1995. National Biological Service, Environmental Management Technical Center, Onalaska, Wisconsin, April 1996. EMTC 96-G001. 66 pp. + Appendixes A-B

Proceedings
Upper Midwest Gap Analysis Meeting and Workshop

January 18-19, 1995
Madison, Wisconsin

Hosted by the National Biological Service,
Environmental Management Technical Center

Abstract

The Environmental Management Technical Center hosted the Upper Midwest Gap Analysis Program meeting as a follow-up to implementation of the Program, which is being coordinated by the Environmental Management Technical Center. Status reports were presented by the actively participating States of Michigan, Minnesota, and Wisconsin. Representatives from the States of Illinois, Indiana, Iowa, Missouri, and Ohio provided reports and participated in regional coordination discussions. Attendees also participated in group discussions about the protocol being designed for processing Landsat Thematic Mapper imagery for land cover classification and the status and distribution of satellite imagery.

Welcome

Frank D'Erchia
National Biological Service
Environmental Management Technical Center

I would like to welcome you to the Upper Midwest Gap Analysis Program [UMGAP] Workshop and to Madison, Wisconsin. It's good to see you all again; I look forward to a very productive meeting. This meeting may be a little bit overdue, but a lot of progress has been made since we last met, and it will all be presented here over the next couple of days.

When you registered, you received a packet which contains the agenda, some information about the National Biological Service [NBS], and other related information. There are a couple of handouts about the National Biological Inventory Infrastructure or NBII, which is a program the NBS is facilitating to make information and data available over the Internet. The National Gap Analysis Program (GAP) also has a Home Page. We'll be implementing much of this on a regional basis, but a lot of data will be available over the Internet.

The last time we met for a technical session was at the Environmental Management Technical Center [EMTC], and we planned to conduct several pilot efforts to classify satellite imagery when we left there. Dr. Tom Lillesand, Director of the Environmental Remote Sensing Center at the University of Wisconsin here in Madison, coordinated the development of a protocol for data processing in conjunction with the Geo Services Section of the Wisconsin Department of Natural Resources. Many of you have received draft copies of that protocol for review. That will be presented in detail here today by Tom, and we will have a sign-up sheet out at the desk for anyone who would like to get on the GAP mailing list for this region. We would like to get your Internet address, if you have one, and then a check-off if you want to receive a copy of the protocol. We will be taking recommendations here and modifying the protocol, and once that's finalized we'll put it out as an internal publication. We also hope to get it on the Internet at some point.

Since we met, we received the first dual-date imagery scenes for Wisconsin, which were the two scenes used to develop this protocol. Dan Fitzpatrick will be talking about the status of the Thematic Mapper [TM] imagery. We've had a rough beginning with that, but we will be seeing TM data coming in more quickly now.

I'd like to briefly discuss the purpose of this meeting. The focus here is to present this protocol for processing satellite imagery, to discuss and review the procedures, and to assess the ability of the three states to meet these protocols or some modification of them. Obviously, we want to develop a comprehensive land cover database that covers the three states. In addition, we have representatives here from the states of Illinois, Indiana, Iowa, Missouri, and Ohio, and they will each give us a status and update report. Representatives of those states have expressed interest in coordinating with us on a land cover mapping effort, so we can develop something that would be consistent across the whole Upper Midwest. This is the time to do that, as most of those states are just beginning in this effort and we are in the process of just beginning to process data. I hope we can develop a strategy to coordinate, because we will have a very usable database in the long run.

The discussion today will also focus on the classification system and the status of the National and State Gap Analysis Programs. We have the classifications system that was proposed in the protocol. Don Faber-Langendoen from The Nature Conservancy has reviewed that system and will be presenting some information to us related to that. The Nature Conservancy has been contracted by GAP to develop a comprehensive classification system that can be used nationwide, and we're looking to be able to crosswalk into that system.

Mike Jennings is here from the National Gap Analysis Program. Mike works with Mike Scott at the University of Idaho, where the National Program is managed, and he will provide us with a status of the National Program and answer any questions you may have. State representatives for the individual states of Michigan, Minnesota, and Wisconsin will present the status of those states' efforts.

The other item we may address, which is not on the agenda, is the funding for this effort. The most successful effort to date to develop a land cover database in this area has been the WISCLAND [Wisconsin Initiative for Statewide Cooperation on Landscape Analysis and Data] effort in Wisconsin. The WISCLAND effort has been very successful in gaining partnerships and is moving along very well. The Gap Analysis Program plays an active role as a partner in this effort, and because of the successful coordination with other agencies, WISCLAND has attracted many partners and contributors. We've also gotten some generous funding in the first year from the U.S. Environmental Protection Agency [EPA], through the Great Lakes National Program Office [GLNPO], to kick off this effort. After some delays in our first year in NBS, we received our funding toward the end of the fiscal year, and now we have cooperative agreements in place with the three states. This year we are in a little better shape, because the funding is secured and we are basically trying to determine how to distribute it. As we all know, the GAP funding itself is not enough, as the WISCLAND effort can attest. So, we will basically be talking to the partners and those interested in cooperating with us, either here during the meeting or in the next few weeks, few months, and into the future, because the success of this program will depend upon the partnerships we develop.

We have close to 60 people coming to this meeting; I had originally thought we would have about 30, so we doubled, and I think we are going to have a very productive meeting!

At this time, I would like to introduce the first speaker on the Upper Midwest GAP database development, Daniel Fitzpatrick. Daniel was hired by the EMTC as a biodiversity coordinator, to oversee the Gap Analysis Program effort.

Upper Midwest GAP Database Development

Daniel Fitzpatrick
National Biological Service
Environmental Management Technical Center

Though I'm here to talk about the development of the GAP database, I've been at the EMTC for only 2½ months now, and quite a few of you in the audience have been part of the GAP development from the original meetings which started, I believe, almost 2 years ago. So, actually, a lot of you know more about this than I do.

The Gap Analysis Program in the Upper Midwest is a subset of a regional biodiversity assessment that is the outcome of a series of meetings involving land use planners, resource managers, and ecologists. A couple of working groups were formed, the Upper Midwest Biodiversity Assessment Group and the Upper Great Lakes Biodiversity Committee. Out of the work of those groups, a seven-step design was developed, which some refer to as GAP II or GAP Plus, for the assessment, management, and monitoring of biodiversity. The official national GAP effort is essentially the first three steps of that seven-step design.

In the first three steps of GAP II is the assessment of ecological unit characteristics, including the development of an ecological unit map that all three states agree upon, so that unit boundaries which cross state boundaries will not become a problem. This effort was commissioned by the Upper Great Lakes Biodiversity Committee and for the most part is finished. I would challenge the participants in that effort to see it through to the development of a digital Geographic Information System [GIS] format with a resolution comparable to the resolution we'll get from the satellite images. This ecosystem map is already being used, as you will see discussed later, in the protocol for the satellite image processing. You will see some interesting overlays with that. Another layer that was identified as a necessary component of GAP II was that of pre-European settlement vegetation. Those coverages, I believe, are already developed, based in part on some innovative work being done by the Forest Service using Public Land Survey data and GIS techniques. Again, this is for GAP II and not specifically a part of the official national GAP effort.

Most of the next 2 days will be dedicated to reviewing and discussing the status of the current vegetation or land cover map development. I won't get into the details of that classification effort now. Dr. Tom Lillesand will discuss that more thoroughly later today. However, I would like to address those end users who have concerns regarding the suitability of satellite-image-derived classification to meet their needs. We are pursuing a classification following an image-processing protocol that should give us a uniform vegetation map across three states with a classification system that will be followed by all three states, and which is repeatable and highly defensible. When it gets down to determining whether or not vegetation classes are specific enough to meet end user needs, please keep in mind that we are not doing anything in image processing that precludes extracting particular classes and applying additional techniques or ancillary data, such as Forest Inventory and Analysis (FIA) plot data from the Forest Service, to extract stand structure or other additional attributes that end users might require. It is a hierarchical, iterative process, and we are now doing just the first step in building a base map. Nobody should be expecting to use our current vegetation map to do on-the-ground site management. That never was the intention. We also need to keep in mind that we don't have the luxury of being able to develop the best or the finest resolution current vegetation map of the Upper Midwest before we can take any kind of action to implement ecosystem management. We are building the database incrementally so that we can begin to use it as quickly as possible, and that's part of the process.

Another layer in the GAP II database is that of species-range maps. This is also an official component of the national GAP effort. There is no single best approach to developing species-range maps, and I am neither a wildlife nor vegetation ecologist. My background is in GIS and forestry. I am therefore going to be relying upon the cooperating states and their ecologists to direct me in the development of species-habitat models and the development of a species-range mapping protocol appropriate for the Upper Midwest.

The last component of the official national GAP effort is that of a public-ownership- and management-status map layer. This layer is a compilation of existing maps by all the individual public land managers into one comprehensive map of all public ownerships. We are hopeful that ownership boundaries for tribal lands and for industrial forest land will also be included, but management status categories for those lands is of course proprietary, so will not be likely included. The development of this layer will be incremental. After sending out the first draft for review, I'm sure we can expect reports of omissions and commissions.

These are the basic components of database development: current vegetation, vertebrate species distribution, and public ownership. At this time we are only beginning the vegetation classification effort, so there isn't a lot more to say about the other components. The question does come up, however, and Frank has alluded to some of this; what will we do with the database when it is done? How will people have access to it? The answer is that it will become public domain, and to the extent possible, we intend to put the current vegetation coverage at the full resolution of 30-m pixels on the Internet for anyone to download. This will be in addition to generating vector maps to meet national GAP standards. Precisely how we disseminate the data is yet to be determined. The Internet is a rapidly evolving entity. Anyone actively using it knows that there wasn't really much data to access 2 years ago. What data formats and dissemination protocols will look like 2 years from now is anyone's guess.

I will be authoring an Internet World Wide Web Home Page for the Upper Midwest GAP, and I think we will also be offering to sponsor a Home Page for the Upper Great Lakes Biodiversity Committee and the Upper Midwest Biodiversity Assessment Group. I also intend to develop an electronic newsletter, and so that people can work through some of these issues as they evolve, we'll try to facilitate communications by hosting an Upper Midwest electronic bulletin board or discussion list. I hope that maybe some of the adjoining states can get together with the three states that are already involved in this program and determine where they have common boundaries and common interests to see if Upper Midwest GAP can include a wider audience.

Land Cover Classification System Development

Don Faber-Langendoen
The Nature Conservancy

Thank you for inviting me here. I would like to update you on the work of the vegetation classification. One point is to clarify how it can be crosswalked to the protocols being developed for GAP, the other is to indicate that it is part of the larger, ongoing regional biodiversity assessment.

What I'd like to do is break the talk into two parts. One part is to introduce the full scope of the classification system, particularly as it applies to the Upper Midwest. The second part will look at how the classification system fits more directly into some of the protocols being developed.

The vegetation classification, in general, has four important features. One, it emphasizes biota, particularly vegetation, with a conservation focus. Second, it is hierarchical, based on existing vegetation. Third, it is intended for worldwide application. That is part of the reason for its complexity, much of which isn't needed at local levels. And fourth, it is a component of The Nature Conservancy and Heritage database. I would now like to expand on each of these features.

The first feature is that it emphasizes biota, particularly vegetation, with a conservation focus and within that it has two major objectives. One objective is that the classification, in part, is an attempt to understand vegetation patterns of ecological systems. In a sense, that is its scientific objective; to describe the pattern of the vegetation. The second objective is to provide a systematic framework for establishing conservation priorities. By having this framework, we can prioritize the kind of patterns in the world that need protecting. Whether we are talking about white pine-hemlock stands and their pattern of succession, or oak stands succeeding to maple, the system is designed to capture the kind of variation that exists today.

The second feature is that this system is hierarchical, based upon existing vegetation. The Nature Conservancy has at the very top level of its classification three major systems. The one which I will be addressing today is the terrestrial system, including uplands and wetlands. The second system is the aquatic, within which are the marine and fresh water subgroups, with the marine subgroup being currently more developed than the freshwater subgroup. The third system is the subterranean.

Within the terrestrial system, we've developed six sub-levels. The *class*, *subclass*, *group*, and *formation* are all physiognomic-type levels. Physiognomy simply refers to the structure of the vegetation, its outward appearance. For example, at the *class* level, forest is defined as trees greater than 5 m tall and having a canopy cover of 60%-100%. The *subclass* level breaks out the phenology of the leaves into evergreen, deciduous, and mixed. The *group* level further breaks down leaf types, the temperate and subpolar needle-leaved evergreens being one example. Within the *group* level, climatic names are often used as they are reflected in leaf types; temperate or subpolar as related to the needle-leaved character, for example. The lowest of the physiognomic levels, the *formation* level, is used to describe a combination of environmental factors that influence structure. We also refine some of that structure. For example, the *needle-leaved evergreen forest with conicle crowns formation* is distinct from a forest with rounded crowns. All these distinctions are in part designed to assist with mapping the vegetation.

Below the four upper physiognomic levels, *class*, *subclass*, *group*, and *formation*, we have two floristic levels which further refine the units. First is the *alliance* level, based on canopy cover species, where the uppermost strata of vegetation is emphasized. The *black spruce alliance* is an example. At the *association or community* level, we have reached the lowest level in the classification hierarchy. At this level, we define the units by describing the entire assemblage or grouping of plant species that co-occur in a loosely organized way. For example, the *black spruce-feathermoss association* describes the canopy and also the understory species. While it is certainly possible to classify vegetation to levels of greater detail for conservation purposes, we feel that this is the level that we would like to get down to, and it is also a very widely used level.

To clarify how structure is captured in the classification, let's compare forest and woodlands. Woodlands are defined as trees over 5 m tall, having a canopy closure of something like 25% to 60%. That is, the crowns are not touching, but they're not widely scattered either. If we compare evergreen forest and evergreen woodland, we'd come down to the needle-leaved types at the *group* level, and at the *formation* level, we have needle-leaved evergreen forest with rounded crowns as an example. We can also look at the *alliance* level, jack pine forest with a very dense structure, closed canopy compared to jack pine woodlands, with a more open canopy. Jack pine is used in both, but the structure of jack pine woodlands is different from that of jack pine forest, so the two jack pine *alliances* are distinct. At the next lower level then, below the *alliance* level, the *jack pine-blueberry-feathermoss association* is a different grouping of species than the *jack pine-red pine* with the *bur oak-* or *pine-oak-understory association*.

The second feature of the classification system is that it is hierarchical based on existing vegetation, and the emphasis on existing vegetation is important to remember. We're classifying what's out there now based on the combination of floristics and physiognomy, so successional stages that are floristically distinct are recognized as distinct categories. We interpret successional patterns, taking into account pre-European settlement and future stand development, but it's what's out there today that we classify. Because the classification system is based on existing vegetation, crosswalking to land use/cover mapping projects requires very minimal effort. I should point out, too, that building knowledge of existing vegetation also contributes to ecological land classifications and visa versa. Ecological land classifications are a good way to interpret the existing vegetation, so a lot of interaction can occur between ecological land classification and TNC [The Nature Conservancy] vegetation classification.

A third feature of TNC classification system is that it is intended for worldwide application. The hierarchical structure of the system is modified from the UNESCO [United Nations Educational, Scientific, and Cultural Organization], which developed a document called the International Classification and Mapping of Vegetation. The TNC has tapped into that document and modified it. The UNESCO system was developed for mapping at a scale of 1:1 million; countrywide surveys were the intent. We have modified the UNESCO System so it is applicable at a scale of 1:24 thousand, which is the scale that TNC Heritage Programs most often use for conservation planning. This means that the TNC system is also a benefit to our partners elsewhere in the world. We have programs in Latin America where there are tropical rain forest systems, and we also have partners where there are tropical dry forest systems. All types of ecosystems are incorporated into the UNESCO system, so we have the ability to address the world's vegetation. Again, on the local level we don't need that complexity, but that's where the TNC frame work is established.

Finally, the fourth feature of the system is that it is a component of the TNC Heritage conservation databases. So the system, whatever its worldwide perspective, is tied down to very specific databases here in the Midwest. Through inventory and sampling, classifications are developed, community types are recognized, and these are built into various descriptions. How do you recognize these things in the field? What's the condition of the stand that's required to meet the definition of the type? What are the minimum size requirements? How do you map these things? These are all components that we hope to flesh out through the GAP effort. We hope to do that to the *alliance* level for the Midwest types that you've seen. Right now they are often just names on paper, but we would like to be able to describe those so that the specification for recognizing them can be spelled out. We also then develop global ranking schemes so that we know how rare the types are across the states and around the world. That also leads us to the inventory part, where we do the element-occurrence tracking in the field. We develop what we call element-occurrence records, so that the location of these elements in the landscape is tracked by the TNC Heritage databases.

With respect to the use of the TNC Heritage classification for GAP mapping protocols, one of the things we would like to accomplish is to resolve where the differences in vegetation categories between WISCLAND and TNC Heritage classification exist. Also, as you have now seen, there is more detail in the TNC Heritage classification than can be realistically interpreted from TM Imagery. We can resolve which TNC Heritage classification units can and which cannot be detected in Landsat TM imagery, and in doing that we will begin to know what we can learn about existing vegetation through GAP efforts. It is also important to add that our interpretation of rare vegetation types depends very much on map scale. If elements are in an area too small to be detected at a given map scale, then by definition we cannot determine their rarity. In fact, rare elements are often the hardest things to detect at coarser map scales, so they are often missed.

Let me present where we are with the definition of vegetation types in the three-state area. At the top level of the classification hierarchy are usually nine *classes* of vegetation. No [upland] sparse or shrubland *classes* are currently known from the Upper Midwest, however, so only eight *classes* are being used in the three-state area. At the other end of the classification hierarchy, the *association* level, are 85 forest *associations* and 20 woodland *associations*, for a total of 222 *associations*. Of those, 28 are thought to be rare in the three-state area, including a number of upland prairie *associations* and wetland *associations*.   Those 28 *associations*, by the way, have been described in a report funded by the U.S. Fish and Wildlife Service (Bauer et al. 1994).

Looking at this in terms of the hierarchy itself, I've shown the number of units recognized at each of the levels in the TNC classification. Again at the upper level, are eight *classes* in the Midwest, and I have just broken down how many kinds of units occur at each of the levels. It should be noted that very little added information is at the *subclass* and *group* level. This is a reflection of the application of a worldwide hierarchical system to the Midwest condition, so some of these levels aren't needed very much for the work you are doing here. Note that in the three-state area, the biggest increase in the number of units is at the *alliance* level. If you can get the GAP classification down to the *alliance* level, you will be covering a lot of the kinds of vegetation in the Midwest. The *alliance* level has been suggested as the level that the GAP protocols can achieve.

In general, the greater the detail of your classification, the less protection those classification units will appear to enjoy. For example, if you are looking at protection or conservation of evergreen forest across the Upper Midwest, all the National Forests and many of the park units will protect that kind of vegetation. If however, you are asking about jack pine versus white pine evergreen forests, then the apparent conservation status changes. If you ask whether there is jack pine-feathermoss forest versus jack pine-blueberry forest, the degree of conservation tends to further decline. So, clearly the degree of mapping detail affects the interpretation of the conservation status of the map units or vegetation types.

There are two ways to improve the interaction between TNC classification and the UMGAP protocol. One is resolving how to define the top level, that is the class level, because the TNC hierarchy flows out of those top level definitions, that is forests, woodlands, shrublands, etc. If those upper levels are different, it certainly will complicate the interpretation of all the units down below. So that's one area to look for some resolution. Second, a way to improve resolution is to do what the UMGAP protocol is proposing; that is, to stratify the classification by TM scenes and by ecoregion (subsection). If you can sort out which vegetation types are in each ecoregion, you will greatly reduce the potential for mismatching the vegetation to the polygons.

As an example of that, if you look at the western Minnesota grassland region, particularly the northwestern part of it, and ask what kind of woody vegetation is there, you find one of the types that is present is the oak savanna. There are two dominant oaks, bur oak and northern pin oak. The only other major tree species in the uplands is aspen, so two major tree types are identified or found there. Quite a bit of open vegetation is found in that part of the country because it is prairie, so the forests are not very closed. So this kind of vegetation fits into what we call sparse woodland, some call it savannas or barrens, but it is scattered. Look then at how the vegetation classification would be applied within this ecoregion. If you look at the forest *class*, there are no evergreen or mixed forests, only the deciduous forest. In terms of *subclasses* is a cold deciduous forest and low-leaf submontane, broad-leaf cold deciduous forest. At the *alliance* level is a potential for only three kinds of forest *alliances*. Some maple-basswood are in that region, there is aspen forest, and there is bur oak forest. So by knowing that you are in this ecoregion (subsection), the number of vegetation units to be concerned about in terms of interpreting their TM signals are relatively few. But these three forest *alliances* are quite different from sparse woodland *alliances*. We only recognize two sparse woody *alliances*, an *aspen tall-grass sparse woodland alliance* and a *bur oak tall-grass sparse woodland alliance*. At both the forest and the sparse woodland levels, species in the aspen and the bur oak *alliances* are shared. If the sparse aspen woodland and aspen forest *alliances* are collapsed, you will lose the distinction between those kind of aspen types. The same way for the bur oak. If you collapse the *class* levels of sparse woodland and forest, the kinds of vegetation types are lost. Now if that's all that's possible (with the UMGAP protocols), then this is simply a reality to be noted.

Frank D'Erchia: Thank you. I think you made the point that the effort here is to get to the *alliance* level the best we can with satellite imagery. Beyond that, these add-on activities are where we need more detail, we get aerial photography, or we go on the ground or whatever it takes to break out those levels. We are trying to separate the idea of getting to that level for three states immediately with the data imagery funding available, from follow-up projects. The database is going to be dynamic, it's going to continue to grow and be enhanced.

Status of the GAP Analysis Program

Mike Jennings
National Biological Service

It's really great to be here and see the cooperators and people coming together from the three states. I think it represents something important and unique to those GAP projects that are running on a multi-state basis. There are a few others; the Mid-Atlantic states are operating as one GAP project. The six New England states started out as one project and have now gotten to the point where they have a very general land cover map. Now they have broken up into three individual efforts—Maine, Vermont, and New Hampshire as two separate entities and Connecticut, Massachusetts, and Rhode Island as another.

These multi-state cooperators are all taking the same generalized land cover map and are now going down to the alliance level of mapping with that.

These larger multi-state regional efforts represent a real phenomenon happening in the United States. A natural process is going on that coincidentally mirrors the political process we find ourselves in today at the national level; the reversion of greater power to the states, a lessening of Federal regulation, Federal oversight, Federal involvement in state-level government, and state level affairs. Those are two fairly coincidental things, and from the point of view of the natural resource profession, this has been coming for a long time. It's related to the development of the science of landscape ecology and the understanding that we needed to deal in larger, more biologically meaningful units. Obviously, the coupling of geography and ecology with the use of remote sensing for the purposes of large area management is related to the need driving it—dwindling natural resources and a variety of environmental crises which have been bringing people together. I see this kind of thing happening generally around the United States, even in areas where there are single state GAP projects, which describes most of the GAP efforts. States are also regionalizing on their own in the Southwest, the Southeast, and in the Northwest. So, you've gotten a jump on a lot of those efforts by approaching GAP as a regional program from the beginning. A regional program has some real up-front costs in terms of organization and getting people on board, and it is more awkward and difficult. It seems to me you are now on a threshold, though, to really cut loose.

This year's effort was a lot more intensive than last year's, but in a lot of ways more simple too. I also wanted to mention that I provided a pretty hefty background document to the first meeting of the NAFTA [North American Free Trade Agreement] Environmental Commission. When NAFTA met in their headquarters in Montreali, they were extremely interested in large-area mapping and the kinds of coordinated efforts that GAP is doing. I mention that here because of the obvious relationship between the Upper Midwest and Canada. I think that the dialogues are going to be ongoing.

I was really pleased with Don Faber-Langendoen's presentation on vegetation classification, and I am very interested in following the dialog on that. Compared to other places in the country, that subject can be explosive and divisive. I didn't get that feeling here at all. It generally seems as if people are on the same track in working out the two major issues that are raised. I'm delighted about that. That is also a very exciting development. That there has never been broadly accepted and consistent vegetation classification system for the United States is amazing. The response I get from that fact, over and over again is, "You mean that it hasn't been done? I thought that all those scientists went out there and did that a long time ago." We are now moving toward that. We funded a project a couple of years ago in New England through the TNC northeastern office, to develop a regionalized classification to ensure that the labels were consistent at the alliance levels and the attributes for those labels were consistent. This classification is having an enormous impact on natural resource management in the 13 Northeastern States, and I expect that we'll see a similar thing happen in the Upper Midwest as that system is used more and more.

At last year's meeting of the Ecological Society of America in Knoxville, Tennessee, we established an ad hoc group to discuss the idea of establishing a permanent committee within the Ecological Society of America to essentially be the gatekeeper for this taxonomy and to provide an arena where people can duke it out at the national level. The analogy here is to the American Ornithologist Union that essentially acts as gatekeeper for the taxonomy for birds. Every now and then, they revise that taxonomy and decide whose proposed changes to accept, and what not. Even though people disagree, they come away from that arena with a basic sense of agreement, and the amplitude of those disagreements will dampen, I'm sure, over time. It's an extremely exciting development, and it's very important. It's important enough so that this summer at the next meeting of the International Association for Vegetation Science, a portion of the meeting will be devoted to exactly this issue. It is being addressed because this issue has been raised by TNC and by GAP. That's turning out to be an important partnership and has very broad ramifications, so it's really an honor to be part of that.

I also want to briefly touch on another topic. On the national level, we recently received a review document that was commissioned by the forest and paper industry to review GAP analysis as a scientific concept. This is something that certainly had our attention when they let this contract almost 2 years ago. At that time, people speculated that the industry was hiring their experts to come out with their ideas for the eventual core challenges or legislative issues that might come out of discussions based on GAP information. The upshot of that commission is that the report was very supportive of the scientific concept. The reviewers laid out a number of issues that I thought were very constructive, important research issues for the future—the validity of the concept of species richness as an applied conservation measure, for example. That will be an entirely legitimate and very important research issue for years to come. We are hoping that in generating the maps under GAP for much larger areas and having them be consistent across at least the lower 48 states, we can begin to address those issues and work on them over time. We read this report with a great sigh of relief as it did endorse the direction that we are going, the basic scientific concepts, and it helped focus a future research agenda for us.

I also want to touch on GAP implementation; the issue of what we do with GAP data once we are done, and where we go from there. One of the other nongovernmental organizations that we have worked closely with, and that has been a staunch GAP supporter from the beginning, is the Defenders of Wildlife. The Defenders has just produced a fairly visionary document about where we could go with GAP once these datasets are constructed—educational outreach and related products, public involvement in decision making at the local level—a real wide variety of things, and it's fairly well done. So we still have quite a process to go through. GAP is not driven from the top, so it's going to depend strongly on cooperators. How the data are used is going to largely depend upon states and state agencies, who along with Federal agencies working at the state or regional level, are our number one clients. Many exciting things are going to come about, and again, the single greatest issue is producing data that is comparable across state boundaries. One of the things that I hear again and again is that as the Federal role in regulation is reduced over time, the importance of good, consistent information that is comparable across state boundaries increases. So states and state agencies can make the best decisions for their states, while having a regional context within which to make those decisions. I strongly believe that land managers here in the Upper Midwest are going to end up querying databases about the distribution of a certain element that they are responsible for managing. Managers are going to end up querying databases in many other states and be able to do so pretty effortlessly in the long run. All that will also relate to the development of the National Biological Information Infrastructure, of which GAP data is a part.

Though most of you probably are already aware, I want to briefly mention the GAP Home Page on the World Wide Web. It's going to be the center point from which people will be able to access state data. The Home Page has a map of the United States so you can click on a state and if there are data for that state, the map will route you to that node. There is also a bulletin board discussion, job announcements, standards, and a lot of other information including the complete text of the GAP monograph. Also, we are just about to put out our newsletter in which we will review what the different GAP projects are doing and present a variety of articles.

I want to mention also that we are getting under way with some pilot projects for an aquatic component of GAP analysis. We have pulled together a small brain trust of people to help think the idea through and to determine the approach we need to take for developing information at the national level. We are now at the same point as we were when we started terrestrial GAP. We do not yet have a commonly recognized aquatic classification system, though a variety of good systems exist. We hope that we can continue to work productively with TNC in addressing the issue of an aquatic community classification system, and I suspect we will end up with a kind of two-track structure that will use physical structure down to a certain level, and then break it out into biological components that seem to make sense. We still have a long way to go on that.

Finally, I want to mention a couple of things about the development of the Multi Resolution Land Characteristics [MRLC] Consortium. I can't really begin that discussion without recognizing Don Fields from the National Oceanic and Atmospheric Administration [NOAA]. Don is with the Coastal Change Analysis Program [CCAP] and is one of the original MRLC people that helped moved that group forward to make it all happen, and he helped us decide on imagery dates. To reiterate briefly, MRLC consists not only of GAP, but also the CCAP, the EPA EMAP [Environmental Monitoring and Assessment Program], the U.S. Geological Survey [USGS] NAWQA [National Water Quality Assessment] program, the EROS [Earth Resources Observation System] Data Center, and the North American Landscape Characteristics [NALC] Program. We hope to move forward a little bit with the MRLC effort by next addressing the development of a library of ground-control points for the United States that will be useful for a wide variety of purposes.

The state of the art for large-area mapping really is not conclusive as far as accuracy assessment is concerned. One of the ways that we can help that issue is by providing access to a large library of ground-control points. With that, a variety of different kinds of accuracy assessment calculations for whole maps, parts of maps, or classes of maps can be performed. If a user is not satisfied with the performance measure that we may attach to a map or any part of a map, they will be able to go back and redo it, or at least understand how many additional field points they may need to collect to achieve the data quality they need to do their analyses. So, that is the next sort of issue we hope to address through the MRLC.

So with that, I would like to open the floor to any questions.

What sort of cooperation or linkages have there been with the EMAP Landscape Characterization Group?

The EMAP landscape characterization group was part of the MRLC Consortium. They matched our funding for the purchase of the satellite imagery for the lower 48 states, which is now being marketed by the Earth Observation Satellite [EOSAT] as the best of the United States. There is extremely close coordination with EMAP landscape characterization, and they have bought into the TNC-UNESCO vegetation classification system. As a matter of fact, that classification system is being reviewed by the EPA directorate right now for the potential of being recognized as a kind of EPA standard. That is basically due to the relationship between GAP and EMAP's landscape characterization effort. The other relationship is, of course, funding. EMAP landscape characterization provided funding for GAP last year, and as a matter of fact, for the Upper Midwest Project, it really put GAP back on the tracks. Funding from the Great Lakes National Program Office really made a big difference to the Upper Midwest GAP budget last year. I am working with EMAP on developing some cooperative funding again for 1995. That hasn't quite happened yet, but it's imminent, so I would say that the relationship between our two programs couldn't possibly be closer.

You mentioned deemphasizing the Federal role in some of these conservation efforts to get beyond GAP as a entity itself, or are you heading towards voluntary private land participation in GAP or maybe a habitat conservation assessment approach?

Well, I guess there are two tracks there. One is simply providing good information to the plethora of programs that are operated by nongovernment organizations, by State agencies, and by Federal agencies. One track is simply to provide the datasets and to involve those groups in their use and in the feedback and the betterment of those datasets. Two, I mentioned the development of an implementation plan with an outreach component, and that is yet to be developed. One example is in Washington state, where there are hundreds of elementary and secondary schools involved in the GAP project in one way or another. Class trips go out and actually do groundtruthing. The data need to be quality checked, but it is a useful kind of community involvement. The state of Tennessee has a very large effort underway in the development of natural resource information for county level planning and strong support from the counties to use GAP data to help them put their counties into context regarding the distributions of the different elements of biodiversity. So, those are some examples. My earlier reference to the Federal role was really just a generic reference to a lessening of Federal regulation and the kinds of things we can expect to evolve under the existing political climate.

Do you have the address for the GAP Home Page Uniform Resource Locator on the World Wide Web?

Yes, it's http://www.nr.usu.edu/gap/gaphome.html.

General Overview of Draft Image Processing Protocol

Tom Lillesand
University of Wisconsin-Madison

The current version of the protocol document is 1.3. The intent is to have documentation in writing on how to approach the classification process. Our long-term hope is that we will end up with a systematic product, based on having a written protocol to follow to approach the development of that product. My dear wife said, "Well, you have this meeting that you are getting prepared for, what is it all about?" I showed her this document. She said, "This is a lot like reading instructions for a microwave." You know, you have 40 or 50 pages, the first 9 of which list hazards: Don't do this, don't unplug this, don't stand in front of it. What you really want to do is put the food in there, push a button, and get it out warm. I think all of us would like to take the TM data, put it in a computer, push a button, and get it out classified in a way that we would like. So, there is a lot to the analogy, but the intent here is to be able to say, "Gee, I've got to do this systematically in such a way that there is some hope for correspondence among the states, some hope for replication downstream." Believe it or not, we are all really tied into the concept of doing this very well once, when really, the database is going to be a dynamic one. We want to make sure that we have a repeatable process to look at the meaningful changes over time.

Our philosophy has been to try to develop a classification system and procedure that would be applicable over the very large areas in the three-state region. If we were to look systematically at remote sensing literature, we would probably find that a tremendous amount was published in the early days, where we only looked at the center portion of a digitized air photo, or at a small portion of a quarter of the Landsat scene at one time and made all these conclusions about how great it was. Scaling up really does give us a lot of fits with the image processing involved. We suddenly get lack of continuity of the atmospheric conditions. We get involved with the natural gradients along the terrain. So, we are trying to develop a procedure that can be utilized in that context. Much of this working document extends from the WISCLAND initiative referenced earlier. The WISCLAND basically started just slightly ahead of the GAP initiative in the state of Wisconsin. The WISCLAND stands for the Wisconsin Initiative for Statewide Cooperation on Landcover Analysis and Data. So WISCLAND is indeed a multi-agency activity involving State, Federal, and in the case of Wisconsin Power and Light, the private sector, with the concept of developing a systematic database. We would all share in its construction and cost and, therefore, have a better composite product than if individual entities were to approach this task individually. The WISCLAND is chaired by our State Cartographer's Office, and Bob Gurda, who was here earlier, chairs the steering committee. In addition to these signatories, the EPA and our Wisconsin Department of Transportation have contributed substantially to this.   In terms of the Wisconsin input to this, the protocol has involved a number of people in addition to myself. One of the things I would like to make clear is that we want to recognize the involvement of all the principles in this activity, many of whom are not in Wisconsin. Indeed, many have already been substantially involved by virtue of their previous and continuing input in the format of the classification system and the protocol.

What we really want as an objective is to systematically map a very large area. How large? Looking at a Landsat Multispectral Scanner [MSS] mosaic of the United States and the facts and figures, you can see that it's a pretty good chunk of territory, equivalent essentially to the area of France. It is 41% forested, in terms of the cover type. So, it is a very complex image processing situation when we are trying to reliably map vegetation over that area. Looking at Wisconsin to give you an idea of the contrast within just our state, in the context of the existing database we see the level one LULC [Land Use and Land Cover Digital Data] classification of the forested area, the lake-laden northern portion of the state, and the agricultural-urban area in the south. The WISCLAND set out developing an approach where we would take our current database, which lacks the currency we would like and lacks the spatial resolution in terms of the minimum mapping unit that we would like, and be able to produce a higher-quality product. At the same time, GAP is coming on board and, as a consequence, WISCLAND is envisioned as the Wisconsin complement of the tri-state Upper Midwest Gap Analysis Program effort.

One of the guiding philosophies of the protocol is the fundamental interrelationship between the data resident in a GIS and using the GIS to help extract information from the TM data and, in turn, depositing that information into the GIS. We will see this continual interplay at different levels of detail as we discuss the protocol downstream. We were trying to approach this to yield a procedure that would be accurate, detailed, would exploit the primary information of GIS, be explicit (that is repeatable), extendable across the region (and perhaps in great part elsewhere), efficient, adaptable to new technology and developments, evolutionary, and anticipate land cover changes. So, that is all the motherhood and apple pie that we were trying to meet.

The latter point is an important one—the anticipation of land cover change detection activities. To demonstrate the kinds of changes we are realizing, consider the Madison area in 1986 and 1994. That is an 8-year difference. There doesn't appear to be a whole lot of change there. It's the same mapping codes and the same image. If we visually concentrate on just the southwestern portion of that scene, we would begin to see that some new highways and some new subdivisions are under construction. If, on the other hand, we actually ratio those two data after proper registration, we can see that in the western portion of the Madison area we have tremendous change, with 1994 basically being bare soil or built-up land, as opposed to being vegetated in 1986. You can see, just in the city where we are currently residing, the tremendous amount of land cover change that has transpired. If we look at that change on the kind of scales used in the GAP analysis, we will be very surprised, particularly in areas such as this, at really how dramatically the land cover is changing. Estimates are at a minimum of 100 acres a day in the state every day of the year, so it is not a subtle change. In 5 years, we want to be able to repeat this analysis and look at those changes. This procedure shows one date in green and one in red. So if things don't change, it is yellow, but things that are most recently cleared are red and those things that were there before are different colors of dark green. So, here we can see a tremendous number of red areas, which represent the change between those two dates. That is the kind of thing that we want to do. We want to make sure that it is real change we are measuring, not offset of a pixel by two pixel units, because it's going to be a perceived change rather than a real one. We also want to have reasonably high quality data in our database, because if we repeat this experiment downstream, we likely only want to classify those areas that are now red, rather than reclassifying the entire area. So, there is a host of strategies we should be thinking about in terms of treatments in the change detection.

I think it is important to talk about the basic framework or the basic components for the protocol. The first is the thorny issue of establishing a classification system—a classification system that really is responsive to the needs of the Upper Midwest Gap Analysis Program, appropriate for the midwestern cover types, and reflective of what is doable using the given TM data source and available image processing techniques. We also wanted something that would be extendable because the TM data is only one small piece of the very large picture associated with continual acquisition of GAP and GAP-related data. So, we want to make sure that we can anticipate complementarity and crosswalks and so forth. Bob Goldmann will say more about the details of the classification system and the crosswalk as we go along. I think it is important to state here that we have two levels, a base level which we can reliably meet at 75% classification accuracy and interpret with well-timed TM data, and an extended level, which given the best of all worlds, we can produce with well-timed data. So, this scheme recognizes two levels of classification accuracy. It was based on the systematic analysis of 31 previous large-area land cover classifications in the Upper Midwest. So, there was a systematic study that Bob will talk about later.

The Protocol: Use of two-date imagery was available to look at phenology as a discriminator of cover type. We are looking at pre-classification segmentation of the image in order to improve classification accuracy. One cut is urban versus rural. Why? Bare streets and bare soil have very similar spectra. It's a silly thing to be confused by that. Simply lasso the urban area and classify it independently, and you have a better classification accuracy. The second cut is uplands versus wetlands. Cattails and corn look identical in the spectra. Eliminate that problem by delineating all the wetlands and classifying them as an independent group. Then bring the urban classification and the wetland classification in with the nonurban upland classification. Put them all together, and that is the final classification. So, we have stratification by upland versus wetland, urban versus rural, and in the nonurban upland areas, an additional stratification is by ecoregion (where we recognize the composite, physiographic variability, the soils, and the vegetation that is out there). We'll take it one more step and use Dennis Albert's map of the ecoregions of the three-state area in conjunction with other similar maps to then look at an image. Stratify the original image data, based on our visual interpretation aided by the aforementioned ecosystem maps. So, that is the general basis on which we do the stratification. In terms of the methodology used in different locations and different parts of the protocol, we use supervised, unsupervised, and a technique which is a hybrid, referred to as guided clustering, basically derived from the Minnesota experience in classification of forest vegetation types in northeastern Minnesota by Bauer et al. (1994). I am pleased to say we have some experts in guided clustering who have been actively involved in the Minnesota program in the audience.

Another approach we are suggesting is obtaining the training data in a more systematic way than is normally used to obtain accuracy assessment data. We set out to design a system that was geographically representative, that is to say systematic. It was, however, stratified in some sense geographically. It also had some element of randomness, with all the things that a statistician would look for, tempered by the reality of having to either field-verify these sites or to draw them out of the database. We are referring to that as the "stratified systematic non-aligned sample design" for obtaining polygons that we will use for both training and accuracy assessments. This design will improve upon the logistics of actually obtaining meaningful statistics.

We are also looking at continual development of multiple methods of measuring and communicating classification accuracy. Error matrices, users' or producers' errors, K-hats, and Z-scores are all important, but so are sample units and sample size. So might be the second most probable class we are calling jack pine. What's the probability that it's Red Pine? Maybe it is just a very close call. On the other hand, it may be that those two classes were quite dramatically separated in terms of their probability. We go through the classification, and we compute the probability density value for every pixel, but we don't do anything with it. We are suggesting it should be stored so that we have an inherent measure of the probability of it belonging to the class. We also figure out what the next most probable class is and perhaps the next one beyond that, so we have the marginal probabilities. In this way, when somebody cranks us through a GIS ecosystem model downstream, they will have a feeling for the relative probability of cover types in that model, one to the other. Wouldn't it be nice to be putting out our classification images with those relative probabilities as parallels? We can view the database in three dimensions—the most probable things stand out, or the reverse is true. Then we have a cartographic portrayal of that. With its elegance, there is no reason why we couldn't put it in some intensity-hue-saturation transformation, and show jack pine as green, but if it is really well classified, we show it as dark green in the image file. I'm suggesting there is a range of ways to look at it, and a lot of research to be done just in the development of alternative measures and in communicating the accuracy. One of the things we have suggested is that the polygons used for accuracy assessment have their vertices stored, so that with any cartographic product somebody picks up, they will know where the training and test polygons were derived. This gets back to what Mike Jennings was talking about, this concept of a national database. Preservation and spatial location of the data used for training and accuracy assessment is something envisioned in the polygon. Also envisioned is anticipation of land cover change and collection of radiometric normalization data as we are going through the current images rather than doing it 5 years downstream.

Let me illustrate some of these points. We mentioned the two-date analysis in a general sense. We are looking for image data that shows one date in the growing season and then, in the case of forest vegetation, the fall senescence or, alternatively, in the agricultural area, we might choose one date in the spring, looking at crop calendar differences. So, we have two dates and if we are restricting ourselves to the reflected bands, we've got six bands in each date. That gives us 12 sub-images. That is a lot of data. One of the first things we did in the pilot project, since we were last together, was to look systematically at a number of combinations of those bands, stacking them all together in a raw sense, doing independent classifications, choosing the best of each, computing principal components, putting all of them in one band, and taking separate dates and putting them together. There are a lot of permutations on how one can use two dates to do the classification. The one that seemed to perform the best, given the study areas we used, was independent computation of the first three principal components of each of the two dates and then putting those two together in one file to do the classification. Again, Bob or Heather has the details of how that was assessed, and I encourage you to ask questions or to talk to them. The first scene that we are working with here in Wisconsin is in the northwestern portion of the state. In this case, these are the two dates, October 1 and May 13. I think what is more important than showing the details of the principal components is to convey the concept that we have so much data volume that we have chosen principal components as a means to reduce the dimensionality of the data.

The stratification: Again, we stratify the data into spectrally consistent ecoregions. We are referring to them now as not ecoregions necessarily, but as modified ecoregions or photomorphic ecoregions where the analyst has decided that they have consistent spectral characteristics. Other forms of masking are the uplands versus wetland and urban versus rural. If we look at that stratification then, we are saying the classification of Wisconsin will be done in chunks. A stratum is a chunk, a photomorphic-modified ecoregion. All of the state will be classified using separate strata. Then, all the strata will be stitched together to give us the continuous classification. We envision that the same thing will be done in all three of the states. Within this stratum, however, we are now going through a secondary stratification delineating urban areas versus rural, then classifying those independently and delineating wetlands versus uplands and classifying the wetlands independently.

Now we are referring to the procedure for the nonurban uplands as guided clustering. We're typically using unsupervised classification for the urban and wetland areas. Tomorrow, Jana Stewart will be presenting the results of her research to document how we got to our current place with stratification. Along with the Department of Natural Resources [DNR], she looked at a number of different ways of performing the stratification; that is, Albert's ecoregions, Omernik's ecoregions, surficial deposits, etc. Obviously, depending upon whose map you use, you get a different view of the environment. Her study area uses three different maps relative to how we might stratify it, so there is a range of choices. Earlier, Dennis Albert's ecoregions were shown, and his map is the one that we are suggesting as the point of departure now for doing this initial stratification which is based on climate, physiology, soils, and vegetation. Using Albert's ecoregion maps and other information, we stratify the scene into spectrally consistent classification units. It is remarkable how consistent these are. Indeed, the Albert map really does a fairly decent job. It is, however, very clear that the Albert map has to be modified slightly to represent the true spectral variability the analyst is seeing in the scene. So we are modifying the boundaries to actually divide the scene into spectrally consistent units for classification.

By now, you are probably trying to keep track of what is an ecoregion, a modified ecoregion, a spectrally consistent classification unit, a stratum, and a photomorphic unit. I have introduced a lot of different terminology. To clarify that, let me just go back through it again. We use the units based on Albert's ecoregions, modify their boundaries to reflect the imagery, and classify each of the modified regions as a stratum. For each TM image, we have a number of spectrally consistent classifications. So you might find that we refer synonymously to these regions as modified ecoregions, spectrally consistent classification units, classification strata, photomorphic strata, and spectral physiographic regions. This means that when we take a TM scene, we will likely have different phenology, different atmosphere, different ecology, and different physiology, all of which are integrated into the delineation of the boundaries within which we are going to classify a chunk of the real estate as a unit. So there is indeed a great deal of feedback between Albert's original map, the image, and the analyst interpretation of the image that goes into the classification, all with the aim of getting good, within-unit classification accuracy, because there is between-unit spectral confusion. Let's remember why we are doing this. Janet will amplify on that fact tomorrow, but we are trying to get into areas that have the spectral confusion minimized among types. We do so by exploiting prior information about how the cover types are distributed across the area to be classified.

Stratification takes time, but the payoff is classification accuracy. Through experience, we've realized that if you want just general classes, stratification isn't worth it. The more generalized the classes you're seeking, the less important it is to do the stratification. If our whole objective in life is to separate land from water, we aren't going to be doing all of this, are we? So, think about that conceptually, how that cascades up. The farther we want to push in terms of classification detail, the more important this might be. On the other hand, we might over-stratify things to the point where we've got inadequate sampling sizes and so forth. So, there is a trade off in terms of actually affecting this in a protocol, but Janet will say more about that tomorrow.

Urban areas are separated from rural using TIGER [Topographically Integrated Geographic Encoding and Referencing]-line files, in addition to National Aerial Photography Program [NAPP] photography. Individual classification of the urban area is divided into two classes, high-intensity and low-intensity urban. In addition, there would be a third class within an urban area that might represent such a thing as vegetation that neither fits the high-intensity or low-intensity. That third class is not classified with the urban but is returned to the parent image and classified with the nonurban upland types.

Wetlands versus uplands: We at our facility have had a lot of experience trying to classify wetlands using satellite data and have had a lot of disappointment. Wetlands are spectrally complex, spatially complex, they are prolific, and they are important; but cattails look like corn. So the suggested procedure is to overlay the TM data with the National Wetlands Inventory, in our case, the Wisconsin Wetlands Inventory. Shown here is a TM image over which the polygons, coming from the Wisconsin Wetlands Inventory, are used as a mask to separate uplands from wetlands. The mask is applied and the wetland types are separated or classified independently, using unsupervised classification. To emphasize the importance of stratification, we take a portion of the second strata in our first image. On the left is the uplands database, on the right is wetlands. In that particular area, you can see a substantial proportion of both. The tendency is to throw a TM scene at the problem and say, "Give us the best classification you can." The important point is to do that stratification first and then independently classify those land cover types by groups.

I mentioned the statistically based sampling. Let's hit a little bit again on the philosophy that has led to the sampling procedure proposed in the protocol. One aspect is simultaneous collection of data for both training and accuracy assessment. The actual statistical prude would say, "Don't do that. Don't have the same crew even be involved. I don't even want them to meet one another before the classification and before they assess accuracy because then you have to eliminate all bias and everything else." The problem with this approach is the need to send two crews to the same chunk of territory, potentially. Another problem we found is that in supplying the land cover classification sheet to two people from the same facility going out in the field, each is going to interpret the classification system differently. So, with complete independence of training from accuracy, a potential problem is a difference in how personnel will interpret the classification system. The interpretation has nothing to do with the performance of the digital classifier, so that is a very important point in the concept of doing training and accuracy assessment simultaneously. The other thing is that we're midstream, but we are already dealing with data that were collected in 1992. We're 3 years out already. If you separate these functions, over the time of the collection real changes can happen on the ground. Then you've got that thorny issue:

"Gee, we were out there training and it was not a class x, now it is. . ." and so forth. So, by combining the functions, you do tend to minimize the differences that are due to time as well.

In addition to that, just consider the logistics of deploying field resources to collect data. Here in Wisconsin, we do have the benefit of WISCLAND to arrange for local cooperators to assist the State and Federal agencies in verification of the polygons we use for training and testing. I mentioned the statistical characteristics of the design that were sought—representation of the geography, but yet an element of randomness. To the extent possible, we wanted to avoid periodicity in the landscape, and we wanted to balance statistical purity with the application. We came up with what I would refer to as a nested design. That nested design is not just something in the abstract; it is tied to the 7.5-min quad series under the provision that a lot of our GIS data are organized in that way. Also, the NAPP photos are taken on a quad basis. It takes 10 NAPP photos to cover a quad, and they are systematically acquired at the bottom of the quad, in the middle, on the top, and filling in, in between. The actual photos are quarter-quarter-quad [QQQ]-centered. We also looked at a design so if somebody said, "We would like to use the same technique, but we're only interested in our county. We're interested in much higher resolution data, and we don't have 1:40,000 NAPP photography, but we have 1:5,840 black-and-white-infrared," it can still be used. In other words, the sample design is a geographically nested hierarchy over a range of scales.

What is in the protocol and being applied at this juncture is taking the state and, at the first level, breaking it up into 7.5-min quads, of which there are 1,154. Every quad will have one sample allocated to it. So, in terms of the systematic nature of the design, we are making sure that every quad in the state is sampled. Within the quad, however, we divide the quad up into 16 QQQs. We sample one of the 16 QQQs at random. In this case, this yellow area is the one to be sampled. Every QQQ has a quarter-quad-centered NAPP photo. Within that photograph, the data analyst delineates polygons using certain criteria. Those polygons should be homogeneous, be greater than 5 acres in extent, represent the spectral variability in the scene; and if other things are equal, they should be proximate to roads if they don't look substantially different from other counterparts in the scene, so they can be readily field-identified. If, while analyzing the TM image, it becomes clear to the analyst that a lot of the spectral variability in the surrounding area is not present in the sampled QQQ, polygons that capture those other classes are also delineated at that time. These polygons then become the fundamental units that we use for subsequent training and accuracy assessment. We are over-collecting the number we need for training so we reserve a portion for the polygons of accuracy assessment.

How do we determine which QQQ we're going to use at random? This shows the procedure. In red is the delineation of an individual 7.5-min quad. So, this red boundary here is a 7.5-min quad. We have six quads by four. Within those, we have the QQQs shown as the black boxes. Treating the upper-left-most quadrangle, we need a northing and an easting. That is, which quarter-quad are we going to take north and south and which one east and west. We roll a four-sided dice, so we get a one through a four at random. In this case, we roll a one. That means the northing for this entire column will be the first row. Then, we roll again, and we get a two. That means the easting will be two. What that does is ensure a randomness in the northing and easting and because of the replication of the four units per quad sheet, it also guarantees that our samples will be all spaced a certain amount. So, it is systematic, non-aligned sampling. So, that is how we determine which photos will be sampled.

What does that mean for getting the job done in an average-sized county? Our definition of an average-sized county is 248 x 248 pixels, which is close to the average county size in Wisconsin. There are four rows by six columns of 7.5-min quads. It is an area that is about 37 miles east-west by 34 north-south. If we sample 1/16 of that, it would be a 6.25% sample. We delineated polygons within that sample. We're sampling far fewer than 6.2% of the points in the image. We have 24 QQQs that would be sampled in that county. We are estimating roughly 30 polygons per image, that's 720 per nominal county. We are nominally saying we are going to divide the polygons in halves between training and accuracy, but that division will not be done at random. We want to make sure that we have adequate representation of the spectral variability in the scene, as expressed in those polygons used for training. Seventy two counties times 720 polygons; that is roughly 52,000 polygons. In certain portions of the state, that will be the case. All polygons will be field-verified by DNR staff or other State and Federal agency cooperators. This protocol also gets into the concept of the role of existing information in terms of the forestry inventory data—the other land cover data in Michigan, for example—so that the actual on-the-ground field visitation may not be utilized in all situations. We hope that the sampling methodology and the philosophy of the collection of the data will be consistent, though the data sources may vary among the three states.

This protocol is for the forested area, the nonagricultural area.

In the agricultural area, we must determine the situation in 1992 relative to crop times. If we indeed are after corn versus alfalfa versus other crops, we would sample anywhere between one and six Public Land Survey sections per QQQ. Those sections then would be chosen again to represent spectral variability and to have a number of fields in different agricultural types. If we get six or seven fields per section, we would have 156 in this particular area. We then put in a request to the Agricultural Stabilization and Conservation Service [ASCS] to identify the crops that were on those fields. ASCS then provides the information about crop times. Now, it would be really nice if they were able to point to a field and say, "Well, here's the whole field history for the last 4 years." The ASCS information is organized by owner, that is to say by tracts. So an intermediate step is required. The owner has to be identified for the field in that section, because an inquiry to ASCS has to be on the basis of the ownership, and the same owner could have multiple holdings. We want to make sure that we identify the holding in that particular section so it is properly done. It is not trivial to go back and recover this information. I guess, Don, you had to face this issue in the Lake Michigan ozone study. How many sections did you use? Did you stratify it by county?

Don Luman: "Twelve whole sections per county. That was all the states were willing to do, because of multiple ownership within the sections."

Tom Lillesand: So that is the agricultural sampling strategy.

I mentioned the hybrid guided clustering approach derived from the Minnesota experience. Again, guided clustering will be used in the classification of the rural uplands. We use unsupervised classification principally for the wetlands and urban areas, although, we might follow the urban areas with some form of supervised classification. The point is that in the rural uplands landscape, we would use guided clustering, which is a hybrid technique. It avoids some of the disadvantages of using either unsupervised or supervised independently. It works as indicated here. Let's assume that we are working with class jack pine. The analyst has delineated on the photos several polygons known to be jack pine. They've been field-delineated as jack pine, or we've gotten it from our database as jack pine. In this case, let's say we have these multiple polygons and we extract the pixels from the image for those areas that we know to be jack pine, nothing else. We then cluster the pixels into different spectral classes. So, we might have a range of clusters—let's say we have 10 of them, but all of those clusters are known to be jack pine. What we do then is evaluate those clusters. Some of the polygons, low and behold we didn't know it; right in the middle is a whole patch of marijuana. It looks different from jack pine. It will come out likely as a strange cluster because it has a different spectral characteristic than what we were attempting to evaluate. So part of our analysis might be that we throw that cluster out because it is really spurious spectral information. The other beauty is that the unsupervised approach looks at the inherent spectral variability in the category known to be jack pine and gives that to us automatically. Rather than an analyst having to find all of the symptoms of jack pine throughout the image, we say, "We know this is jack pine, tell us where the spectral classes are." So the unsupervised part of this approach will break up the original data into clusters and then evaluate and modify the signatures for class jack pine. This can be done in a number of different ways: pre-classification of a small portion of the scene, looking at histograms in one, two, or three-dimensions max, or looking at statistical measures such as Transformed Divergence or Jeffries-Matusita [JM] distance.

This process has been repeated for every one of the vegetation categories that we want to classify. We now go to Red Pine. We now go to whatever and repeat the program. We then get a set of individual signatures for all of the classes. We then do a holistic assessment of all of those signatures, refine them as a composite group, then apply supervised classification in the form of maximum-likelihood, using those cluster centers as the basis for statistics to classify the scene. Again, the advantage here is that we don't spend a lot of time labeling clusters. We make sure that the system will give the spectral variability that's inherent in the scene to the extent possible.

I mentioned we hope to envision multiple-accuracy assessment measures such as the error matrices, marginal probability, and density values from the maximum-likelihood algorithm, where it's applied. We want to look at these different ways of portraying accuracy. I really think a great deal of work must be done in parallel with a classification best describing, portraying, and measuring the classification accuracy. There are some really thorny issues. One of the things that I want to be able to do, and hope we can do, is to take that protocol and ask analyst 1 to classify stratum 1, then ask analyst 2 to classify stratum 1 using the same protocol and data. I would like to have a measure of between-analyst variability in terms of quality control.

Tomorrow, Dave Nagel will talk about the wetlands classification. We're using our inventory data as the basis for stratifying in that classification and for assessing accuracy. When we go to the ground and measure the accuracy resident in the Wisconsin's Wetland Inventory database, the correspondence is only 80%. So, what we treat as the gospel has some error in it—sometimes simply because we misclassified something. But on the other hand, there could be a change because of succession. Now, we want to look at this 10 years out and we want to compare, in a model, Time 1 with Time 2. We take the probability of being correct at Time 1, the probability of being correct at Time 2, times the probability of being correct with what we measured, or used as a standard correctness and pretty soon we'll say, "There's a lot of stuff out there. We don't know how accurate it is." We must avoid that. We have to recognize the differential accuracy among types in the classification system. We constantly have to be suspicious of the quality of what we are using as "groundtruth" and measure the accuracy of the accuracy measure, if that makes any sense. We have a great deal of work to do.

One of the things that we were also suggesting in the protocol was to try the sampling unit as a polygon, but we get to the center of the polygon and that gives us a point. At that point, look at the majority land cover type in a 3 x 3 array, and that's what the classifier has called that point. That is then the identity of the polygon. The minimum sample size that we hope people would use for expressing accuracy is 50. Now, where did that come from? That came from a fair amount of discussion with people who have been involved in CCAP, Russ Congalton and others who are saying that experience indicates 50 is the absolute minimum, and for very large areas you should certainly use more. This may mean that in order to have any sort of statistically valid expression of accuracy, we may have to aggregate strata or generalize classification systems. We don't want to have people using these techniques and saying, "Well, this is shown to be this accurate," when we only have 13 points to measure the accuracy because it just is not the same as having an adequate number of sample sizes. The point here is that the classification will be really accurate at the most general level. So, we have to be able to design our accuracy assessment data collection at the base level, the most detailed level. The polygon level will be verified in the field or some other way. If we had that database, we could then meaningfully use it to aggregate those data in a way that corresponds well statistically with what we are trying to assess, and that may be the accuracy in a county, in a stratum, in the state, or in the three-state area. All these levels of assessement have different implications in terms of the sample design. But if we assess correctly at the base level, we should be able to be flexible in data management downstream. Also, while people are bringing up all of the scenes all of the time, we might as well be looking for points that are, or are soon to be, spectrally invariant, such as large areas of deep, nonturbulent lakes or large paved surfaces that we assume aren't going to change. So, we collect those data now so that 5 to 10 years downstream when we want to repeat this program, we know already where the Time 1 sites were for radiometric normalization. At a minimum, we will then automatically capture Time 2 sites and do some sort of a regression between the two or some other more sophisticated analysis to be able to radiometrically normalize Time 1 with Time 2. Then, hopefully, we can do a subtraction or ratio or something to show only areas of change. We will then reclassify only areas of change to determine what went from one class to another one. So, it is important we think about this now, and these ideas are envisioned in our protocol.

We have come a long way from the microwave analogy, throw it in there, and push the button. Start with finding where the clouds are, delineate the cloud cover, then remove it from both images. Delineate urban areas using heads-up digitizing, using NAPP photography, and using the TIGER DLGs [Digital Line Graphs]. Compute the principal components for each scene separately, put them together and do an unsupervised principal-components clustering, and find high-intensity urban and low-intensity urban. The third class goes back to the parent image. We then delineate our spectrally consistent classification units, our photomorphic units, and our ecosystem regions as modified by the analyst. Within those, we delineate the wetlands using, in our case, the Wisconsin Wetland Inventory digital data. We compute independent principal components for the wetlands and use unsupervised classification to delineate the wetlands types. For those nonurban uplands within the spectrally consistent classification unit, we employ guided clustering. Then in those areas that have clouds, we extract the principal components for the cloud-free date. If one of the two dates is cloud-free, we apply it only to that date and use those statistics to classify those areas. Then combine the urban, wetland, upland, and cloud-covered classified datasets. Where we have clouds on both dates, we scurry for some other form of information to fill in those holes. That is not a simple process that we are seeing here, folks. It is formalized and, hopefully, defensible. Time will tell whether it is repeatable or not. We will certainly have some measures of that. But if any of you have the misperception that highly accurate and detailed image processing is pushing a button and buying the latest version of somebody's software to spit the map out at the other end, that is not the case. What we're trying to do is to say, "Here are the rules. Here are the protocols that we hope to employ to get us down to this 14th step, so that now we finally have something that is appearing for all intents and purposes to be rather systematic across at least the three-state area."

The 5 acres minimum mapping unit comes in here. Are you assessing accuracy on the pixel base?

Good question. Yes, and it's really kind of confusing, isn't it? When we assess accuracy, we will be starting with a database that is at the full TM resolution, that is to say the 30-m resolution. Envision a polygon that is out here and we know from the field that the polygon is indeed the jack pine. Now, it happens to be one of those things that we're going to say, "Now what did the classification call it?" We go in that polygon in the center, it gives us a pixel. At that one pixel it may be, for some reason, marijuana. Now, we don't want to say the classification called it marijuana. It may call that one pixel that, but eight that surround it are all jack pine. So, our sampling unit will be a single pixel, whose identity will be the majority of the pixels in a 3 x 3 array around.

Wisconsin Report

Bob Goldmann
Wisconsin Department of Natural Resources

My talk will parallel Dr. Lillesand's in that we are following the protocol that has been developed, so we will be reporting our progress with that particular protocol. We have done a number of things to date. The majority of them have been preliminary research dealing with the pilot projects, helping to test methodology, and defining the methodology that went into producing the draft protocol. We are in the process of moving into full-scale production in conjunction with this.

The hot topic which we did first was to develop the classification scheme. As part of this, we looked at 31 different classifications done in the Upper Midwest area. We noted the categories and attributes associated with each. We looked at putting those together in the minimum and extended classification. We also developed a crosswalk to the modified UNESCO classification, developed definitions and recording schemes, and we're looking at developing the color scheme as well. The classification assigned digital numbers to each class. In conjunction with that is a minimum classification with its corresponding crosswalk to the modified UNESCO as we interpreted it with the classification scheme.

Getting into the data, WISCLAND is dealing primarily with NAPP photos that we're acquiring for use during groundtruthing, as well as the TM imagery which is being used for the classification. We anticipate full delivery of NAPP photos by the end of this month. As for the TM imagery, to date, we have received one scene with dual dates. Today, we've also just received the adjacent scene for path/row 25/28.

We have developed and started on the groundtruthing activity for WISCLAND. For the most part, we've structured the groundtruth to have the observations occur along roads. We've identified the number of sample points and have chosen the methodology which Dr. Lillesand went over this morning. I would like to get into some of the specifics of the groundtruth form that the field personnel fill out as they are identifying the polygons to be used for training and accuracy assessment. We have also developed a series of definitions for the corresponding classification. A point to note here is that the classifications were delineated so that they would provide assistance to the groundtruth field personnel. They are primarily used as a guide for the field interpreter to help fill in the corresponding information on the groundtruth sheet. They are not intended to be a definition list as you would see in the UNESCO System.

What have we completed so far with groundtruthing? All of the groundtruthing to date has been done by the team members. No volunteer help has yet been enlisted because we did this just this fall, and we had only a month left of growing season. A tally sheet shows that so far we have identified 1,600 polygons off of 71 photos. Bear in mind that is 50,000 polygons off of a total of approximately 2,000 photos. We have a little way to go yet. You can also see by category the tallies and the number of polygons that we've hit so far. As part of this, we have separated out the agricultural areas. That groundtruthing is just getting under way now this winter. We will be contacting each county ASCS office to obtain a copy of their ASCS crop reports and a copy of their photomap.

As Dr. Lillesand said this morning, ASCS indexes crop reports by tract. To groundtruth an agricultural field, we have to identify the tract number for the fields that we're interested in, go to the corresponding crop report for that year, find the tract number, and then identify the fields and the cover type they contain. As you can imagine, this will be a pretty labor-intensive process. We have not started this effort yet, but will commence as soon as we get past this series of meetings.

Also as part of our project, we have an ongoing infrastructure development to provide us with the equipment and hardware we need. We looked at the TM preprocessing and determined the best way to collect large volumes of information in addressing the classification. We have also developed a series of ARC Macro Languages to streamline the data-entry process. The progress to date is a series of ARC/INFO (ESRI, Redlands, CA) ARC Macro Languages that allow us to first collect metadata on the files being created, and second to gather more detail about the metadata for the file itself. We have also developed an interface that lets us enter the corresponding training and accuracy polygons into the database. You will see examples of all of these in the demos tomorrow. As part of this entering process, we also have a series of metadata or data fields for each individual polygon. So each individual polygon will have a series of individual entities that can be used for future purposes outside of the training classification for this particular project.

As part of our project doing the wetland stratification, we're using Wisconsin's Wetland Inventory, which is analogous to the National Wetland Inventory, but specific to Wisconsin. We are funding the conversion of that data into a digital format, and are in the process of bringing that over into a raster format so that we can use that as a masking agent for the WISCLAND classification. We are anticipating full completion of the state by the end of calendar year 1995. Dave Nagel will give you more detail about the wetlands classification tomorrow, but we have assessed the geometric accuracy of overlaying those digital overlays on the imagery and have determined that it is indeed usable.

One note is the software with the revision changes of ERDAS Imagine (ERDAS, Atlanta, GA). Moving from 7.5 to Imagine has caused delays and necessitated a number of work-arounds within the project. As part of our update, Dan Fitzpatrick has asked us to discuss where we sit with the public ownership-management status layer. We are currently developing the ownership layer from 1:100 K USGS [U.S. Geological Survey] digital line graph files that will be augmented with our own DNR information.

Minnesota Report

Ken Boss
Minnesota Department of Natural Resources

So far, one of the challenges we have been facing in our GAP analysis work in all three states is data delivery. In Minnesota, we do not have a dual-day image pair to work with yet, so we haven't been able to execute any image processing to date. I'm going to be telling you about some of the preparations we have been making for the eventual execution of some vegetation mapping in Minnesota. I'll start out with a description of some of the ancillary data sources that we are gathering and will be working with. I'll talk briefly about a related project currently being executed at the Natural Resource Research Institute at the University of Minnesota in Duluth. I will have a few words to say about some adjustments to the WISCLAND image processing protocol that we might find necessary to implement in Minnesota. Then, I will conclude with a brief introduction of the area where we intend to conduct our pilot study.

One of the most important data sources we will be working with is the National Wetlands Inventory (NWI) which exists in Minnesota in a digital form for the entire state tiled to 7.5-min quads. For those of you not familiar with this program, the polygons and the dataset were developed from 1:50,000 NAPP photography and then transferred to base maps. The photography used for this project in Minnesota was gathered over 1974 to 1984 time frame. It is a bit out of date, but it is the best thing we have to work with. It is fully edge-matched across the state, although it is not map- joined, so we will have some processing overhead to deal with in terms of putting these tiles together in the units we will be classifying. We have full access to all the NWI attributes through the digital dataset, which will possibly allow us to pursue a guided clustering approach to wetlands classification in Minnesota, as opposed to the unsupervised classification approach being conducted in Wisconsin.

Probably the single most important source of groundtruth data that we have for forested areas in Minnesota is what we call our cooperative stand assessment [CSA]. This dataset was developed from a state DNR program and involves stand-mapping of all state-owned forest lands. The polygons in this dataset were originally delineated from 1:15,840 aerial photography, using the Public Land Survey section corners as ground control. Ninety-five percent of these data are currently available in digital ARC/INFO coverage and the remaining 5% will be available by the end of the calendar year. The oldest data in that database are 1976 vintage. The majority of the data were gathered in the mid-1980s and a good portion are even more current than that. All of the attributes in this dataset are ground-collected. They include such things as over- and under-story species, size and density, age class, height, volume by species, and a wealth of other data. Most of our forests are in the northeast part of the state. There are some significant oak forests down in the southeast as well. Dedicated state ownership of forest lands actually amounts to quite a bit of forest. Approximately 4.5 million acres are state-owned out of a total of 16 million forested acres across the state of Minnesota. We have data in the CSA dataset for all of that forest. So we've got ground-collected data on over 25% of Minnesota's forests available through the CSA dataset, which will undoubtedly prove to be an invaluable resource for GAP work in Minnesota.

A nice complement to the CSA data is found in what we call the County Biological Survey [CBS] dataset. This dataset is put together by the folks in the Natural Heritage Program at the DNR's Wildlife Division. This is a digital polygon coverage of natural community types in Minnesota often referred to as "remaining natural vegetation." In forested areas, this corresponds in many respects to CSA data. In fact, one of the things that the Natural Heritage people do before they go out to conduct their survey in an area is to crosswalk existing CSA data to their natural community types through a computer-based decision tree they have developed. The real strength of the CBS dataset, as far as GAP is concerned, is in non-forested areas where we don't have CSA data to fall back on. So, that will be our primary source of information for non-forested, non-agricultural cover-types. The polygons in this dataset were originally delineated on air photos and then transferred to 1:24,000 scale-base maps. All the attributes were first photointerpreted and subsequently ground-checked. The survey is completed or in progress for 30 of 87 Minnesota counties. If you would compare the completed CBS survey to the CSA coverage, they do actually form a nice complement in terms of spatial extent. We have covered most of the state with these two datasets, exclusive of agriculture areas. The CBS, like CSA, is not a wall-to-wall dataset. Actually, the attributes are collected only in areas considered to be natural vegetation communities. For example, Rice County in south-central Minnesota is primarily agricultural, so the CBS polygons are only the natural vegetation communities.

Another data source that we have at our disposal is the FIA data that comes out of the U.S. Department of Agriculture [USDA] Forest Service. This nationwide inventory is conducted in Minnesota at triple the intensity of a standard FIA and the state DNR works in close cooperation with the Feds to conduct this survey. It is executed approximately once every 12 to 15 years. The original plots are essentially selected from a dot-grid matrix overlaid on aerial photography. In Minnesota, that resulted in about 44,000 plots. Those plots are then photointerpreted, 12,000 of them were identified as forested areas, and those 12,000 plots are subsequently ground-checked in the inventory. I believe these plots were most recently visited during 1986 to 1990 time frame. Each plot is a 10-point cluster plot that covers approximately 1 acre. Fifty-nine different attributes are collected at each plot including forest types, stand age and size class, basal area per acre, and a number of subplot records for individual trees and shrubs. We have Universal Transverse Mercator [UTM] coordinates for these plots that are accurate to within 50 m or approximately two pixels on a TM scene. Obviously, 1-acre plots are of inadequate size for training purposes in image processing, but we do feel that they could make a good independent data source for testing the accuracy assessment. Furthermore, in Minnesota we have the Ecological Classification System, which is actually just an extension of Denny Albert's work in Minnesota. This exists as digital ARC/INFO coverages at the province, section, and subsection levels in Minnesota and will form the basis of preprocessing stratification for the scenes. Since they are developed on the regional basis, their use will also facilitate edge-matching at state borders. The next step of ecosystem delineation in this hierarchy is the land type association, which is currently under way in Minnesota and is expected to be completed in digital form by the end of this calendar year. Although we don't plan to further stratify the scenes according to those land type associations, we will use them as a guide for the placement of training polygons so that we are certain to capture the variability in spectral response across cover types throughout a subsection.

We will be making use of several additional datasets. Probably most important for agricultural classification are the ASCS data that reside at the county level. These data will form the basis of both our training and accuracy assessment datasets in the image processing procedures, much along the same lines as Wisconsin's procedures. Because they do reside at the county level and are archived according to land ownership, a fair amount of time and effort will be required to actually gather these data and get them into a usable form. We haven't begun work on that yet. Some of the other datasets available to us are a couple of land use and land cover mapping efforts. These are both in digital form, the

Geographic Information Retrieval and Analysis System dataset from the 1970s and another Minnesota coverage that was executed in the mid-1980s and is about half complete right now. We will probably use these as kind of a general reference to guide our work and also to help identify urban areas that need to be lassoed for the urban-rural stratification.

Last, being the Resource Assessment Unit for the state DNR Forestry Division, we are the main archive site for DNR aerial photography, so we have got aerial photography coming out of our ears. We have NAPP photographs acquired in 1991 and 1992 as spring, leaf-off color-infrared photography for the entire state. We have National High Altitude Photography from the 1981 to 1984 time period at a slightly smaller scale. We have a number of our own DNR forestry projects, which are flown with a 9 x 9 format camera at a 1:15,840 scale. These were traditionally flown as black-and-white- infrared summer photo projects. In the last several years, we have converted to fall color-infrared as a means of highlighting distinctions between various cover types through the differential senescence. We also have a good source of supplementary photography that covers about a million acres per year. This is all small format, 35-mm photography both color and color-infrared flown on an as-requested basis. We currently have in-house large format photography that covers the entire forested portion of this state from about the year 1983 through the present. We have found fall color-infrared (as opposed to the black-and-white-summer-infrared) extremely useful in distinguishing forest-cover types. Northern hardwood stands are readily distinguished from aspen cover types in photography like this. Tamaracks that are coloring can be easily distinguished from surrounding black spruce cover types. Bill Befort, our Remote Sensing Coordinator, could tell you about the trials and tribulations of trying to acquire photography like this because of the extremely short window of opportunity for acquiring photos, and having to depend on the weather to cooperate.

I would like to talk a little bit about a study conducted at the Natural Resource Research Institute (NRRI), University of Minnesota, Duluth, using multi-date Landsat imagery to show differential senescence in forest-cover types. Many of you here are probably familiar with this project. I will just briefly touch on their approach. Over the last few years, NRRI has been doing large-area forest cover classifications using Landsat imagery of multiple dates in a hierarchical classification approach. The study I am discussing was published by Mladenoff and Wolter in the September 1995 issue of *Photogrammetric Engineering and Remote Sensing* (Wolter et al. 1995). The study was commissioned by the USDA Forest Service and was conducted in the area of the Chequamegon National Forest in northwestern Wisconsin. The classification results are being used in a model to predict the presence or absence of various songbird species and, in this respect, is similar in its aims to GAP. In essence, this approach employs one thematic map or scene from early summer 1987 as a baseline and four meticulously chosen MSS scenes of various dates to capture certain phenological events. As in flying the aerial photo missions, it can be very difficult to find that window of opportunity to grab those phenological events, especially considering the 16-day Landsat repeat cycle, so it was necessary in this study to go as much as 6 years on either side of the baseline TM date in order to get suitable imagery to capture these phenological events. The hierarchical classification approach that the Forest Service used allowed them to classify 22 forest cover types, with a reported accuracy of approximately 80%. Basically, they used the June 1987 image to threshold the image between forested and non-forested cover-types and then again to further threshold the forested area to hardwood, conifer, and mixed stands. Then on October 8, 1980, they used the MSS image to classify oaks from the hardwood portion of the image. This was a time of year when oaks were still holding their leaves, while the other hardwoods had already dropped theirs. They then ran a maximum-likelihood classification on the oak portion of the image to further delineate between northern red oak and northern pin oak. Next, they compared a September 13, 1985, image to their baseline 1987 TM image to classify black ash through an image-differencing procedure. Black ash in September has already dropped its leaves, whereas the other hardwoods are still holding theirs, allowing the Forest Service to distinguish that cover type. Next they used an early May 1992 image to capture aspen leaf flush at a time when the other hardwood species were still dormant. Finally, a winter scene from February 1988 distinguished tamarack from other surrounding conifers, that being a time of year when tamarack has dropped it needles. Some state ecologists have been using the results of this work in Minnesota as a guide for where to concentrate their efforts in the County Biological Survey and have found it to be quite useful data.

I would like to move on to the next part of my presentation, which relates to a few adjustments that we would like to see made to the WISCLAND image-processing protocol for use in Minnesota. First of all, if feasible, we would like to consider the possibility of incorporating work previously conducted in the state. The first of these adjustments would possibly incorporate the project I just described. Another possibility is considering the recently completed University of Minnesota Remote Sensing Lab classification of the seven-county Twin Cities area. Incorporation of such work into GAP would necessarily depend on at least two things; first and foremost, the compatibility of the classes used in the projects, and second, an accuracy assessment that would facilitate comparison of those earlier classification efforts done using the WISCLAND approach. This would ensure that the classifications are executed to a level acceptable for GAP. Such an assessment might well be executed in that original study area in the Chequamegon.

Another area of adjustment which I have been alluding to in the description of the reference datasets we will be using is the idea of using existing in-house datasets for reference and image training. Minnesota lacks the resources both financially and in terms of cooperators to conduct the sort of stratified systematic nonaligned sampling scheme that Wisconsin has constructed. Fortunately, the state already has a wealth of reasonably current ground-collected data that we can fall back on. A majority of it is already in a digital form. So, we would propose selecting training sites from existing field data collections based on criteria which would include such things as the currency of data, the size of polygons, the homogeneity within polygons, etc. At least in training data selection we are really after a representative sample of ground-collected data and not necessarily a random sample. Accuracy assessment sites might be a little trickier because they do need to be randomly sampled in order to generate valid statistics, but it may be possible to develop some algorithms that would select polygons from our existing datasets in a random fashion. The FIA dataset could form an independent set for accuracy assessment of forested areas.

Finally, are a few issues regarding target classes. These adjustments are based largely on input from the ecologists in the DNR Natural Heritage Program. First and foremost, we have not identified a constituency for a detailed 1992 agricultural information dataset in Minnesota and are not sure that it would be a wise expenditure of our time and dollars to try to classify all the various crops in the fields. Unless somebody comes forward and suggests that they could really make good use of such a dataset, we would propose to classify agricultural areas at a more general level. Some of the forest types that have been identified as classes for WISCLAND do not commonly occur in Minnesota. Other forest types not identified for WISCLAND are important in Minnesota. Finally, we would hope to extend wetland classes with the aid of the digital national wetlands data that we have in-house through a guided clustering procedure.

To close, I would like to talk a little bit about the area we intend to pilot. We overlaid the polygons representing the Landsat scene borders on the ECS [Ecological Classification System] ecoregion subsection coverage, and we propose to pilot the Chippewa Plains subsection. We have a couple of good reasons for piloting this area. One is that the County Biological Survey has recently completed their survey in that area, so we would have access to their data in conducting our pilot in that area. Second, if you look at how this subsection falls in relation to the province, you see that it is largely internal to the province in which it lays; so we wouldn't have to deal with the thorny issues of the transition zone across provinces in the pilot, saving those issues for a time when our image processing protocols and procedures are better established. That is basically all I have right now. If there are any questions, I would be happy to entertain them.

I heard of one process which Minnesota is looking at to eliminate clouds from their scene and that they are using the thermal band.

Yes. In other projects that we have conducted at our unit, we have always made use of the thermal band to mask cloud and cloud-shadowed areas. We don't generally use them in our classifications, but they have proven to be very useful for masking out clouds. Whether or not it would be worth the effort to try to get the MRLC [Multi Resolution Land Characteristics] Consortium to go back and ferret out band 6 from all of the images that they are supposed to be shipping to us, I can't answer.

Michigan Report

Mike Donovan
Michigan Department of Natural Resources

and

Pete Joria
Environmental Management Technical Center

Pete Joria: As Frank mentioned, the Michigan GAP project is pretty much in its infancy. I have been on board since mid-November and actually just met my co-worker Mike Donovan today. We do have some information for you today. We tried a rather preliminary approach because of the different data available for Michigan. I'll address the specific attempt and then Mike can give you a broader overview of GAP in Michigan later.

The original proposal for GAP in Michigan was quite different from those procedures being implemented in Minnesota and Wisconsin. Layers depicting land use in Michigan already exist within a statewide database called MIRIS, which stands for the Michigan Resource Information System. However, the information in the layer is dated; it dates back to photography acquired in 1978. One initial proposal was to update MIRIS with a manual interpretation of Landsat TM data as a back drop. Manual of course means visual, so one drawback is that only three bands of the TM data can be used simultaneously.

Some background on MIRIS; it was created under the Michigan Resource Inventory Act of 1979. The purpose of the Act was to create a statewide computerized database of information pertinent to land utilization, management, and resource protection activities. Data in the MIRIS system were compiled by numerous local, State, and Federal government agencies, not just a single agency. The accuracy of the information is relative to the standards and specifications developed by each particular group, as well as the scale at which the data were compiled. Therefore, the quality of the information may vary geographically. MIRIS data are available as paper maps or digital copy. Digital files are created in an Intergraph Microstation vector format, but the Michigan DNR provided vector files in an ARC/INFO format for this study. The minimum mapping unit is 2½ acres, half the size of the proposed minimum mapping unit for GAP training and assessment polygons. MIRIS land cover categories are identified by a code of one to five digits. The most general categories are shown here, except for forested areas. Most urban and forest polygons are identified to a three-digit level similar to Anderson level 3. For example, single-family residential for urban, or, in forested areas, northern hardwood. The fourth digit labels polygons to the predominant species. The fifth identifies size class and density. In the study area I was looking at, I never saw a polygon labeled with more than four digits, or beyond the species level.

The study area was Midland County, which is in the east-central portion of Michigan's Lower Peninsula. It includes a mix of forest, agriculture, and urban land uses within an area of about 1,300 km^2. The TM scene date represents very early spring. Forest leaf-out had just begun. The grasslands and some crops were already highly reflective in the near-IR [infrared]. Slide five shows the study area as a three-band TM image, although here it appears more as a single band image, within the boundaries of Dennis Albert's ecoregions. Midland County falls entirely within the Saginaw Bay Lake Plain. Subsection 313 is characterized by mesic to wet-mesic forests, swamp forest, wet and wet-mesic prairie, and emergent marshes. Lake effects from both Lake Michigan and Saginaw Bay may have delayed spring leaf flush here. Saginaw Bay is just to the northeast of the county. The closer look at the TM image of the study area is in bands 4, 3, and 2. The Au Sable State Forest appears as the dark green area in the central and north-central portions of the county. The predominant forest cover type is lowland hardwood, which includes ash, elm, silver maple, and cottonwood. Originally, when I saw this image, I thought that the generally dark appearance was because it is a lowland forest type, with high absorption by water. But taking a closer look, it is evident that the dark appearance is also because of the early spring date and pre-leaf-out conditions. The bright red areas along the Pine River, which flows from the southwest corner of the image to the city of Midland on the eastern edge, are identified as crop land, although areas near Midland with similar appearance are identified as outdoor recreation. In the urban category, that includes city parks. The MIRIS coding scheme often identifies land use, not land cover. It is sometimes difficult to translate to a scheme based on interpretation or classification of TM data. While some areas identified as agriculture appear to have growing crops by this date, most are still bare soil and appear as the white areas, mainly along the southern edge of the scene.

This slide shows the MIRIS land cover layer for Midland County in the form of an ARC/INFO coverage. The level of detail of the MIRIS data is obvious, with over 7,500 polygons identified in Midland County. Some artificial boundaries are also obvious in the image, in particular, boundaries between USGS 7.5-min quads. Many of these boundaries remained when the MIRIS data for individual quads were joined together in ARC/INFO. Typically, the boundaries are not because of different interpretations of adjacent polygons, but are an artifact of the joining process with some overlap and sliver polygons occurring between quads. Occasionally, however, the boundary reflects more specific labeling by an interpreter, for example, in an area of lowland hardwoods where the predominant aspen is identified, next to a polygon with a more general lowland hardwoods category.

Three major difficulties were encountered with the pilot project. First, as I have noted, May 10 is simply too early in the spring for an update based on manual interpretation of a single date scene. The scene might be more appropriate in a multi-date digital classification with a second scene depicting peak growing season conditions. Second, it is also obvious that TM imagery by itself is not sufficient to visually identify new land cover types, though drastic conversions, such as vegetation to urban types, may be detected. The 1:40,000-scale NAPP photos may be the only statewide photos available. Third, and most important from my perspective, is the amount of time required to visually update individual polygons. In Midland alone, 7,500 polygons occur. Midland County represents a little more than 1% of the area of the Lower Peninsula, so a conservative estimate of the number of polygons within the MIRIS data system occurring in the Lower Peninsula would be about a half million.

Current status of the Michigan project is, first of all, we need a second scene depicting peak growing conditions. A temporal window in the Lower Peninsula is particularly short because of the gypsy moth defoliation problem there. I have been talking with Frank Sapio of the Michigan DNR and acquiring data depicting the extent of defoliation since 1990, and some years it has approached 700,000 acres. We also would like to refine the GAP-MIRIS crosswalk to make sure we can incorporate MIRIS information into whatever approach we do end up with for Michigan. Our main interest in the GAP meeting today and tomorrow is to meet with our fellow workers and try to adapt the GAP protocol to Michigan and still retain the Michigan MIRIS information as part of the protocol. It has a lot of possibilities, particularly in the guided clustering done in the upland areas. A lot of information there can be put to use. That is all I have right now, I would be happy to entertain questions.

Is the MIRIS database for the land cover of the entire state including the peninsula?

Yes, I believe that is correct. We just have Midland County in an ARC/INFO format, though. It would have to be converted for our uses.

Isn't there a fairly recent land cover classification for the Upper Peninsula?

Yes, for the Upper Peninsula there is.

Any other questions?

Then, I would like to present Mike Donovan from the Michigan DNR.

Mike Donovan: Pete did a good job of covering the activity that the EMTC is involved in. Participating in that activity was one of the things listed on our agreement with GAP for the year. We have three other things on that agreement, and I will cover those. The first one is to extend the land cover classification effort in Michigan's Upper Peninsula [UP]. We have just completed a UP-wide land cover inventory using 1991 TM imagery. We did this project for the specific purpose of identifying lowland conifer types associated with white tail deer wintering areas. We did this processing with a really specific goal in mind and ended up with 30 categories of land cover. About half of those are in lowland conifer types, so we spent a lot of effort in identifying lowland conifers. We spent less effort in identifying the other upland types. We have categories similar to the WISCLAND category for those upland types but more on the order of Anderson level 3 for the non-conifer types. In the conifers, we go down to the species level and make an attempt at canopy closure with two broad categories, greater than 70% and less than 70%. I couldn't fit it on the plane, but we have produced a 52- by 36-inch seven-image mosaic of the UP. Part of our agreement with GAP is to take that effort and try to expand it. We will go into non-conifer types and go for greater detail, trying to match the GAP protocol. This meeting will be a good kickoff for that because we now have a protocol document we can work with. There is a lot of overlap between that protocol and our work in the UP, and there are areas of divergence too. So, we need to negotiate the differences and decide how much we can do with the budget allocated for GAP.

The other thing we had identified as a first-year effort was to develop a public ownership layer. In that effort we have, as part of MIRIS, a state land ownership layer that will be complete this year. Right now they are wrapping up the UP for all our state ownerships. We also have National Forest lands, National Park Service lands, and all those other public ownerships. From initial assessments and discussions with people at numerous meetings, I've concluded that most of that effort will actually be one of data conversion, since much of that information is already in digital format. We just need to convert it and get it into the same coordinate systems and the same software package. So, that is going to be one of those large data conversion efforts. We also have The Nature Conservancy and other conservancy group lands, and some of that is already mapped within our Heritage program.

The third effort part of our agreement with GAP was an analysis of the state survey intensity for threatened and endangered species. We thought it was important that we know how good our Heritage database is, so we wanted to go back and look at all the surveys conducted and figure out where it says we have gaps in our survey. We are still working on how we are going to approach that effort, what makes sense in terms of looking at those surveys, and how to portray that information. So, we are early on in that phase.

So, those are the three things that we had agreed to with this year's budget and our progress on them. There is another very significant effort that will be wrapped up in June and that is that our Heritage folks are mapping presettlement vegetation for the state through interpretation of the original Land Office Survey notes. That effort is about 60% to 70% done now. Quite a few counties are completely done and available as part of our MIRIS program, others are not. So, that is basically where we stand on that. Any questions?

Was there much intersection with that eastern UP effort between Forest Service, DNR, and private land holders?

Actually, there is now. That dataset is available and our Forest Management Division people are involved in that effort. Another significant note is that the field office of The Nature Conservancy threw in a little extra money for the eastern UP and the Northern Lake Huron Bioreserve Project, so in the eastern UP we do have greater detail of the classification system, especially in the wetland types.

Could you refresh me again, was the UP classification unsupervised or supervised, single date or multidate?

It is a classification based on a single date of imagery, and it combines unsupervised and supervised approaches. I won't call it guided clustering, but it was similar. Maybe a better term is hyper-unsupervised. In other words, the contractor did an initial classification, masked out those areas we were not interested in, the upland types went back into lowland conifers, and then he just basically did an unsupervised classification. So, it is unsupervised, but he went into a tremendous amount of detail in masking out other types as he went along. When he felt he hit the specified accuracy or within contract specs, he masked out that vegetation class and went on to the next class. There is a document that details that whole process. One of the other things is that the window of opportunity in Michigan is very narrow for good imagery and in that project, the scenes for the western UP were of poor quality, and we would really like to acquire some better imagery. TM scenes for the west end of the UP had a high haze layer in the atmosphere. Good cloud-free, late growing season imagery for the entire UP just doesn't exist.

Illinois Report

Don Luman
Illinois Natural History Survey

After hearing about the three-state consortium here with Minnesota, Michigan, and Wisconsin, this is going to be a bit different in that Illinois' landscape is quite different. Eighty-four percent of our 56,000+ square miles is agricultural. Tom Lillesand mentioned in his overview of his protocol that the three states had approximately 41% forest. We have only about 9.5%, so we are predominantly an agricultural state. With the exception of Iowa, we have the largest amount of this type of landscape within the Upper Midwest, which is probably going to change the discussion quite a bit in terms of how we are approaching GAP in Illinois.

Before I start with that, I would like to give a few introductory comments as to who we are. My name is Don Luman and I am with the Illinois Natural History Survey, which is one of several divisions within the newly renamed Department of Natural Resources. We merged with a few other state agencies to, as the Governor said, reduce the footprint of government in Illinois. So, we are now the Department of Natural Resources. The Illinois Natural History Survey is one of the oldest and largest surveys in North America. Our chief, Dr. Loran Nevling, was a member of the formation committee for the National Biological Service and serves on many subsequent committees. The Illinois Natural History Survey co-signed a Memorandum of Understanding with the National GAP Analysis Program in January 1994. Just this past fall, we submitted a formal proposal for funding, and based on what Mike told us this morning, we are one of those that is anticipating funding for our first year.

In Illinois, we find ourselves in quite a different situation. Referring to Wisconsin, Tom Lillesand mentioned this morning that they were seeing about 100 acres per day of changing landscape. It is interesting because the National Resource Inventory just produced by the USDA has estimated that Illinois has been losing approximately one township of agricultural land per year in the last 10 years. This seems to be sustained and if all aspects are correct, it looks like it is increasing. We are experiencing something on the same scale as Wisconsin in that we have our most important landscape, in some regards economically, losing ground literally each year. Yet, the state does not know where this land is that we are losing. We can surmise it is mostly around urbanized areas but we actually cannot point to where it is. For these reasons and others, the state decided to join the MRLC Consortium last year. We have done that and hopefully we will be involved in a formal GAP analysis this year. Informally, we have been working for more than a year, about 16 to 18 months, preparing for this eventuality of having a funded project.

I would like to break the discussion up into three parts to give you a status of where we are and to tell you about some of the things we are involved in. Number one, which was probably the most important and actually one of the more difficult periods, was to acquire or to complete the TM data acquisition for Illinois. I know this has been an issue that has been mentioned by the three states heard previously. It was very gratifying to have both the NBS office in Idaho and EROS [Earth Resources Observation System] Data Center [EDC] cooperate to our contributions. Some were a little more serious than others in that we felt we knew the best dates for our landscape. Through a lot of compromise and sharing of information, the MRLC database for Illinois has been completed. Not only completed but we have, except for one scene, a full multi-temporal statewide database in-house. We have approximately 24 scenes already present. Does that mean we got ahead of everybody else? No, we just decided to get some objectives met by the end of this past year. We requested that MRLC forward the system-corrected

TM data to us, and we mutually agreed that we would not require them to do the geometric registration for us. We took that task on ourselves rather than waiting several months because EDC was just getting their staff up to speed. By doing it this way, we actually received almost the entire database in one shot in mid-April of last year. It was system- corrected, therefore, the full geometric correction had not been done to the data. It was interesting that about 2 to 2½ years prior to this, the state of Illinois had acquired a statewide purchase of TM data from the EOSAT. Those data were all precision processed, map oriented. So, what we had been doing is co-registering the new MRLC scenes to the statewide purchase. The fit was remarkably excellent and most of it refits well within a pixel or two. I should mention that we had one noticeable hole in the multi-temporal dataset and just last month the USDA Forest Service from Minneapolis responded to a desperate request on our part and purchased the remaining scenes for us. Finally, the multitemporal coverage was completed.

Second, during this time of acquiring the data, we developed an internal classification system. It is not the GAP classification system, but it is a system which is based on our landscape and tailored for our specific kinds of landscapes within Illinois. It is about a 16-class classification system that emphasizes predominant landscapes present within Illinois. For example, in terms of agriculture, we are doing what other states do in enlisting the ASCS crop compliance information from our state. It is useful that the Federal office for the state is just a few blocks from our agency. We have a statewide sample. From a sample of 102 counties, we were allowed a sample of 37 counties. Within those 37 counties, we collected the 35-mm compliance data along with what we refer to in our state as the cut sheets and the 35-mm photography 12 quarter sections per count for 37 counties. One thing I learned from experience and I think was valuable was that we had to do this prior to the Landsat data arriving. We did not know exactly which years we would be getting, so we asked the state Federal office to supply us with the 12 quarter sections for 4 crop years. The big problem is that the organization of the data within the county office ranges from fair to very poor in our state. I am told by our Natural Resource Conservation Service (NRCS, formerly SCS) office that because of the predominant amount of agricultural land in our state compared to some surrounding states, we actually do not have enough staff in the county offices to handle the burden of work generated by these new sign-up programs. As a result, we tested one county in northwestern Illinois, where we knew the district conservationist very well, and asked him to take 12 sections of data and develop it for multiple years, 1 year, etc. In terms of agricultural discrimination, we are actually at the crop-type level. We could actually separate at the various types of crops, but we have been advised by the NRCS office that it really is not terribly useful within the state. We are quite easily separating row crop areas from pasture land areas and small grains quite consistently.

We have found that in our state in the last 10 years the color of the state of Illinois has changed. Ten years ago, back in the early 1980s, the farm bill was still new. Minimum tillage was still not something that was big in our state. Clean tilled crops were the big thing, and in the spring the state looked rather dark, almost black. Now, when you look at Illinois in the spring with the MRLC data, it is bright cyan over almost the whole state. This is because of a level of crop residue in the row crop areas. The majority of the state of Illinois now practices no- or minimum- tillage techniques. What we have been doing and refining, through a year-long study with a graduate student, is to see if Landsat TM data could be used to measure discreet levels of crop residue in row crop areas. The answer to that is yes, and you can do it consistently for a small area. You cannot do it consistently for an entire state. Therefore, rather than the three levels that the NRCS wishes to see, we simply divided row crop into two types, row crops that are retaining a high level of crop residue, that is above 67%, and row crops that are exhibiting or retaining low levels of residue or no residue. We find that models the physical landscape, because the areas of highly or susceptible erosion lands are now practicing high residue. Areas that are depositional in nature do not have to practice residue management, therefore, they show up as a land form on the landscape. It is kind of an interesting and potentially valuable tool. Our row crops are being separated by a level of crop residue. Why are we doing that? Because we have the USDA/NRCS as a cooperating agency with GAP, so we need to share by having some things that they would like.

The classification scheme that we have is about 16 classes. It seems that we are doing pretty well with that. We are just finishing with our third scene area and moving on to the next area. We did validate one thing the three states have found out already; we tried to experiment using a single date to see if we could get by with it. The answer was absolutely not. The GAP initiative has to be multi-temporal. That has clearly been proven. So, we have embarked on a full multi-temporal classification. We are actually in a production mode right now processing a scene, two dates about every 6 weeks or so. Now, that doesn't mean it is finished but it is pretty much finished at the time. We are also separating urban and rural areas. We have found that in our state we have an interesting mix. We have a predominance of agricultural land, but we also have one of the largest metro areas in the United States—Chicago. You cannot process those two landscapes together; they have to be separated. We have found within the TIGER data is an interesting buried feature that is not used a great deal. Within blocked areas, within urbanized regions, and within what are called places and urbanized areas, the Census Bureau, for the first time since post-1990, designated each census block as either urban or rural. There is actually a flag they throw. If you create a mask with those characteristics, rather than just using the incorporated boundary or what people call the municipal boundary, you pick up some important extra blocks that are actually urbanizing. There is a down side too, because you may also exclude some areas, but we found it to be a balance. If you want to try to get a little more of the urbanized area to do the urban-rural masking, this is something we found to be really quite useful.

Another difference in our state is that we are not processing the large cities, the metro areas above 50,000, with the small towns and villages. The larger cities are processed separately. We found we can do a full classification of urban areas for the larger cities, but the small towns and villages do not seem to have the same structure. Therefore, if we treat these separately we find (much like what Tom was talking about where he is doing his principal component analysis and taking two categories of urban) we are actually doing that for the small towns and villages. We then simply imbed those in the rural classification, and it comes out really well. We are quite happy with the urban classification.

In terms of ancillary data, it is not terribly different from what you have already heard today. As I have mentioned, we are relying heavily upon Soil Conservation Service [SCS] data to help guide us on our agricultural classification. We are not using those data in a supervised approach. We are doing it a little differently in that all of the fields delineated here on the cut sheets are actually not digitized, but points are selected looking at the raw imagery heads up. Points are sampled within each of those known fields and stored in an ARC point file. At the end of the clustering and classification, we bring half of those points back in, overlay them all in the final classification, and use those to tell us what focal spectral classes are in certain crop types. This works very well. We find that we can identify about 150 spectral classes through this approach. They then can be labeled as specific crop types, and that works out quite well. So, we are actually using it in the post-classification labeling. Earlier last year, if we had one date already for the MRLC, we were then constrained to use the same year for selection for the second date, which really restricts the choices for the TM data. When using this method, we had the SCS data and could actually follow a specific field from spring to fall and could see the change in its senescence over the crop period for one crop year. The SCS has told us that in classifying agricultural lands we should not go more than one entire crop year. We may not be able to go spring to fall in 1 year, so going fall 1 year to spring the next year is not recommended, but it is tolerable. We have found this works out pretty well. So, if you have the same scene, if you're going for multi-temporal data, you have spring to fall, and you happen to be on the same property, the SCS data can really monitor that phenological change in the fields. We have about a 20% sample of the color-infrared NAPP, much like our surrounding states. No agency within our state has ever purchased the entire statewide database of color-infrared imagery. It still resides out at Salt Lake City and EDC. We did purchase a 20% physical sample. We have 100% coverage for the black-and-white map at the University of Illinois Map and Geography Library, and that is excellent, but it is inferior to the color-infrared for doing landscape examination. We have run an experiment this past year using Digital Ortho Quads [DOQ] for one county to literally check the classification heads up for a Landsat classification. In other words, we have the DOQ on the screen. These are 1-m resolution imagery from the quarter quads, and when we point to a building on the NAPP, it will point to where it is in the classification scale. That is really slick, but we currently only have five counties within the state that are converted to the DOQs.

Finally, I will allude back to Tom's talk. He was talking about how this Gap Analysis Program should be adaptive to new technology and evolutionary. Well, we took him at his word. We happen to have a very wonderful situation at the Illinois Natural History Survey. We were located on the campus of the University of Illinois in Champaigne, which is also the home of the National Center for GIS applications and another agency closely allied to it, the Beclan Institute for Advanced Sciences. Over the past year, the Department of Electrical Computing and Engineering and the Natural History Survey have co-shared an experiment to try a new technique that is in the remote sensing literature back into the 1970s but is not applied to regional classifications. The term that is often used for this is vector-field segmentation. It is image classification using pixel classifiers that typically do not use contextual information. So this spatial domain, which is in the image and visible to the human eye, is geometrically homogeneous areas that are largely ignored by remote sensing algorithms. Yet, the research has shown that inclusion of spatial structure can improve discrimination. One of the approaches that has been used in the 1970s, as I said, is referred to as vector-field segmentation, and this approach uses what are referred to as massively parallel deterministic relaxation algorithms. In essence, it approaches the partitions of a multi-temporal, multi-scaler image dataset into regions that correspond to homogeneous objects on the landscape. Through a process that we refer to as photographic generalization, the original image data is actually preprocessed. The result is that the common problems usually found with land cover maps at the end of the process, such as noise, salt, and pepper, etc., are no longer present in these images and in the classified maps that result from them. So, as a result of a cooperative study in which we did a fair amount of experimentation, we are now using a massively parallel algorithm at the Beclan Institute for Advanced Sciences at NCSA for preprocessing the MRLC data in a production approach. Before we even start clustering, this is done. Again, the result is smoother classifications and much less noise than seen before. When you look at the image and flash up spectral class one, the whole field lifts up in one chunk rather than a portion of a field popping up. So, the problem is in the evolutionary part of this. There is nothing in the literature that will verify the use of this for large regional inventories. I was talking to Tom over lunch about it, and what we have decided to do at the Natural History Survey is to put a sample image on Mosaic with a clustered image that has not been segmented, then a clustered image that has been subjected to segmentation. Now, this algorithm is free for the asking, it just requires one little thing; you have to have a super computer to run it. Professor Bresler, who has pioneered this at the University of Illinois for multi-temporal datasets, has indicated that if there were national interest in the algorithm, much like clustering now at EROS where they are using a standardized algorithm through spectrum software, it would be possible to place the algorithm on one of the NCSA super computers.

The program could then be accessed over the Internet. I am not sure what you would do at the other end with this program, but our experimentation has shown that this type of approach should be investigated. It shows some real promise especially in multi-temporal processing.

With that, I will close by saying we are kind of on a production schedule. In order to ensure that funding is continuous for the staff at the Natural History Survey, we will have a statewide preliminary land cover classification finished by mid-year. We fully intend that we will have to refine it again, but primarily through the segmentation technique we are able to move through scenes quite a bit faster than we have before. It takes about 15 hours on a massively parallel computer to do one whole scene. If you translate that to your basic 386 PC, it could be months, years, whatever. We fully intend to have a 16-class statewide classification finished by mid-year.

Through the help of two colleagues of mine that are here, Mark Joselyn and Tom Kompare, the DNR last year produced a compact disk of many of its photometric databases for the state of Illinois. I think there will be a volume 2 coming out soon. Mark knows more about this. He knows how it was produced and the cost. He has volunteered that if people want to send him an E-mail, give him a note, or something along those lines, he will send you one of these. If you are from an institution, there is no charge. If you are a private agency, it costs $100. That is the way it is being handled in this state. This is specific to an ARC/INFO workstation but Mark has found that it does work pretty well on the PC version also. So you can just stick it in your CD-ROM drive and start drawing immediately, which is really quite an enormous advance in our state.

Indiana and Ohio Reports

Forest Clark
U.S. Fish and Wildlife Service
Bloomington, Indiana Field Office

I want to thank Frank for inviting us here. We have several people from the universities working on the project in Indiana. I have been to two or three meetings on the Upper Midwest Gap Analysis Program now, and I always go back with volumes of notes of things to do when I get back. I want to thank Mike Jennings for his vision of the manifest destiny of GAP, which gives me the courage to actually go back, look through those notes, and try to figure out what to do with them. I am just going to take a few minutes to talk about Indiana and Ohio.

Ohio Report

In Ohio, there really is not a GAP project going on right now, but Jonathan Bart from the NBS Cooperative Unit at Ohio State University has asked me to say a couple of things. They are very interested in GAP in Ohio, and John is doing some work that parallels to the GAP project. His research interests are using GIS technology to predict distributions of endangered species. He has proposals right now to do some of that work, particularly with aquatic organisms. While there is not a lot going on in Ohio, John is keeping abreast of what is going on with GAP. He is very interested, and, when the opportunity arises, he will be ready to move forward. So, I guess I will leave Ohio there and let John speak for himself at the next meeting.

Indiana Report

We got started with the Indiana project about a year and a half ago with GAP analysis doing a pilot project on a site down in the southeast portion of the state. The site was called Jefferson Proving Ground, which is a military base that was undergoing base closure procedures. We selected this as a site for our pilot project for a couple of reasons. One, we wanted to try out the GAP methodology in Indiana to see if we could make it work. We wanted to see if we were going to be able to do GAP analysis. Second, we wanted to produce some products for the Department of the Army that might facilitate the base closure procedures. I think we met the goals of the pilot project pretty well. We did produce the basic GAP data layers at Jeffersen Proving Ground, or actually the study area which was a little bigger than Jefferson Proving Ground. We produced a vegetation map with about 12 classes using one May date imagery. We produced a managed-area coverage, which was not real interesting because we determined that the entire base was a managed area, although it was a very interesting area to work in over all. Using GAP in conjunction with a lot of more traditional biological surveys, we found that it produced a very biologically rich area. We had threatened and endangered species data and produced some wildlife vertebrate models. We tried the GAP methodology there, as well as National Wetlands Inventory. We also produced some products for the Department of the Army out of that pilot project. As a result, this site may become the next National Wildlife Refuge in Indiana. Although that has not been established as such, it looks pretty promising.

Currently, we are going back and revisiting the pilot project and moving on to the statewide project for which we received funding in November. This is just a little explanation of how things are laid out in the GAP project in Indiana. We do not have a cooperative research unit, so we administer the project out of the Ecological Services Office of the Fish and Wildlife Service, then the two universities are where all the work is done. We have several cooperative agreements in place and a few others that are currently being negotiated.

Because funding is limited to national-scale kinds of products, in the Indiana project we are going to produce the national GAP analysis data, but we are going to try to work at finer resolution on small parts of the state. The way we decided to approach this is by working on small subprojects, metaprojects, within GAP analysis, for which we have been starting to get some cooperators and a small amount of funding. Four projects are currently under way. The Indiana Department of Natural Resources is sponsoring the updating of the vegetation map. This is probably going to be at about the 11- or 12-class level that we worked on for the pilot project. The Nature Conservancy. We are really excited about working with The Nature Conservancy on some of their bioreserve projects. One project is currently getting under way, and a couple more are under discussion. We are currently working on the listing of the Northern Copper-Bellied Water Snake out at the Ecological Services office, and we're using GAP products to facilitate that effort.

One of the main focuses of the project and the reason we got into the GAP project in the Indiana Field Office in the first place was to try to use GAP as some kind of basis to facilitate our habitat protection, in particular habitat restoration efforts. Indiana is a lot like Illinois in that there are a few natural areas, but it is mostly agriculture. One of the things that we are particularly interested in is looking at the restoration of some of that area in the most logical, effective, and efficient way possible.

Missouri Report

Tim Nigh
Natural History Division
Missouri Department of Conservation

I work for the Missouri Department of Conservation as an Ecologist in the Planning Division. I am going to tell you a little bit about where the Gap Analysis Program has gone in Missouri. It has really undergone an evolution and become what we call MORAP, which stands for Missouri Resource Assessment Partnership. MORAP is an interagency effort to produce information, largely digital information, and technical support for natural resource planning and management in the state. When GAP came on the scene in Missouri about 2 years ago, many initiatives related to biodiversity conservation were already going on and many ideas about biodiversity assessment were already being formulated. Many of these initiatives were spurred on by a Missouri Biodiversity Council, which is very similar to the Upper Great Lakes Biodiversity Council. It is a council of 14 agencies and organizations, mostly State and Federal natural resource agencies, as well as organizations such as The Nature Conservancy and the Farm Bureau. The mission of the council was to coordinate the conservation of biodiversity and to implement the objectives in the Missouri Biodiversity Report.

So, the council had groups working on biodiversity conservation initiatives when the Gap Analysis Program came. The following are some of the working groups: (1) The Regional Goals Working Group. This group was looking at ways to get agencies to cooperatively develop and implement ecoregional plans for specific ecoregions across Missouri. After about a year of work, the group decided that we should use Coordinated Resource Management [CRM], a process that our agency was working on to develop these ecoregional plans and their implementation. According to Resource Management, it is an interagency effort. We have 12 agencies that have signed a Memorandum of Understanding to participate in the development of these ecoregional plans. We have divided the state into 10 planning regions based on 4 service subsections, ecological subsections. The goals were to develop a plan for each of those subsections which will address the conservation of biodiversity, address recreational use of natural resources, address education and interpretation of those natural resources, and finally, will integrate natural resource commodity production. We have actually started two of those plans and have really good interagency support in doing that. As we approach those plans, we are finding that they are going to require a lot of information. (2) The GIS Working Group, which investigated how we can use GIS and related technologies to develop information and support the CRM effort. The GIS group came up with a pilot project proposal that largely focused on one of the subsections in the Ozarks. They had some recommendations about the types of data layer that we wanted to develop and how to use them. (3) The Ecological Classification Working Group, which met for over a year. They recommended we apply the Forest Service national hierarchical framework to the ecological classification in Missouri. They also wrote a project proposal. We all went to this council and gave our individual working review report and the council said we had a lot interrelated interests and to come back with a cohesive plan that integrated them. So the working groups designed MORAP.

I will now explain the history of where MORAP came from and how GAP has evolved into MORAP. The Gap Analysis Program agreed to incorporate its activities into the MORAP project. At first, MORAP was a project but now we are calling it a partnership. It is proposed to develop and provide information and technical support for natural resource planning and management in Missouri, largely to support the CRI effort and its implementation. Right now, MORAP is in a proposal stage with articles of participation developed that everybody is willing to sign, etc.

First, MORAP will establish a technical facility to coordinate the development and application of digital information layers and also to provide technical support to agencies that are going to use these layers. We will actually build and develop a facility that will be staffed by a coordinator and three specialists: a GIS ARC/INFO specialist, a remote sensing specialist, and a systems analyst. We are looking to develop a facility dedicated to the production of these resource layers and their application. We then identified about 15 major projects, which are packets of information that we want to develop in a digital format. The top six are what we consider the major projects that are going to take a lot of effort, time, and money. Current land cover is one of the major informational areas we are going to need for resource assessment and planning, but we also want to get statewide soil mapping digitized. We want to develop a statewide ecological classification system down to the ecological land type and phase-layer level and have it digitized and available. We want to have an aquatic resources inventory of some sort. We want all statewide contours from the 7.5-min USGS quadrangle digitized. We want historical vegetation for the whole state. Then, there is other supporting information such as (1) animal/plant distribution data that will be part of standard GAP analysis, (2) public land ownership, (3) recreation and interpretation information, such as the location of trails, camp sites, and other outdoor recreational facilities, (4) special features like heritage sites, and (5) the potential environmental impacts of mines, gravel operations, etc. Resource commodities take a lot of the information from the FIA-type information and mining information, etc. At least for the county or zip code level, we have an idea of what kind of commodities are being produced at different spatial resolutions. Finally, (6) human dimensions, what are people's attitudes about resources across the state? What do they want to do, what do they care about? We have a whole staff dedicated to polling people about natural resources. We have a lot of that information, we just need to make it spatially available.

We have all these different information layers we want to develop. We treat each as an individual project, and each will have an interagency project team dedicated to it. The 14 agencies are free to identify which of the projects they have the most expertise in, or the most interest in using the information for. They will sit on that team and develop a project plan which will identify what we will have done, by when, by whom, and how it will be paid for.

The job of the technical facility is to facilitate the completion of those project plans, and to make sure that certain standards are met so that all of this information can come together at the facility. This information can then be used for CRM and its implementations and to produce national GAP products. The Gap Analysis Program agreed to take its money and efforts and incorporate them into the MORAP project, if we agreed to provide the national GAP products after 3 years. We want to go to a higher resolution because we believe it is needed for actual statewide resource planning and management. We want to incorporate a lot more information than GAP analysis does in our assessment and planning process.

MORAP is slated for 6 years; in reality it will probably take 8 to 10 years. In the first 3 years, we will develop statewide low-resolution products to support CRM. The CRM is rolling now and they need a lot of these things now. When I say low resolution, I mean things like statewide land type associations. We have already got subsections, we are going to break those into land type associations. We are using land type associations as land units on which we develop our objectives within each ecoregion. Another example of low-level resolution is level 3 landcover classification in the first 3 years for the whole state. We are also going to take a pilot region, one subsection in the Ozarks, and develop high-resolution products: higher resolution land cover classification, higher resolution ECS [Ecological Classification System] down to the ELT [Ecological Land Type] level for that pilot region, and soil series mapping instead of soil association mapping. In 6 years, we will do the high-resolution data for the whole state.

The beauty of the whole thing is we are going to ensure that the data layers will be compatible by bringing it all together in one facility. Another neat thing about this is we have an interagency planning process to do the biodiversity assessment and to actually develop plans to address some of the challenges we have in each part of the state. It is an example of GAP analysis kind of rolling into a state level resource assessment and planning process.

As for the cost, we will be more absolute about it as each project plan is developed. We went through a process to estimate a cost for each project. To build and staff the facility, we estimated 10 to 12 million dollars over a period of 6 years. That includes a million and a half to actually staff the facility and buy the equipment. Another 10 million is needed to support all of the projects themselves.

There are some important differences between the MORAP approach and GAP analysis as far as the assessment. Number one, it is based on the ecological framework. We are moving from ecoregion to ecoregion, subdividing those into smaller ecological units, and then basing all of our assessments on those ecological units, much as land cover mapping is being done. Within each, we are asking ourselves, what do we have? We are using information like the historical vegetation map. We are asking, what has happened to it? What do we have left? What are the major resource challenges in each of these ecoregions? We then use all this information to answer those questions. Finally, we will ask, where are the best opportunities to achieve certain regional objectives? How well have we covered those? Where do we put more effort? That is the GAP part of it.

In addition, it incorporates socioeconomic factors. Most will agree that we are not really going to conserve biodiversity unless we incorporate human utilization of the resources. So, we are trying to incorporate recreational use, commodity production, and resource education and interpretation into the whole conservation equation. Finally, we have this planning process to use the information to develop plans to make sure it hits the ground.

Where are we today? We had a MORAP Executive Steering Committee, that met yesterday before I came up here. These are the high level, assistant director types who are deciding if their agency will sign the Articles of Participation. They met yesterday and reviewed the Articles of Participation. Everyone was very positive about it and ready to sign; they were just bringing the Articles back to make sure everything is all right. Next, we will start forming project teams.

I am the coordinator of an ECS Project Team that will develop specific project plans, outline all of the who, what, where, when, and how much parts of each project. There is also an Aquatic Resources Project Team, a Historical Vegetation Project Team, and a Current Vegetative Land Use Cover Project Team. There is a Technical Committee that has outlined the amount of equipment we are going to need at the facility and has come up with some data standards, etc. There is a Search Committee that is developing the job descriptions for the people who will be at the facility. This committee will be in charge of the advertising and hiring for those people. The National Biological Service in Columbia, Missouri, which is now going to be the Midwest Science Center, is constructing a new building to handle this kind of information. NBS said they would dedicate a large portion of their new building to the MORAP project; therefore, we have a place for it to go. The University of Missouri-Columbia has agreed to establish this as a university center so all of the money and staffing can be funneled through the university. So, that is where we are with MORAP. It is really rolling and it is the ultimate evolution of where GAP analysis would go in each state. We will see if it actually works. It is kind of a complex beast.

Iowa Report

James Giglierano
Iowa Department of Natural Resources

The Iowa DNR attempted to do a land cover inventory of the state a number of years ago, and we repurchased a full set of multi-temporal TM imagery for the state. So, we are trying to reinvigorate our land cover program and we thought cooperating with the GAP program would be a good thing.

There is a lot going on in Iowa besides GAP, and SCS is also very much interested in producing some kind of inventory. We are trying to figure out how to make it work with all of the different agencies, and I am interested in hearing what other states are doing. The other thing that I am worrying about is the land cover classification. We had done quite a bit of work before we got knocked out. We were having a lot of trouble with dividing our classes and splitting them up into more detailed classes; our accuracy started going way down. So, I am trying to pick up some new tricks here. As far as I can tell, the GAP program in Iowa is just getting started so this is really good for us. We do have a NAPP site with a lot of the same data that Illinois has.

Preclassification Stratification Research Results

Jana Stewart
U.S. Geological Survey

I am going to be talking about stratification of Landsat Thematic Mapper imagery based on regional landscape patterns to improve land cover classification accuracy of large study areas. This was work I did to complete my Master's Degree at the University of Wisconsin under Tom Lillesand, as well as part of my coop program with the USGS National Water Quality Assessment Program [NAWQA].

The research was essentially supported in terms of satellite imagery by the University of Wisconsin and in terms of the ancillary GIS data layers by the USGS NAWQA program. The purpose of the work was to deal with the spectral variability that may occur within the landscape during image classification by stratifying large areas into spectrally consistent subareas. Within the context of this work, I had three objectives. First, to produce the best land cover map I could using a combination of techniques, such as stratification, hybrid guided clustering, and ancillary GIS data layers for classes that are easily confused. Second, to evaluate the utility of two different stratification systems and base this comparison on K-hat values and Z scores. Third, to compare stratified and non-stratified classifications.

The study area I selected was the Fox-Wolf River Basin in Wisconsin. This is part of the Western Lake Michigan Drainage Basin, which is currently under investigation by the NAWQA program. The Fox-Wolf River Basin is one of three river basins within our study unit and covers about 5,000 square miles.

I acquired a single day Landsat Thematic Mapper scene for June 24, 1990. One reason this area was selected was because of NAWQA's need for land cover in this area. Second, there was the single date imagery available, derived from the WISCLAND map project which was a precursor to WISCLAND. It was funded by SCS and was looking at the potential of statewide land cover mapping back in 1991. Current land cover is needed by the NAWQA program as we investigate the status and trends of surface and ground water quality within this study unit and within 60 study units around the entire country. We are looking at natural and anthropogenic effects that may be affecting the surface and ground water quality. While GAP may be asking, "Where should we focus our conservation efforts?" NAWQA is asking, "Where should we concentrate our water quality efforts?" For example, most of the northern area is forested and the southern area is mostly agricultural. Within the bottom third is the Fox-Wolf River Basin, so it is predominately agricultural with the northern part being forested. We are looking at some of the contributions of agriculture, for example, nutrients, pesticides, surface, and ground water in agricultural areas. In the forested areas, we are interested in effects of mining and clear-cut logging. In the urban, we are interested in more industrial things, such as polychlorinated biphenyls and dioxin that are in the water and sediments and are going into Lake Michigan. We are not just interested in the bad. We are finding fairly clean water in some of these areas. We are also sampling pristine areas and doing a lot of habitat work to find out what is good about them. If people have good management practices, we want to find out what their practices are.

The first step in this work was to calculate the first three principal components of the six reflective bands, then to overlay the TIGER line files and with heads-up digitizing, mask out the urban areas. So, my work was essentially done on the nonurban component. The stratification and classification occurred only on the nonurban portion of the image. The study area covers about 5,000 square miles and is characterized by geomorphology of a recent glacial origin. It is important for any large area or regional land cover classification to understand such things as geomorphology that may effect the landscape patterns. By gaining an understanding of this, you can really get an understanding of the spectral patterns that you might be seeing in the imagery.

I will give you a little bit of history. The glaciers came in from the northeast. The bluer area around Lake Winnebago was basically covered by the glacier. Then, as you move into the transition zone where you start seeing a little bit more green and red, this is pretty much where the glacier stops. So, we have got an area of terminal-end moraine that runs up through there. To the very far west, we have outwash plain. In the Glacial Lake deposit to the east is really prime agricultural land. It is clay surficial deposit and has very intensive corn, alfalfa, and dairy farming. As you move into the moraine, we see a lot more forested areas and land not quite as productive for agriculture, but still some dairy farming. As we move a little further to the west, we move into the outwash plain which is part of the Central Sands section of Wisconsin. It is pretty much sandy surficial deposits. A variety of things are going on here: we have some cash cropping under center pivot irrigation, we have some Christmas tree plantations, and we have some non-productive land, just grass lands. It is really important to have an understanding of what is going on out there before you decide what kind of stratification to use.

I started looking at a number of different stratification schemes. A few that I looked at were physiographic provinces, at the top left. The top right is generalized surficial deposit from Wisconsin Geological and Natural History Survey. To the bottom left is Omernik's ecoregions, a national ecoregion map. To the bottom right is Dennis Albert's regional ecoregion map. I overlaid these with the imagery and, based on my knowledge of the area and the spectral variation that was occurring, I selected two of these. I selected the regional ecoregion map and the surficial deposit map for Wisconsin. I must mention too, I did overlay the STATSGO data, and it looked very good. However, my study area was smaller than the area STATSGO recommends using their data on. They recommend using it on study areas that are greater than 6,000 square miles because of the nature of the sampling they went through. They ran transects across counties to identify soil types and did interpolation from that. Therefore, they do not recommend using it on very small study areas.

The regional ecoregion map I mentioned was produced by Dennis Albert of the Michigan Natural Features Inventory. It was under a request of the Upper Great Lakes Biodiversity Committee and funded by the U.S. Forest Service, the North Central Experiment Station. It takes an hierarchical approach and is based on maps of soils, physiography, climate, and uses vegetation as a check. It subdivides the Fox-Wolf River Basin into six different units. There are the two units on the far left, the Upper Wisconsin-Michigan Moraines and the Central Wisconsin Sand Plain. They are pretty much sandy surficial deposits. The division between those two, however, ends up being what we call the tension zone in Wisconsin. The tension zone is the northern limit of certain species and the southern limit of other species. Some of the species occur in both units but it is the limitation for some species between that, so it is a climatic zone essentially. The other three units on this sort of transect would be mostly clay. The middle unit is more of loam surficial deposits. The Northern Lake Michigan till plain in the middle tends to be more glacial moraine.

The generalized surficial deposits map I used for stratification came from the Wisconsin Geological and Natural History Survey. It was interpreted from maps of the geologic atlas. It was created for the state of Wisconsin to look at which soils were more susceptible to ground water contamination. It was almost a little too detailed to do a very good job of stratification, so these were generalized (interpreted photomorphically).

The sampling pattern I used for my groundtruth and reference data was a modification of the pattern that Tom talked about for WISCLAND. I overlaid the 7.5-min quads over my study area and divided it into QQQs. Within each 7.5-min quad, I randomly selected one QQQ for training data and one for reference data. Within each of those areas for the agricultural areas, I selected three to four public land survey sections, where I requested data from the ASCS offices. For the forest types, I did the groundtruthing pretty much on my own along with some help from the WISCLAND map project. I didn't have three person years to do all that the WISCLAND does, so I actually do not have the detail of data for the forested areas that will be occurring in WISCLAND, although I have similar type of data that they will be obtaining from ASCS. I used a minimum sample size of 50 for each of the cover types I was interested in. I requested more data than I needed for a number of reasons. Each county varies in the amount of information they provide, and not everyone reports all of their crop information. Therefore, in one public land survey selection, ASCS may not have all the crop reports for all the land. In some cases, errors may be associated with labeling information, so it was nice to have some extra information to play with.

I used hybrid guided clustering, and I heard about this through Marv Bauer, who is from the University of Minnesota. I know many people in the audience are more expert in hybrid guided clustering than I am, but essentially I went through step-by-step for an individual strata and delineated training-set polygons for one class within each. Then I clipped those polygons out, calculated statistics, ran an unsupervised clustering ISODATA [Iterative Self-Organizing Data Analysis Technique](Tou and Gonzalez 1974), and varied the cluster number until I obtained a transformed-divergence value which was greater than 1,500. This was sort of a preliminary screening. I repeated those steps for all of the classes. I had a corn image that I clustered, an alfalfa image that I clustered, and so on. After I had done this for each of the classes, I appended all the classes and evaluated all the signatures for all the classes together in one file. Then I set a transformed-diversions value of greater than 1,700. Of course, I ended up with inclusions of other classes within the clusters, so these had to be evaluated. I needed to throw out some of them. I used the ERDAS Quick Alarm which is essentially a quick parallel-piped classifier to help me do some of this evaluation and sort through some of these clusters. Once that was completed, I went through a maximum-likelihood classification. I then collapsed these classes back into the target classes. For example, I had five different corn classes that collapsed back into one corn class. They are separate spectral classes but they are one cover class. Finally, I smoothed with a 3 by 3 majority filter to remove the salt and pepper in the classification, so the accuracy assessment was done then on the smoothed image.

The accuracy assessment was pixel-based with the minimum sample size of 50. It was then evaluated based on error matrices and the kappa coefficient and comparisons for a statistical significance based on Z scores. This is one of the first three principal components of the ecoregion Upper Wisconsin Michigan Moraine. This area was more sand surficial deposits, and it is in the northern part of the study area, north of the tension zone. It tends to be more forested with a mixture of agriculture. I had classifications of row crops, forage crops, hardwood forest, mixed forest, coniferous forest, grasslands, and water. Again, I was limited to the number of classes I could have in the forested areas because of the groundtruth I had. That is why it is so important to get out there and be able to get a lot of groundtruth for the forested areas. The Wisconsin Wetland Inventory [WWI] was not completed for this area, so I was not able to do a wetlands or upland classification. Thus, a lot of the wetland areas will end up either in the grassland or the forested areas. In the large green chunk that you see there, the square boundary is not an anomaly, it is the Menominee Indian Reservation. This reservation is about the size of an average county. It is really a very beautiful, pristine forested area. It tends to be a fairly mixed forest. They have won national awards for their logging practices in the Menominee Reservation. They haven't created any type of monotypic forest at all, most of it being a very natural forest. To the north, you see a forested area which is part of the Nicollet National Forest. You can see that this has a bit more mixture of the grasslands and agriculture throughout that area.

The first objective I mentioned was looking at the utility of using hybrid guided clustering, ancillary GIS data layers, and stratification. In the 10 strata that I classified—three surficial deposit strata, six ecoregion strata, and one non-stratified classification—my overall K-hat ranged from 0.77 to 0.93. The lower scores were in the sandy surficial deposit area. The most confusion was between grasslands and forage crops. For example, a lot of old fields are in the sandy areas. There are a lot of old alfalfa fields that are out of production and somewhat going to grasslands, so we had confusion between classes like that. The higher end tended to be areas that were smaller strata. One problem I found in having too small an area of strata was that you do not have enough of all the classes to have a sample size of 50 for each. In those cases, one of these classes may have been missing, which can result in a little higher score. I would say on the average, there were no real significant differences between the strata.

The non-stratified ended up in the middle because the training sets I used for the non-stratified were a composite of the training set used for individual strata. It averaged out the low and the high. This didn't really test whether or not stratification worked. I essentially was using the best training set possible. One way to test whether or not stratification would work would be to actually blindly go in here, take the whole non-stratified classification, and come up with training sets without taking into consideration the different strata. I didn't test it that way, but I will show you another way I did look at it. I did composites of the ecoregion surficial deposits and the non-stratified classification with roads. The little grid that you cannot see very well is the roads overlaid on the imagery. The urban areas are the magenta areas that I burned into this. I didn't actually classify the urban areas, I just burned them in as one class because my interest was just in stratification for nonurban and the surficial deposits composite. The accuracy assessment was done on the individual smoothed strata but when the composite was made, I went back to the unsmoothed data, stitched them together, and smoothed the composite at the end. My conclusions for this objective were that hybrid guided clustering was very fast and efficient. I thought it was less subjective and it got at a lot of variable classes such as alfalfa in the agricultural areas. It varies quite a bit through the season because it is either mature, cut, or has recently been cut. With that kind of thing, it is very difficult to come up with training sets supervised, and it is very complex to do it strictly unsupervised. So, by specifying these training areas with polygons like you would in a supervised classification, you can break out all the spectral differences with the clustering. The ancillary GIS data layers were very useful for stratification, as well as for masking out easily confused areas.

My second objective was to compare the two different stratification schemes, but I really did not have adequate groundtruth to actually make a good comparison between the surficial deposits and the ecoregion stratification. I also wanted to see how well stratification worked versus non-stratification. To test this, I classified an individual stratum with a different stratum training data. Then, I took the reference data from that original stratum to do the accuracy assessment on that. I compared it to the original classification for that stratum. I did 39 classifications total, and I looked at all 30 ecoregion combinations. I tested each ecoregion with each of the other ecoregions' training data and ran the accuracy assessment with the original stratum reference data. I did this with all the surficial deposits combinations. I didn't actually do all of the combinations of non-stratified with ecoregions or non-stratified with surficial deposits, but I did do a few. The results of this were that in the 30 ecoregion comparisons, 20 had significant differences. Nineteen were significantly worse classifications than the results of the original training data. One classification was actually better than the original, but this turned out to be one of the areas that was over-stratified and had a limited number of classes. Ten out of 30 actually had no significant difference. In the surficial deposits comparison, there were four significant differences out of six, and those four were actually significantly worse than the original. Two had no significant differences. In the two non-stratified ecoregion comparisons, both were significantly worse than the original. Finally, no significant difference was found in the non-stratified surficial deposits comparison.

In conclusion, I feel comfortable saying that stratification can be used to improve land cover classification of large areas. Keep a few caveats in mind. First, the more generalized the land cover classes, the less important stratification is. The more detail you want, the more important stratification is going to be. There may be some boundary differences because of scale differences in the mapping, and for this reason modifying those boundaries may be important, as is suggested in the WISCLAND protocol. I know people have said if the people making the maps actually had the satellite imagery in front of them when they created the maps, perhaps they would have drawn those boundaries slightly differently in the first place. Finally, it is possible to underrepresent classes by overstratifying. For example, I had areas where I had broken it into such a small strata that I wasn't able to classify all of the classes I had in other strata. When I went to make the composite, I ended up with somewhat of a discontinuous classification because I did not have all the classes represented in one of the strata but had all of them represented in the adjacent strata.

I would make a number of recommendations. Of course, I only had a single date to work with, which was June 24, 1990, but as far as a single date, it was certainly not an ideal single date. I would have selected something later in the season if possible. Even better than that is to have the option to use the multi-temporal imagery for both agriculture and forested areas. Second, is to investigate principal components analysis, which WISCLAND has done. For the multitemporal imagery, they use the first three principal components of each of the two images, combined to a six-band image. I would suggest additional training and reference data, because for my work I needed more forest cover types and more detail in that particular area. The ASCS data were very useful, and the nice thing about it is if you are using imagery that is not the current year, you can actually get that information for those fields. Boundary modification based on photomorphic units is an important thing to look at because of the scale differences in mapping. You do not want to overstratify boundary adjustment to higher order, so in Dennis Albert's ecoregion map, which is hierarchical, it may not be necessary to go to the lowest level of ecoregions in order to do this classification. You may be able to go to the middle level. Dave will be talking about stratification into upland and wetland. I had problems with confusion between grasslands and other crop types. If I would have had a wetland stratification, I would have been able to separate these out a lot better.

Finally, another way to test this would be to do a classification of the entire study area without any prior knowledge of the stratification scheme and select training areas not based on the stratification scheme but on a different method.

You said that stratification may not be as necessary to use in the more generalized classification. Where is your break-off as to general?

Well, if you wanted to just say agriculture and forest, I would say it is not that important. As you start breaking down the forest or agricultural types, it was still useful for looking at row crops versus forage crops and looking at different types of forest mixtures. Even at that level, stratification was good. I had done that previously on an agricultural area and found that I could easily distinguish a lot of different row crop types. More often, the problem is that they occur at such a low frequency compared to some of the other cover types that it is difficult to get a sample size that is large enough. For some reason, you end up creating a row crop category just because the sample size isn't large enough, but I would say even for this area, the row crops from the sandy areas to the clay areas, it was very useful to have stratification at that level. The forest cover types wouldn't be as critical at this general level but as you try to break it down any further it is going to be important.

How do you handle what you have done to stratify. . .What do you do about the blending between classifications? How do you bring those, how do you resolve those boundaries?

If all of the classes aren't represented?

Well, I should say, do you have boundaries along the strata?

Right, that is why I did the smoothing. I stitched together the strata before they were smoothed rather than after. I then smoothed with 3 by 3 majority filter over the entire area.

And that took care of all of it?

Yes, it actually worked pretty well.

Wetlands Classification Research Results

Dave Nagel
Wisconsin Department of Natural Resources and the University of Wisconsin-Madison

I will be talking about my research work this morning. The title is Preclassification Image Masking to Improve Wetland Classification Accuracy. This is done in support of the WISCLAND project. This project was partially funded by a grant through the Forestry Department at University of Wisconsin-Madison.

The purpose of my research is to test the proposed methodology for the classification of wetlands for the WISCLAND project. The methodology is if we use the WWI boundaries, which are very similar to the National Wetland Inventory [NWI], to cut the wetlands out of the TM imagery and classify wetlands separately from the uplands, we can then put the two back together and have a full classification. Hopefully, this process will help improve the wetlands classification.

To give a little background on the WWI, it was first completed in 1984 and was created using 1:20,000 scale aerial photography, which was black-and-white-infrared. They delineated all wetlands 2 acres and larger. The maps are reproduced at 1:24,000 scale, and each map covers one township. There are wetland delineations on the map and categories of wetlands are defined. Those categories are somewhat similar to the NWI. The reason for doing this stratification is that often land cover types in the wetlands look similar to those in the uplands. For example, you find deciduous trees in wetlands and you will also find those in the uplands. It is very hard to make that distinction. As Dr. Lillesand mentioned earlier, other categories often get confused, such as corn and cattails and the sedge meadows and grassy wetlands with the meadows on uplands. These are just some accuracy results of some studies that have been conducted at the Environmental Remote Sensing Center.

My general hypothesis is that wetlands categorical accuracy can be improved when wetlands and uplands are separated before classification, as opposed to classifying them together in the same image. One objective of my study is to assess the categorical accuracy of the WWI. Another is to determine what classification accuracy we can obtain within the wetland boundaries using the TM. I also wanted to estimate the minimum-size wetland that could be accurately classified. Can these 2-acre wetlands really be classified as accurately as the larger areas? I also wanted to look at whether or not TM could be used to detect succession in the wetlands over time.

When studying vegetation in Wisconsin, it is important to know that there are two floristic provinces here, the northern hardwoods in the north and prairie-forest in the south. Basically, you find different vegetation species in the north than you will in the south. There is an area of transition in the center called the tension zone. I chose two study areas, one in the north in Marathon and Portage Counties, and one in the south in Dodge and Jefferson Counties.

The Marathon-Portage study area is 144 square miles. The landscape is primarily recessional moraines and pitted outwash. It is composed of four townships. The wetlands are primarily deciduous forest, but there are also numerous stands of coniferous forest, tamarack, and black spruce. There are a number of shrub communities here that have important species, such as small woody evergreen broadleaf shrubs, bog, rosemary, and leatherleaf. There are also some coniferous shrubs, tamarack, and black spruce. There are few emergent wetland areas in this region, mostly sedge meadows.

This is the TM imagery for the Dodge-Jefferson study area. It is the same size, 144 square miles. Primarily, ground moraine are here from the glaciers. Wetlands are in the river basins—mostly emergent wetlands and grassy wetlands. There are also a lot of cattails and sedges. There are some shrubs too, red osier dogwood, willows, and some deciduous forest, black willow, and cottonwoods. A few isolated stands of tamarack are also in this study area.

The data that I used for the Dodge-Jefferson study area were TM imagery acquired in August 1991, and for the Marathon-Portage area I used a June 1990 image. I also used aerial NAPP photography for the accuracy assessment. In addition, I had the WWI for both study areas. The general methodology was first to obtain the TM imagery, the NAPP, and the WWI, then to rectify the TM imagery. The WWI for the study area that I was working in was not digitized when I started, so I had to digitize it. When digitizing, the wetland vectors did not fit well on the TM imagery because the WWI was drafted on non-rectified aerial photography. Therefore, I had to transform the vectors to fit. Then, using that vector file, the wetlands were pulled out of the TM imagery so that the TM image was with wetlands only. I classified the TM imagery and did an accuracy assessment on that. I then tested the hypothesis, did a classification of the wetlands and uplands together in the same TM scene, and did an accuracy assessment of that. Finally, I compared the accuracy for the two classifications.

The first step was digitizing the WWI maps. The root mean square was about 50 to 70 m when I was doing the digitizer set-up, which is greater than a pixel. When I overlaid the WWI boundaries on the TM imagery, the fit was not great, although it was pretty good. When I did the transformation choosing ground control points, things seemed to fit much better, within one pixel. We were then ready to pull the TM wetlands out of the TM image. This is a procedure the DNR is using now. They are scanning the WWI maps and then a transformation takes place in order to make them fit the map projection we are interested in. So, for people using the NWI, I am not sure if your data will be rectified to fit the TM imagery, but it was an important step in this case.

Now, we had two files that were supposedly wetlands only. I wanted to make sure these were indeed wetlands only and uplands were not included because of some misregistration problems. So, in each study area, I took 100 random sample points in the wetlands and decided whether each point was indeed a wetland or an upland. This distinction was based on the pattern and tone of the pixel that was selected and the pixels around it. For the Marathon-Portage study area, it seemed that 94% of the pixels we thought were all wetlands actually were. Uplands inclusion was 6%. For the Dodge-Jefferson study area, we found that 95% of the pixels were actually wetlands.

I will split off a bit and talk about the Marathon-Portage study area and the classifications there. I collected groundtruth primarily for labeling the clusters that came from the unsupervised classification I did for this area. I collected 96 points; these were points that were along roads that I had found in the WWI. I looked at each of those 96 points on the ground and decided what type of vegetation was there. Then I came back and compared what I had found in the field with the categories that were on the WWI maps. I found an overall correspondence of about 78%, so in this case the WWI categorically was about 78% correct. For this study area, the classification was unsupervised and iterative. I started out with 30 unsupervised classes. If a number of classes did not seem to represent known categories; they were brought back out and extracted. A classification was run on those again until we had a decent product. I ran a 3 by 3 majority filter over the classification when that was done. There are seven categories: emergent wet meadow, bog vegetation (which is basically that low woody shrub), needle-leaf shrub, deciduous shrub, coniferous forest, deciduous forest, and open water.

The next step was to do an accuracy assessment on the classification. I selected 250 random cells in the classified image. I then located those same areas in the NAPP photography and used photointerpretation to determine what the vegetation was at that point. In some cases, it was really hard to determine using the NAPP. In that case, I went to the 1:20,000 WWI photography. If I still couldn't determine what it was, I just looked at the WWI map to make that decision.

Overall, I found the accuracy for those seven categories to be about 77%, which was not as good as we would like. Notice the three shrub categories: bog vegetation, coniferous shrub, and deciduous shrub. Those three are not in the minimum WISCLAND classification. Those are just some extended categories I thought I might be able to define. When I lumped those three categories to reflect the minimum WISCLAND classification for wetlands, I came up with an overall accuracy of 82%, which is a little bit closer to our goal. The next step then was to do the classification when the uplands and wetlands were together in the same file. I classified the same seven wetland categories and added an additional generic class called upland. This was again unsupervised. To test the accuracy of this classification, I did two things. One, I wanted to test the ability of TM to do the upland/wetland separation. So, as a reference file, I used the WWI and compared that with the TM classification where all the wetlands were coded as one class wetland and the uplands as upland. In doing this comparison, I saw a 77% agreement between those two files. That is what we can expect in distinction of upland from wetland using TM.

The categorical accuracy in the wetland areas, when the wetland and uplands were classified together, was 52%. This is lower because a lot of things that were actually wetland got classified as upland.

When I combined those shrub categories again, that accuracy improved a bit to 57%.

For the Dodge-Jefferson study area, again the first thing I did was go to the field and collect groundtruth. I collected 64 sample points in this case. I compared what I saw on the ground with the WWI again and found an accuracy of 79% for the WWI categorically. I again did the classification and, in this case, I used the guided clustering to try to improve the accuracy a bit. This is the classification for that study area. I only classified five categories, which is the minimum WISCLAND classification: emergent wet meadow, deciduous shrub (there are only deciduous shrubs here but in the WISCLAND it would be called the lowland shrub), deciduous forest, coniferous forest, and open water. I did the accuracy assessment slightly differently. In the first case, I took the points randomly from the classified file, and a lot of times it was difficult to interpret what that point was on the NAPP. In this case, I went to the NAPP first and said, "I know that this is a deciduous forest, and I know that this a shrub area. I will use this for reference data." That is a little bit biased, but it improved the accuracy of my reference data. The overall accuracy for this classification was 90%. That was an improvement from the first, probably because of a combination of reasons: the guided clustering, changing the accuracy assessment procedure, and less complex wetlands in this area. I have finished the upland and wetland classification for this study area. The upland/wetland separation was 63%, and the categorical accuracy was about 55%.

One of the other things I wanted to do was see the minimum-size wetland that could be accurately classified. I looked at wetlands of different sizes: one group, 2 acres; another group, 5 acres. I took 50 sample points from each and looked at their accuracy using the WWI as reference data. The smaller polygons, 2 acres, had an accuracy of about 50%, and the larger ones about 70%. I did the same thing for the Dodge-Jefferson study area and got accuracies of 64% and 72%. This is important for the users of the data to remember that even if the accuracy of the wetlands in the WISCLAND classification is 80% or better, maybe those small wetlands do not have quite as good an accuracy.

My findings to date are the geometric accuracy of the digitized wetlands was adequate to allow us to use this procedure. The categorical accuracy of the WWI was determined to be 78%-79% in this case. Since then, we have gone into the field with the WISCLAND project, and I found it to be about 80%.

Was that at the full level of the WWI classification or the collapsed version?

Yes, that was at the full level. So, you can't really compare the classification of WISCLAND accuracy with the WWI accuracy, but it is important to know if you are going to use the WWI for accuracy assessment or guided clustering.

Wetlands less than 5 acres were less likely to be classified accurately than the larger ones. The preclassification image masking did improve the accuracy of wetland classification quite significantly. I also found that guided clustering was effective for classifying the wetland if there is adequate groundtruth available. For this study, I did have enough groundtruth for some of the areas that we will be doing in WISCLAND. I know that Minnesota is considering using guided clustering in their wetlands. I guess I am a little bit reluctant to do that for our study, but maybe that is something we will discuss this afternoon. I mentioned earlier that one objective was to try to detect succession in the wetlands using TM and quickly found that I wasn't going to be able to do enough field work to do that. Wetlands do not succeed like the uplands from grass, to shrub, to trees through time. Sometimes, they go from trees to shrubs depending on changes in hydrography, so it was a little bit more than I could chew. Are there any questions?

In your analysis, did you try to determine how many wetlands were left in the upland category after you masked?

No, I didn't. It would have been an interesting thing to do. From our point of view now, we are considering the WWI as gospel to make that distinction between upland and wetland.

Why bother with a TM classification if you are treating it as gospel?

It would be good to have a consistent minimum mapping unit across the state. Also, if we have TM and we plug into WWI, we go from 30-m pixels to fairly large polygons in a lot of cases. Sometimes, in the TM imagery, a lot of the subtleties between categories within wetlands can be picked out, whereas the photointerpreter does not have the time to make all of these little distinctions. Sometimes, those things can come out in the TM. The categories in the WWI do not always lend themselves to the straight classification scheme that we have with WISCLAND. There are a lot of mixed categories. Often they do not go to the most specific category, and they generalize up. So, there are a number of reasons.

Another point is that the majority of the state is based on 1979 photography, and there has been a lot of succession. We are not trying to replace the WWI, we are trying to supplement it. It is a major DNR program and those data are used for regulatory purposes, so there is some sensitivity. You may not be trying to replicate it for a good reason. In thin budget times, we need to make sure our program capabilities aren't in vain. The WWI program is potentially on the cutting line, which could be a real problem. You simply cannot get around the detail with TM data that you can with aerial photography. A *lot is involved and we have been able to get the visuals from the WWI completely updated by the end of this calendar year in digital format so that the next time someone does a study like Dave Nagel did, they will not have to digitize the hard copy. That data existed in digital form but had not been updated, so he was not allowed to make a copy of it. When it is regulatory data, these kind of considerations come to the foreground.*

Landsat Thematic Mapper Data Status

Daniel Fitzpatrick
Environmental Management Technical Center
and
Mike Jennings
National Biological Service

Daniel Fitzpatrick: I'd like to start by addressing those of you who have heard about what we have gone through to get this imagery. Though there have been a number of glitches in the image preprocessing at the EROS Data Center [EDC], I am more confident than ever that we have finally worked through all of them, and from now on we are going to get scenes delivered that are registered in the correct datum, with the correct corner coordinates, and the dual dates co-registered.

We are in the queue with most of our requests. That means that EDC is preprocessing a large number of scenes all at once, going through it in a step-by-step fashion. Several scenes go through the first process and then are passed on to the next step. It takes a long time going through the queue to complete a scene, since they are not just working on one single scene at a time.

I have given you the most recent list of preprocessed scenes that has been E-mailed to me from EDC. There are two scenes in path 20 and one scene in path 21 that are estimated to be shipped to us today. Those must be Michigan. There are two scenes in path 22 and two scenes in path 23 scheduled to be shipped. All of these are single date scenes. The scene for path 27, row 28, which is the priority scene that Minnesota has requested, was scheduled to be sent 2 days ago. So far, estimated shipping dates have been accurate.

I am still not entirely familiar with the whole MRLC process; how scenes are selected and what the process is, and I get a lot of questions. There are logical reasons for why things were done the way they were, even though at first it may sound illogical. Since we have Mike Jennings here, why don't we call on him to answer any questions about why the processing is done a certain way.

Mike Jennings:

The other thing to keep in mind is that there have been no institutional barriers or any other kind of agendas in dealing with this. What I am really saying is, we are all on the same team. What we are facing are technological and data availability problems. That it came off at all, to me, still remains absolutely a miracle. We have been able to take this data and at least begin to put a standard clustered dataset into the public domain.

One of the things that I did that was pretty interesting was to go back and flip through the *Journal of the American Society of Photogrammetry and Remote Sensing* and look at some of the history. Tom Lillesand, I read not only your article, but also your testimony before Congress in 1982. The commercialization of the Landsat data was the most disastrous thing that has happened not only to the remote sensing community, but at this point in our history where there is a real confluence between the geography and ecology, it set both fields back 15 years. No question about it. The effort by the MRLC

Consortium was the first serious effort to overcome that. This effort resulted in people now trying to deliver products more quickly and being more responsive.

To get back to the TM scene processing, we sifted through thousands of scenes. I spent hours on the phone negotiating scenes, checking scenes for system errors and cloud cover. Three spots in the country are the real rascals when it comes to clouds: northern Maine, northern Michigan, and central West Virginia. Northern Michigan was a black hole for a long time, and we finally made some tough decisions regarding what scene to choose there. We sidestepped phenomenal bureaucracy to do this also. The decision to do this was at my level. We didn't ask anybody, we committed a million dollars, and we just did it. We just held our breath and did it, and it worked out. We managed to overcome a lot of things in doing this. Now, EDC is taking a very similar attitude to what we did in terms of really not asking, but just simply taking on the job. Now, they are dealing with 531 scenes; that is, archiving, preprocessing, and trying to do so as economically as possible using some new technologies. You really have to appreciate that there have been a few people who have lost some weight over this thing, because they have been working double time. EDC is presently working three workstation positions in two shifts to try to get this out. They are feeling extremely responsible and are feeling a lot of pressure. When I hear a lot of complaining about the scene processing, I am empathetic, but at the same time, there is a giant other side to the whole equation.

We need to recognize that since the commercialization of Landsat, EDC has been out of the digital-image-processing business. They have been working with archives, but they really haven't gotten the resources to keep up with the state of the art. The MRLC effort has allowed them to get back into the game, but we have to be a little patient.

Precisely. Don Lauer, Director of the EROS Data Center, redirected a million dollars of his own budget to bring some of their equipment up to speed and to invest in the MRLC. That was completely unexpected. That was not in buying imagery, that was in gearing up for preprocessing.

GAP projects are developing across the country and are currently active in 37 states. About five states are organized and ready to go, but still lack funding. All of a sudden, around 20 states went crazy with requests for the data once they got the word that there was going to be imagery available at no cost to them. They had graduate students and staff, exactly in the position that you are in, waiting to get imagery. None of this is unique to Upper Midwest GAP. In fact, a lot of priority has been given to this project because it covers a multi-state area, which is really important. Only two other scene areas in the country have received dual-date scene coverage. Just two scene areas, not two projects. There was priority given to get the dual-date scenes out to Upper Midwest GAP, at least for some scene areas, so some work could begin.

In the ferocious negotiations between MRLC and EOSAT, we walked away from the table three different times, to let it sit for a week or so before resuming negotiations. One of the deals was that we would receive a reduced price on the imagery if we provided EOSAT with the full collection, which they could then market. It is called "The Best of the United States," a full set for the lower 48 states. This is the first time that has ever been done. By marketing that set for 1.4 million dollars, EOSAT is recovering some of the cost they lost giving us a good price. What that meant was that we were contractually obligated to get, in first downloads, one single scene for each of the 430 scenes that cover the lower 48 states. Our dual-date scenes didn't come in until later, so this was another problem. This has been an extremely complex process, with a whole series of things that needed to line up. There have been some serious problems, such as using the wrong datum, but it has been amazing to me that it has actually worked as well as it has, for as inexpensive as it has been.

Now that a lot of the glitches have been overcome, we are going to see an output of a tremendous amount of data.

One of the other important things coming out of this is offering a spectrally clustered dataset that is public domain. Two hundred and forty spectral clusters per scene, available as a public domain product is a pretty big deal. Again, those are all consistently preprocessed, and we accomplished quite a bit by doing that. The other thing to remember is that the scene acquisition and processing has to fit in with the real burning needs of all the other MRLC projects.

Open Discussion Lead

Dr. Tom Lillesand
University of Wisconsin-Madison

When Frank and I originally put this part of the program together, we were anticipating the need and having the hope that we could talk about any of the technical details that some members of the audience would not necessarily want to participate in. This is our opportunity to provide input, seek closure, or address issues if you want. Let me indicate some of the things that I thought we might want to put on the agenda for this session and then ask for any other topics people have.

We clearly want to talk about the extent to which we have closure on the classification system. We may still have some suggestions, certainly, but I would like to talk about the classification system. It is certainly of interest to explore, at some level, the consideration about the cloud cover screening, how important it is.

I know the Minnesota folks have had experience both with the cloud screening and with the guided clustering. I hope we will hear something about it, at least the status of their work and the process of trying to implement guided clustering in an efficient manner. We have the fundamental issue of ground-reference data. As has been pointed out among the states, many forms of ancillary data have the potential for utility as training and/or accuracy assessment of datasets. What we want to do is begin a dialog about what constitutes acceptable criteria in the context of the three-state effort for the use of those data.

Last on my list is the consideration of the multi-temporal MSS approach suggested by the Minnesota folks; and in the same domain, what are our criteria for accepting variations on the theme. So, we can take a look at that. Those are the things that I have on my list, but are there others that people want to talk about?

Heather: What are we going to do when it comes to a scene that has an ecoregion that falls equally across the borders in, for example, Minnesota and Wisconsin. What are some strategies for this situation?

Tom: That is a good point. Don Luman expressed a similar concern. He wanted to make sure that there was some degree of compatibility at the Wisconsin-Illinois border as well.

Frank: My perspective and Ken's is to go as far as to classify into the next state. Then try to look at the compatibility with this overlap rather than stopping at the border, because the procedures that are followed are similar. Not so much in that the line matches at the border, but that some of the procedures match. That is something you do, just keep going full scene width and some agreed-upon distance into each bordering state.

Do you know the distance that has been used previously?

No, I don't have any particular number.

When I talked to Mike about this issue, he said he finished the polygon that crosses the border into the next full polygon.

Except there is a river between us and Minnesota.

Frank: In that case, it is obviously not a thing to worry about. It is the same situation with Michigan if you look at it that way. The southern border and the borders between the UP are those areas that matter. Where they exist, we should try to do a couple polygons into it.

To the extent that we have the imagery, even though we don't have groundtruth, it might be worth doing the whole SCCU [Spectrally Consistent Classification Unit] in the next state.

I don't think there is any value or need to go more than one ecoregion into the next state. Continue an ecoregion into the next state or to the edge of the frame, whatever comes first. In Minnesota, that is just a little piece.

Heather Reese: I am thinking of one example in particular we were looking at today. It was one SCCU that falls about 40 miles into Wisconsin and 40 miles into Minnesota. We are using a certain number of groundtruth polygons for Wisconsin's part of the SCCU. I don't know if they have the same date of imagery in Minnesota for that same area. The scene we are working with now cuts off the SCCU.

Well, we'd just go to the scene boundary in that case, rather than doing the whole thing. They would then go to their scene boundary, and we would have an overlap.

Heather: Then, I guess we would have to verify who is going to do which.

Frank: You're both going to do each and then look at the overlap.

Tom: You wouldn't both do the entire SCCU, just the portions on your images. The image overlap area would be done by both groups.

Then you can compare the classification also.

Are you trying to get an applicable training date?

Tom: I suppose that could be, so we could actually process it as a unit. I guess maybe what this suggests is we should look carefully at the dates that we get, and that the coverage is a function of the modified ecoregions or the SCCUs.

If you want to look at classifying states adjacent to each other, I don't think there is any problem with getting the scenes.

Tom: I guess what I am hearing is that we would go as far as the border of an SCCU or the scene boundary, which ever occurs most proximate to the state boundary.

The classification would be based on our side of the border. You would just run it through the rest. There are the upland, the wetland, and the urban components that get more troublesome. You can certainly just pump it through all as uplands and see what the classified status is.

It seems, in some cases, we could be classifying an awful lot of real estate that we really do not have a need to classify. I guess I haven't seen the SCCUs and how they fall into Minnesota.

I don't think any of them cross the river.

Heather: No, they stop at the border when you're looking at the river, the river tends to be the boundary, at least in this first scene that we are looking at.

It is all the way down to the Iowa border.

Tom: Dan, is this something maybe you could facilitate by looking at the coverage?

Dan: I don't think most of them cross the St. Croix or the Mississippi Rivers.

There are two counties that border Wisconsin that do not have a river.

Dan: Right, that is the only part we are really talking about, I think.

Heather: Right, and that is the part we were looking at today.

Frank: The benefits to this go beyond cost because we're both doing this little bit of verification back and forth across the states as to how well we match. We can cut off a lot of the study. We know we did the same thing, so it is all the same, but this gives us a little bit of verification. It is a worthwhile effort if we're going to be doing regional classification.

As long as we don't have to handle a whole other scene.

Tom: It would be real important if Dan were to synthesize this and propose a plan so everybody knows what the objective is, and we could modify as necessary.

Bill Befort: As far as I am concerned, the EMTC can lay down the law. We will follow any guidelines you guys care to put out on that. If it turns out to be a whole lot more work, then I will have a concern because we have got a lot of work in front of us. In terms of having the classification based on our side and our groundtruth and then simply when we run it through, including the whole image, that is really not that much more work. It is primarily computer time. It is a little bit more stratum boundary to adjust, I guess.

Heather: My concern is that if there are other cover types within a patch where we are not groundtruthing, then what are we missing?

It won't be as good, and theirs will not be as good on the Wisconsin side, but we will have an overlap you can use to feather them together for comparison.

Frank: That's all part of the effort, not to get down to the details of being concerned with missing cover types from one side. That would be picked up on the other side.

Tom: So, I guess, the way we would propose to handle that is Dan will look at the image coverage in conjunction with Denny Albert's map, and then get back to the respective states with a plan for who is to do what. We would hope to have overlap, but a portion of it being done by both.

I will give you my take on where I think we are with the classification system, and then I will open it up for reaction from people in the forum. When we originally started this dialog about how to proceed in a general sense, the atmosphere for right or wrong was kind of a "them versus us" situation: "The ecologists know what they want, and the technologists have no appreciation. They are trying to run the show, where they are forcing the outrageous system in the other direction." My sense of things is that we have come a tremendously long way from that kind of situation, which is the result of the continued educational experience from both sides of that street. Having said that, I think when we look at the TNC classification system and its fundamental orientation towards both structure and floristics, we might say that GAP goes in at the alliance level. One might think then, by extension, things could be aggregated up in either system. Regrettably, that is not the case. If you enter at the alliance level, you could get back through different kinds of structures going up in the TNC classification system. Whereas, the WISCLAND classification system is designed to be aggregated up and to improve the accuracy along the way. By design, it is compatible with what we expect out of the technology and the detailed extended classes tend to meet some of the need for additional information classes.

One of the other things that strikes a number of people is that if we classify using modified ecoregions as the basis for stratification after completing the classification, we could make a secondary effort saying, "Ok, now that we have jack pine here, we know there is a high probability it has this structure, given that we have limited the scope of this classification within this region." We would infer a particular kind of structure from that by virtue of having it within that region. We might envision downstream doing some more detailed analysis of particular associations that might be trained upon, looked at, or integrated with other forms of ancillary information. I don't think we have, by any stretch of the imagination, a magic fix between what we have in the WISCLAND proposed classification and TNC classification. We are getting to the point where we are getting some degree of accommodation in how to at least relate to one another. Do others see it that way? Are there flaming exceptions or reinforcements?

I should point out that Bill Befort was kind enough to send me a number of considerations that Minnesota has relative to the WISCLAND protocol, part of which was included in Ken's talk yesterday. They are proposing certain deletions because of their irrelevance in the Minnesota case, as well as certain additions that Wisconsin did not integrate into the WISCLAND protocol. In that situation, we would try, at least for our own purposes, to compile a master list showing all of the categories. We would then star those that are used only in one state or the other and envision handling the Michigan situation the same. It would be nice to have all of the possibilities on one composite list, but then show specifically where they are not utilized in a particular state.

Paul Tessar: While we are doing a little classification system maintenance, we probably also want to take a good look at the CCAP classification to see if we can become more compatible with it. We can either adjust what we are doing or look at definitions, maybe there are some categories that match. The CCAP is viewing the classification from a very large remote sensing key of orientation, so there are probably things in their classification that occur in the Upper Midwest and perhaps we should find that out.

Heather: NAWQA is planning off of a CCAP classification.

Tom: Now that we are in the production phase, we can no longer say, "Someday we have to settle on a classification system." That day is really upon us and now we have got the broad framework. We do want to make sure we get this composite list and any sort of last minute modifications that we might want to make prior to charging off with the process.

As we go through these image analyses, the important thing is that we all analyze with an eye toward preserving information about the character of the spectral classes that we may be rejecting or treating as unique. As we get into this and we find that we have some unanticipated differentiability, it is important that data are not lost because we made a decision now to not have them be part of the system. That is a lot easier said than done, but I would hope that we don't have such a large population of people working on this data that we can't ensure they understand this point.

Likewise, if there are new things we find that seem like they are reliably identified, let's go with it. Someone can always generalize it to a higher level at their option if they do not like it, but if we stomp on it, it won't be there to choose.

Tom: For something like a training set, we could keep track of the nature of the process. Was it an extremely hard process in this scene given the intermix of types, or was it easy? More specifically, what kinds of categories do you think were represented by particular clusters?

I haven't gotten any feedback on Michigan. Did you hear anything, Dan, on the classification system? You probably have to go through the same thing if they have additions.

Dan: What I had envisioned after the scenes are processed, is that we really need to have the three-state committee, including specialists and ecologists to say, "Does this meet our needs for this purpose or that purpose? What is the next step?" and "How might we enhance the data?" or say, "This is what we were able to get, how does it meet the needs of the kind of classification we would like to have?" That is not going to happen until you do the classification. At that point, it will be more critical that we begin formalizing a three-state approach of looking at it, rather than what seems to be an adequate approach that worked within the states at this point.

Bear in mind, too, that when this was initially put together, we had a base minimum. Then we listed a series of other additions to that and we stated that the classification is extendable. We can add classes to it if we are finding them. It is not necessarily that we have all possibilities up front, but what we are going to be able to have in the future. We started this with a base minimum and it is extendable, so as far as having a complete list up front before we get going, I don't think that is necessary.

Tom: We are to the point where if we could just get this composite list of what we have anticipated as extended classes, then maybe in turning that list to Dan and Frank, then consider the needs from there. But you have something that you can treat as gospel until further word and information. It is a little difficult, I am here to help out with the technical aspects in any way that I can, and I am not about to tell Minnesota, Wisconsin, or the EMTC how to do this. So if there is a better way, please holler.

We have mentioned this idea of saving the second most likely class for each pixel. Do you folks have the technical means for doing that in place?

Tom: We have not deployed resources in that direction. With the methods used at our facility, we would certainly have the technical wherewithal to develop the software to do that. That is in a whole area of accuracy assessment that I didn't mention on my list for accuracy portrayal. It would be helpful for us to know if others agree with the wisdom of trying to approach that. Regrettably, we don't really have a lot of research background so that we can support its importance. Basically, out of the classification process, we have a really good statistical explanation of the probability and the marginal probabilities among the classes. It is a real shame that we need to throw away this information. Obviously, we are not using maximum likelihood in all locations of the image, so it is not even uniform across the classification. It is certainly information that, downstream, people are going to want to model with this database. They might want to know with what kind of integrity the choice is made at one time versus another. Maybe an equally or almost equally sure call could have been made, as opposed to one that is quite dramatically different from any other possible candidates. It may be important later to model with these data in the context for information.

Even simply storing the density probability function from the most likely class is very useful, because we find holes in our samplings based on things that are not good fits.

Tom: I am very confident we can do that. I know Pete has been through maximum likelihood more times than most pixels have been through it. It is just that it is a certain amount of his time and a lot of software effort. Originally, we hoped that we might be able to convince ERDAS of the wisdom of it all. I just do not think that is going to happen. If there is a segment that is important, then we will see how to make that happen.

You may want to check out Khoros (Khoral Research, Inc. Albuquerque, NM) at the University of New Mexico. Their public image processing system is very powerful and has the ISODATA maximum-likelihood classifier. You are also provided with the source code, so it may be a modification, if they don't already do it.

*We can't be changing horses in midstream. We are locked into ERDAS and it is one thing for Pete to write a routine that reads a *.lan file, puts out *.gis files, which we substitute for a single ERDAS command, as opposed to changing our whole technical approach.*

Well, with this offer, you get the library as well. They may already do it and all it may be is a data conversion between ERDAS and Khoros.

Ok, so they have got that interface built.

Someone in the Soils Department had written that. So, it is just some place to look. In fact, they may have integrated those conversions directly into Khoros Version 2, which came out recently. It is just something to take a look at.

Tom: Are they the ones that did Spectrum?

Heather: Yes, it is one of their products. I might have a contact person. I was out in Denver at a Spectrum training session, and one of the people from Khoros came and was taking our comments as far as additional things they might program into Spectrum. He might be the contact person to talk to about this.

Tom: Yes, those are all real good suggestions. One way we might be able to approach this is to determine the currently available version and use it on the scene after we have done the training. Then use that version to evaluate some alternatives to really see if this is going to make sense.

Don't we want to use the table look-up classifier anyway for the sake of efficiency?

We are going to be starting, in the next couple weeks, with this new image and classifying. Do we just start classifying with ERDAS?

Tom: I think the answer to that is yes. I don't think that we want to delay everything contingent on the probability that we are going to learn more from additional probabilities and the maximum likelihood.

Once we have the classes and the images, we can go back and do a reclassification fairly easily. From my perspective, the classification is the easy part.

Except that it is an intermediate product that needs to be mixed and matched with data flows from five other sources so that whole operation would have to be repeated then. It is not just a single classification, it is the upland component of that SCCU where we end up with a classified image with 47 different parts that come together, so redoing 27 of them means you have to redo the whole synthesis job, which might be quite involved.

The first side is that we wait several months until we get the answers and get the new code. My sense of things is, I don't think we are going to do that.

I was wondering if you people have actually used the EML or think that is a possibility? If you use EML scripts, you could maybe substitute that.

If the maximum-likelihood classifier code was EML [ERDAS Macro Language]-based itself, but it's a binary. I have talked to their regional sales manager, and they have no interest in that modification. Not many alternatives are available, so it probably wouldn't take all that long to check. From experience, that code is very easy to modify.

How does the Minnesota contingent feel about retaining first-most and second-most likely classes, as well as the probabilities? Is that something we are pursuing?

Yes, I would prefer we did it that way, but we need to do it up front, we can't go back. We have already gone back. We're going back the second or third time, and we are doing the same scene. We do not want to do it a fourth time. So, either we plug it in later for future work or we can get right on it and make it happen and use it for this region that we are about to launch.

Heather: It is not in place right now anyway. There aren't the tools to do this right now.

I guess there are two alternatives: Khoros or the table look-up classifier that we have already developed. If you could make that run and you could enhance it to put out four files instead of one.

That would be fairly easy, the major problem is going to a new ERDAS image format.

You're set up for seven rather than eight?

*No, that is not the problem, it is the problem that the image format in the *.lan file is a fairly simple format and the image file that Imagine uses is fairly complex. So it is the classifier. That is really the only hard part.*

Ok, so in other words, if you use the old file format, you could have this done in 2 days in both departments and with the new file formats, you're talking 3 weeks to 2 months?

Yes, well I can do it in a week with the old file formats, which would be nice.

So, our technology has once again left us behind.

Tom: We aren't going to be able to decide all of this on the fly, other than that this is collectively a goal we want to pursue. If so, it means that we are going to have at a minimum to deploy some of Pete's resources in this direction. We have to look at what the trade-offs are in the existing tools, as well as something that might be a rewrite.

Would getting more documentation from ERDAS on the file formats make that job easier?

Tom: Yes. Frankly, I don't think ERDAS feels that this is a problem. There is not a big market out there saying "Gee, why can't you retain the probability density function values and the second most probable class." I do know in talking to the principals at ERDAS that they are interested, at the very highest level, in the whole concept. I would be surprised if we can't at least get some participation in respect to file structure and so forth. We will push on that.

I am going to check off my list what I call border wars, that is, the classification of borders classification system on the provision that Bob and Bill develop a composite list and supply them to Dan and Frank. Pete will get going on the second most probable class, etc. Could we maybe jump to the cloud-cover issue? First, what the Minnesota folks are doing, and how strongly we feel it is an enviable objective.

I can show you what we have done. It is pretty straightforward. Essentially, we have 4-band TM imagery from 1986 and 1991. We have got bands 3, 4, 5, and 6 and they have all been resampled to 25 m. In this case, band 6 is the essential band for pulling out the clouds.

Tom: Excuse me, is there anyone here not familiar with the bands? Band 6 is the thermal band, (coarser) in spatial resolution. This one is different from the other bands, and it is usually not used for classification purposes.

So, what we have got here is our 1986 image and to that we apply a function, which I am calling the Normalized Difference Cloud Index [NDCI], which is (5-6)/(5+6). We chose that because band 5 seemed to have the best contrast between cloud and cloud shadow. Band 6 seemed to be the best band for pulling out clouds individually. We created an NDCI image for 1986 and for 1991. We take the difference of those two images here, the 1991 NDCI minus the 1986 NDCI. That creates our difference image layer. Then, the twist is sticking that layer back in with the original 4 bands to create a new 5-band image for 1986. I have only shown one branch of this. I guess you could do the same thing for 1991. Typically, we've been having problems with the clouds in 1986, not with 1991, that is why I have left it off here. You stick that back in with the other four original bands. That creates a 5-band image. It is then a matter of thresholding your NDCI image. If it is above a certain threshold value, most likely is cloud, but you will also pick up a lot of other change, such as deforestation or clear fields. Then, you could do the same thing if it is below a certain threshold value. Chances are it is cloud shadows, but you would also pick up water and what not. You now have created a mask and you mask out only those pixels that are above or below your threshold value from your NDCI image, and then do an unsupervised classification, but using your 5-band image. When you do that in Imagine, you create a cluster image. I usually use around 30 clusters. You can simply go through and pick out those that most likely are cloud and those that are cloud shadow. It works pretty well. You do get a few extraneous pixels. For example, sometimes there is poor registration between your two images, so some pixels show up along water bodies. By running a 5 by 5 or a 7 by 7 majority filter over that, since we are dealing with clouds here, it is usually taken care of. You could also do a "clump" and "sieve," but that is pretty computationally intensive. It is much easier to just do the majority filter. We have been pleased with the result. I suppose what you could do is to extract the cloud, and then do a simple translation for the cloud shadow.

Heather: You have got a 5-band file at the end there, so you have got 1986 bands 3, 4, 5, 6, and your difference image?

Yes, plus this difference image.

Which is then basically used to mask?

Yes, well actually the NDCI would be used for the mask for 1986, not the difference image because you could have clouds in both images. That is the cloud screening procedure, which is pretty straightforward and would be simple to apply.

Do you think you could do it without the TM (band 6)?

I don't know.

I asked Kent Hegge and he said that this is the first time anybody has brought it up. They have it, and they could do it. Like Tom suggested, we want to be very careful about how we ask for favors. We can get a pair of thermal bands from the first scene and a pair from Wisconsin's scene, and try it to see if it works. I am not asking you to do too much extra processing.

Tom: Is this because of not having the knowledge of their system? Is this going to be perceived as a monumental request to other data delivery, etc. In which case, it is probably a trick we do not want to play at this point. If it is in our nature, we would certainly want to, but maybe in-between (as Dan suggests) is the concept that we have a couple of testing images. . . But, I guess what you are telling us is that you have got your testing, which has worked.

I guess I would want to see the imagery we are getting to see if clouds are enough of a problem that we need to take a more automated approach. We are proposing just to heads-up digitize the clouds and shadows and pull them out manually. If it is a small enough number of clouds, that is workable. If it is a lot of clouds, we will want to pursue something like this.

Instead of using band 6, we may be able to take one of the others, for example, band 1, which really shows where the clouds are. In fact, band 1 probably gives the biggest view of the extent of the clouds.

Try it that way and then maybe you would not have to earn a favor, and you might still be able to automate it. It is just something to try.

We could provide you with a band 1 and band 2 data full scene if you want to try it and see if it works.

I have got it.

Tom: Would that be a problem to try that out?

No, not at all.

Tom: Why don't we look at that. It sounds like you have already been involved in the dialog about can it be done. The answer is yes, but . . .

It can be done, but we want to be sure that we really need it and why.

Tom: Ok, great. I guess the other thing, similarly from the Minnesota group, is the development of anything that has been developed to expedite the implementation of guided clustering.

What we have done to automate guided clustering is, we have written an EML which will accept your reference data. It would take your reference data as one input, and your second input would simply be your continuous raster image file. It will successively iterate through each of your reference classes, mask out those pixels, and run an unsupervised classification. Right now we have our default set to 10 clusters, and this is something I think you wanted to talk about. Automated, you decide the number of clusters, but right now we set it at 10 clusters. It will then delete the intermediary files. Each intermediate file is the same size as your original, so if you are working with 50-mb images, you don't want to be leaving extra files around. It will do this for all of your classes. In our case, we are working with 15 different classes. That is one step that has been automated. We have been talking about automating another part of this, and that is the signature reduction. I guess I wanted to find out how Wisconsin handled reducing their signature set.

We haven't gotten into it extremely deeply. One approach that I have tried is using the old trick of classifying your training data with the training data. For instance, you have a number of polygons that you know are aspen, and many other polygons in the same file that you know are jack pines, mixed forest, etc. You take your 10 aspen signatures and run those parallel-piped-style on your training data for the aspen class and all other classes. Then from that output classification, you know that the aspen polygon should have been classified as aspen. You look at signature one from your 10 aspen signatures and say you just classified the aspens and it had pulled in other classes like the mixed forest, etc. You can step through each of your 10 aspen signatures then and see if those classified aspen or if they classified other pixels from your training data.

So, after you collected the signatures from all of your information classes into one signature editor, how have you gone about reducing that signature separability?

I haven't really approached that yet, but this first run should actually do that for you because, in its first pass, it does measure each signature set against every other classification type that you are classifying.

Tom: It does and it doesn't though, because you do modify the centers with merging and deletion somewhat. Once you have gone through each of the independent spectral classes within each information class, now you add the composite set of spectral classes for all information classes?

There is bound to be some confusion.

Tom: Yes, there is confusion among those. The other issue is the prospect of having missed some important classes. One consideration might be to cluster the whole image to make sure you haven't missed something. Another suggestion is to cluster the whole image before starting the process so that you are guided in the selection of the polygons when you do the actual training. We had not envisioned that because we were stuck with not having images. We were doing these things and planning groundtruth and so forth without having any image to get started, and now that we are looking at the other prospect, it seems to me we may want to do that last step first. That might assist in the stratification of the field work and guarantee it will fit together. I don't know that we have enough experience with getting to the point of developing those composite signatures. In your work, Jana, you went back through another refinement process where you ended up changing the threshold on the transformed divergence.

Jana: Once I concocted everything into one file, I then bumped up the transformed-divergence value, but I knew that there would be some inclusions and some overlap of classes. I didn't find any real clearcut way to do it other than what I call using my biological and remote-sensing sense of looking at things, with transformed divergence and using the ERDAS quick alarm to parallel-piped, which you have to be very careful about using. I didn't take it as the gospel but I used it as a tool to help me evaluate some of those and to throw out things that I thought were being misclassified based on my training data.

I guess one way that we handled this problem with missing something with the reference dataset is not to use unsupervised classification on the whole scene or even some subsection of the scene, but to make use of thresholding on the classification. To use the signature set that you have generated, run a supervised classification with that signature set, you can generate a distance image. If you generate a distance image, set a threshold value for your distance, subset those pixels, run an unsupervised classification on those, and then add that to your original. Then, you have to assign information classes for those.

Jana: How are you doing that?

You're going back to your aerial photos.

Tom: In this training-set refinement process are two elements of automation. If we can really believe that the JM distance or the transformed divergence is a really good acceptable criterium, there is no reason we couldn't automate the process of seeking a maximum by changing the number of classes.

You could do that right off the bat; it gives you the option of doing a dual-skip factor. You wouldn't even have to do that with a sample and then just have a step defined and then a minimum and a maximum, crank it through, calculate your minimum/maximum mean transformed divergence, pick it from there, go with a full clustering, and then go to the other set.

Tom: It sure would be nice to have software tools to be able to approach this task right now and know what the pieces are. It is just that we don't have the composite. Relative to automation in the training set and the final process of the total composite set, remember that training-set refining program. We might want to dig that back out and take a look at it.

Another aspect to the signature evaluation might be quality control in the groundtruth. If we get fields that are clearly being misclassified, we could be looking at patterns. That is another value of doing the kind of systematic evaluation Dave was talking about. We may find that some of our groundtruth in our reference data is incorrect.

References

Bauer, M. E., T. E. Burk, A. R. Ek, P. R. Coppin, S. D. Lime, T. A. Walsh, D. K. Walters, W. Befort, and D. F. Heinzen. 1994. Pages 287-298 *in* Satellite Inventory of Minnesota Forest Resources. Photogrammetric Engineering and Remote Sensing, Vol. 60, No.3.

Wolter, P. T., D. J. Mladenoff, G.E. Host and T.R. Crow. 1995. Pages 1129-1143 *in* Improved Forest Classification in the Northern Lake States Using Multi-Temporal Landsat Imagery. Photogrammetric Engineering and Remote Sensing, Vol. 61, No.9.

Scott, J. M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T. C. Edwards, Jr., J. Ulliman, and R. G. Wright. 1993. Gap analysis: A geographic approach to protection of biological diversity. Wildlife Monograph No. 123.

Tou, Julius T. and Rafael C. Gonzalez. 1974. Pattern Recognition Principles. Addison-Wesley Publishing Company. Reading, Massachusetts

Glossary of Acronyms and Abbreviations

ASCS	Agricultural Stabilization and Conservation Service
CBS	County Biological Survey
CCAP	Coastal Change Analysis Program
CRM	Coordinated Resource Management
CSA	Cooperative Stand Assessment
DLG	Digital Line Graph
DNR	Department of Natural Resources
DOQ	Digital Ortho Quads
ECS	Ecological Classification System
EDC	EROS Data Center
ELT	Ecological Land Type
EMAP	Environmental Monitoring and Assessment Program
EML	ERDAS Macro Language
EMTC	Environmental Management Technical Center
EOSAT	Earth Observation Satellite
EPA	Environmental Protection Agency
EROS	Earth Resources Observation System
ES	Ecological Service
ESA	Ecological Society of America
FIA	Forest Inventory and Analysis
GIS	Geographic Information System
GLNPO	Great Lakes National Program Office
IR	Infrared
ISODATA	Iterative Self-Organizing Data Analysis Technique
JM	Jeffries-Matusita
LULC	Land Use and Land Cover Digital Data
MIRIS	Michigan Reource Information System
MORAP	Missouri Resource Assessment Partnership
MRLC	Multi Resolution Land Characteristics
MSS	Landsat Multispectral Scanner
NAFTA	North American Free Trade Agreement
NALC	North American Landscape Characteristics
NAPP	National Aerial Photography Program
NAWQA	National Water Quality Assessment Program
NBII	National Biological Inventory Infrastructure
NBS	National Biological Service
NCSA	National Center for Supercomputing Applications

NDCI	Normalized Difference Cloud Index
NOAA	National Oceanic and Atmospheric Administration?
NRCS	Natural Resource Conservation Service, formerly SCS (Soil Conservation Service)
NRRI	Natural Resource Research Institute
NWI	National Wetlands Inventory
QQQ	quad quarter-quad
SCCU	Spectrally Consistent Classification Unit
SCS	Soil Conservation Service
STATSGO	State Soil Geographic data
TIGER	Topographically Integrated Geographic Encoding and Referencing
TM	Thematic Mapper
TNC	The Nature Conservancy
UMGAP	Upper Midwest GAP Analysis Program
UNESCO	United Nations Educational, Scientific and Cultural Organization
UP	Upper Peninsula
URL	Universal Resource Locator
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey
UTM	Universal Transverse Mercator
WISCLAND	Wisconsin Initiative for Statewide Cooperation on Landscape Analysis and Data
WWI	Wisconsin Wetland Inventory

Appendix A

Agenda
Upper Midwest Gap Analysis Meeting

January 18-19, 1995 - Madison, Wisconsin

Wednesday, January 18, 1995

8:00 Welcome, Frank D'Erchia, Environmental Management Technical Center

8:30 Upper Midwest GAP Database Development, Daniel Fitzpatrick, Environmental Management Technical Center

9:00 Land Cover Classification System Development, Don Faber-Langendoen, The Nature Conservancy

9:45 Break

10:15 Status of the Gap Analysis Program, Mike Jennings, National Biological Service

10:45 General Overview of Draft Image Processing Protocol, Tom Lillesand, University of Wisconsin-Madison

11:45 Lunch

1:00 Wisconsin Report, Paul Tessar and Bob Goldmann, Wisconsin Department of Natural Resources

1:30 Minnesota Report, Dave Hienzen, William Befort, and Ken Boss, Minnesota Department of Natural Resources

2:00 Michigan Report, Dale Rabe and Mike Donovan, Michigan Department of Natural Resources and Pete Joria, Environmental Management Technical Center

2:30 Break

3:00 Illinois Report, Warren Brigham, Mark Joselyn and Don Luman, Illinois Natural History Survey

3:30 Indiana, Missouri, Iowa, and Ohio Report, Forest Clark, Bloomington Indiana Field Office, U.S. Fish and Wildlife Service

4:00 Open Discussion

6:00 Social

Thursday, January 19, 1995

8:00 Review

8:30 Preclassification Stratification Research Results, Jana Steward, U.S. Geological Survey

9:00 Wetlands Classification Research Results, Dave Nagel, Wisconsin Department of Natural Resources and the University of Wisconsin-Madison

9:30 Landsat Thematic Mapper Data Status, Daniel Fitzpatrick, Environmental Management Technical Center

10:00 Break

10:30 Overview of Wisconsin Department of Natural Resources Geo Services Section, Paul Tessar, Wisconsin Department of Natural Resources

10:45 Depart for Tour

11:00 Tour, Wisconsin Department of Natural Resources Geo Services Section

12:00 Lunch

1:00 Continuation of Group Discussion Image Processing Protocol, Tom Lillesand, University of Wisconsin-Madison

3:00 Break

3:30 Open Discussion and Future Plans

4:30 Meeting Wrap Up

Appendix B

Participant List
Upper Midwest Gap Analysis Meeting

January 18-19, 1995 - Madison, Wisconsin

Norm Aaseng
Minnesota DNR
Wildlife Division
500 Lafayette Road
St. Paul, MN 55155

David Aslesen
US Army, Fort McCoy
AFRC-FM-PWN
2160 S. J. Street
Fort McCoy, WI, 54656-5162

William Befort
Minnesota DNR
Resource Assessment Unit
2002 Airpot Road
Grand Rapids, MN, 55744

Susan Berta
Indiana State University
Dept of Geography, Geology, Anthropology
Indiana State University
Terre Haute, IN, 47809

Matthew Bobo
University of Wisconsin
125 N. Hancock Apt #1
Madison, WI, 53703

Ken Boss
Minnesota - DNR
Resource Assessment Unit
2002 Airport Road
Grand Rapids, MN, 55744

Warren Brigham
Illinois Natural History Survey
607 East Peabody Drive
Champaign, IL, 61802

Jonathan Chipman
University of Wisconsin
1225 W. Dayton Street, Rm 1219
Madison, WI, 53706

Forest Clark
USFWS Indiana GAP Analysis
620 S. Walker Street
Bloomington, IN, 47403

Robert Clark
Missouri - DNR
P.O. Box 176
205 Jefferson Street
Jefferson City, MO, 65102

Bob Clevenstine
U.S. Fish & Wildlife Service
Rock Island Field Office
Rock Island, IL, 61201

Mike Donovan
Michigan - DNR
Wildlife Div.
P.O. Box 30444
Lansing, MI, 48909-7944

Don Faber-Langendoen
The Nature Conservancy
1313 Fifth Street SE, Rm 314
Minneapolis, MN, 55414

Don Field
NOAA/National Marine Fisheries Service
Beaufort Laboratory
101 Pivers Island Road
Beaufort, NC, 28516-9722

Participant List
Upper Midwest Gap Analysis Meeting

James Giglierano
Iowa - DNR
109 Trowbridge Hall
Iowa City, IA, 52242-1319

Bob Goldmann
Wisconsin - DNR
GEO Services
P.O. Box 7921
101 S. Webster Street
Madison, WI, 53707

Mark Hansen
USDA - Forest Service
North Central Forest Experiment Station
1992 Folwell Ave
St Paul, MN, 55108

Pamela Haverland
National Biological Service
Midwest Science Center
4200 New Haven
Columbia, MO, 65201

Todd Heibel
Missouri - DNR
P.O. Box 176
205 Jefferson Street
Jefferson City, MO, 65102

Dave Heinzen
Minnesota - DNR
Resource Assessment Unit
2002 Airport Road
Grand Rapids, MN, 55744

Brian Huberty
USDA - NRCS
2820 Walton Commons west, Suite 123
Madison, WI, 53704-6785

Mike Jennings
National Biological Service
Idaho Cooperative F & W Research
University of Idaho
Moscow, ID, 83843

Mark Joselyn
Illinois Natural History Survey
607 East Peabody Drive
Champaign, I1, 61802

William Kane
Indiana State University
Dept Geography, Geology, Anthropology
Indiana State University
Terre Haute, IN, 47809

Yung-Tsung Kang
Michigan State University
Center for Remote Sensing
115 Manly Miles Bldg
East Lansing, MI, 48823

Steven Koch
Indiana University
School of Public & Environmental Affairs
10th & Fee Lane, Room 412
Bloomington, IN, 47405

Tom Kompare
I1. Natural History Survey
607 E Peobody Drive
Champaign, IL, 61802

Tom Lillesand
Environmental Remote Sensing Center
University of Wisconsin-Madison
1225 W. Dayton St.
Madison, WI, 53706

Participant List
Upper Midwest Gap Analysis Meeting

Stephen Lime
University Minnesota
110 Green Hall
1530 N. Cleveland Ave
St. Paul, MN, 55108

Don Luman
Illinois Natural History Survey
607 East Peabody
Champaign, I1, 61820

Ron McCormick
Wisconsin - DNR
3144 Vilas Road
Cottage Grove, WI, 53527

Stephen Mighton
USDA Forest Service
310 W. Wisconsin Ave Rm 500
Milwaukee, WI, 53203

Dave Nagel
Wisconsin - DNR
Geo Services
P.O. Box 7921
101 S. Webster Street
Madison, WI, 53707

Tim Nigh
Missouri Dept Of Conservation
Planning Division
P.O. Box 180
Jefferson City, MO, 65102

John Probst
US Forest Service Research
NCFES
5985 Hwy K
Rhinelander, WI, 54301

Dale Rabe
Michigan - DNR
Wildlife div.
P.O. Box 30444
Lansing, MI, 48909-7944

Anna Radue
Indiana University
School of Public & Environmental Affairs
10th & Fee Lane, room 412
Bloomington, IN, 47405

Heather Reese
Wisconsin - DNR
GEO Services
P.O. Box 7921
101 S. Webster Street
Madison, WI, 53707

Kelly Roberts
Indiana State University
Dept Geography, Geology, Anthropology
Indiana State University
Terre Haute, IN, 47809

Tom Ruzycki
Wisconsin - DNR
GEOServices Div.
P.O. Box 7921
101 S. Webster Street
Madison, WI, 53707

Tom Simmons
Wisconsin - DNR
P.O. Box 7921
101 S. Webster Street
Madison, WI, 53707

Jana Steward
U.S. Geological Survey
Water Resource Division
6417 Normandy Lane
Madison, WI, 53719

Steven Westin
Missouri Dept of Conservation
Foresty Division
P.O. Box 180
Jefferson City, MO, 65102

Participant List
Upper Midwest Gap Analysis Meeting

Leni Wilsmann
MIDNR - Natural Features Inventory
5th Floor Mason Bldg
P.0. Box 30444
Lansing, MI, 48909-7944

Jeffery Wilson
Indiana State University
Dept Geography, Geology, Anthropology
Indiana State University
Terre Haute, IN, 47809

Phil Strobel
US - EPA
Great Lakes Nat'l Program Office
77 W. Jackson Blvd (G-9J)
Chicago, IL, 60604

Paul Tessar
Wisconsin - DNR
GEO Services
P.O. Box 7921
101 S. Webster Street
Madison, WI, 53707

Dalia Varanka
USDOI - BLM
Suite 225
310 W. Wisconsin Ave.
Milwaukee, WI, 53203

Peter Weiler
University of Wisconsin
Environmental Remote Sensing Cntr
1225 W. Dayton Street
Madison, WI, 53705

Tim Weiss
Wisconsin - DNR
WM/4
Box 7921
Madison, WI, 53707

Dan Wendt
USDA - Forest Service
North Central Forest Experiment Station
1992 Folwell Ave
St Paul, MN, 55108

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE April 1996	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE Proceedings Upper Midwest Gap Analysis Meeting and Workshop, January 18-19, 1995.	5. FUNDING NUMBERS
6. AUTHOR(S) F.. D'Erchia, D. Fitzpatrick, and P. Joria (editors)	
7. PERFORMING ORGANIZATION NAME AND ADDRESS	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Biological Service Environmental Management Technical Center 575 Lester Avenue Onalaska, Wisconsin 54650	10. SPONSORING/MONITORING AGENCY REPORT NUMBER 96-G001

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Release unlimited. Available from National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161 (1-800-553-6847 or 703-487-4650)	

13. ABSTRACT (Maximum 200 words)

The Environmental Management Technical Center hosted the Upper Midwest Gap Analysis Program meeting as a follow-up to implementation of the Program, which is being coordinated by the Environmental Management Technical Center. Status reports were presented by the actively participating States of Michigan, Minnesota, and Wisconsin. Representatives from the States of Illinois, Indiana, Iowa, Missouri, and Ohio provided reports and participated in regional coordination discussions. Attendees also participated in group discussions about the protocol being designed for processing Landsat Thematic Mapper imagery for land cover classification and the status and distribution of satellite imagery.

14. SUBJECT TERMS Gap Analysis, Landsat, remote sensing, GIS, biodiversity, vegetation communities, land cover	15. NUMBER OF PAGES 66 pp. + Appendixes A-B
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

The Gap Analysis Program (GAP) is a National Biological Service project which is being implemented nationally with the help of more than 400 cooperators, including State and Federal partners, private business corporations, and nonprofit groups. The project seeks to identify the degree to which plant and animal communities are or are not represented in areas being managed for the long-term maintenance of biological resources. The National Biological Service Environmental Management Technical Center facilitates the Upper Midwest GAP, a cooperative effort with the States of Illinois, Michigan, Minnesota, and Wisconsin.