EcoNote 2000-2 **March 2000**

Dealing with Complex Polygon Attributes as PEM Input

by Del Meidinger[1] and Colleen Jones[2]

Introduction

PEM programs use existing digital map data, a GIS (Geographic Information System) program, and software that applies an ecological knowledge base to the data to determine ecosystem units. Combining the digital map overlays through a GIS overlay process results in polygons or pixels and a database that contains all the attributes for each of the combined digital map layers. Ecological relationships between selected attributes and the ecosystem units to be mapped are entered into a knowledge table. Both the database and knowledge tables are then merged together like air and fuel through the venturi of a carburetor. Whereas in an engine the output is the power that drives the machine, in the EcoGen PEM model, the air is the resulting GIS database (EcoPrep), the fuel is the knowledge table, the carburetor is processing software (EcoNGen), and the output power is the resulting map (EcoMap).

The focus of this extension note is a discussion on the use of “complex” attributes, associated with some digital map layers, in PEM. A more complete discussion of specific attributes used in PEM is the subject of another extension note (EcoNote 2000-3).

Input Data Attributes

PEM programs utilize a variety of digital map data. Common data sources include forest cover (or vegetation resources inventory), bioterrain, large-scale biogeoclimatic zonation, and DEM-derived attributes from TRIM data. Other data could include bedrock types from geology maps or satellite image analysis. The polygons delineated on these maps may contain one to many attributes in a linked database. The most obvious attribute of these maps is the map label, but many other attributes can be included in attribute databases in data sources such as forest cover, vegetation resources inventory, or bioterrain.

EcoNotes are produced by the Ecology and Earth Sciences Section, Research Branch, BC Ministry of Forests. EcoNotes are available online at **http://www.for.gov.bc.ca/research/ecogen/furinfo.htm**. Contact Del Meidiger, Research Ecologist at **Del.Meidiger@gems2.gov.bc.ca** for further information.

---

[1] Research Branch, B.C. Ministry of Forests, Victoria, BC
[2] Shamaya Consulting, Smithers, BC

An **attribute** is a single category in the data dictionary for a map attribute database (e.g., basic class or leading species in a forest cover data file). All attributes can have multiple values—a **value** being an acceptable value or class type for the attribute. A description of a map polygon may require combining attributes into attribute groups. An **attribute group** is a set of closely related attributes combined to describe the map entity of interest. For example, biogeoclimatic unit is a combination of four attributes: zone, subzone, variant, and phase. Each attribute can have multiple values and the set of attributes has to be combined in order to describe the biogeoclimatic unit.

Most attributes or attribute groups can be termed **simple**, that is, the value describes the entire polygon. Examples include biogeoclimatic unit or basic class (from forest cover). However, some attributes are combined with a proportion attribute in order to provide a better polygon description. In order to fully describe a polygon, these **complex** attribute groups require the combination of several attributes and their proportion attributes. Examples of complex attribute groups are:

- surficial material, in terrain mapping, where the polygon is described by up to three attributes with associated proportions: terrain component 1, decile for terrain component 1, terrain component 2, decile for terrain component 2, terrain component 3, and decile for terrain component 3.
- tree species, in forest cover, where the polygon is described by up to six species with associated proportions: leading species, leading species percent, second species, sixth species, and sixth species percent.

It is these complex attribute groups that pose interesting challenges when combining attributes in a PEM. Simple attributes can be overlaid with one another and the resulting polygons and pixels easily described. However, an overlay of one complex attribute group over another requires certain assumptions, which may or may not be true.

Use of Complex Attributes

When encountering a complex attribute group, the PEM mapper has to understand certain characteristics of the complex in order to determine how to best deal with it. Are the components of the complex evenly distributed over the polygon (e.g., tree species mix)? Or, are they in some sort of pattern (e.g., rolling ridgetop of bedrock exposures and shallow soils)? Or, alternatively, are they likely in distinct portions of the polygon that cannot be differentiated at the scale of mapping? These characteristics affect the way in which the complex is dealt with.

If the components of a complex group are somewhat evenly distributed over a polygon (evenly distributed component complexes), then the polygon can be split into more than one resultant polygon or pixel by the overlay process and the information content for each resultant can be considered “correct.” Likewise, if the components are in some sort of repeatable pattern over the entire polygon (patterned component complexes), then the use of the attribute data after an overlay does not pose a problem. However, in many cases of complex attribute groups, the components are likely found in certain portions of a polygon that cannot be easily determined by automated methods (composite component complexes).

Using one complex attribute group

When only one complex attribute is being used in a PEM, in combination with many simple attributes, the problem of complex attributes can be easily dealt with. In the case of evenly distributed or patterned complexes, then overlays of any kind can be done. An example is tree species within forest cover polygons. The polygon can be divided several times, for example by slope classes, and the tree species mix (cover type) will still apply to all sub-polygons.

There is only an issue if the complexes are composite component ones (e.g., surficial material). In these cases, indiscriminate overlays of polygons comprised of many simple polygon attributes with the complex ones could result in less accurate assignment of ecosystem units. The PEM mapper would have to assume that the proportions of attribute values in the split polygons are the same as in the original polygon. Using the surficial material example, the division of terrain polygons into sub-polygons during the overlay process requires that the complex of surficial materials be assigned to all sub-polygons. It is very unlikely that all the terrain polygons are uniform in the surficial material complex mapped. In most cases, some portion of the polygon is really one surficial material type, while another portion is another of the types. There is no easy way of determining which part of a polygon has which surficial material.

One alternative is to use the complex attribute polygons as the ecosystem boundaries in the PEM. Then each complex attribute and its proportion would be assessed separately and linked to the predicted ecosystem units. In the case of a complex polygon of 70% Mb and 30% Mv/R, after processing each terrain component separately, the resulting polygon ecosystem label would be 70% mesic ecosystem (/01 site series) and 30% submesic ecosystem (/03 site series). This method works OK if there is enough content from the one complex attribute type to allow for increased accuracy in ecosystems prediction. If, however, other complex attribute types need to be used, then the mapper has to determine how to combine two of more complex attribute groups.

Using multiple complex attribute groups

In using digital maps with complex attribute groups (more than one attribute per category), the mapper needs to address two issues when deciding how to utilize the information. The first is to determine how to use the complexes within each map source, and the second is to determine how to combine two or more maps each having complexities.

This section explores the first issue—using complex attributes within a map source.

Although each complex attribute group provides a reasonable description of the polygon on its own, when used in combination, again, some issues arise that the PEM mapper needs to reconcile. Some inventories (e.g., bioterrain), can have multiple complex groups within one polygon (i.e., surficial material/surface expression, soil drainage, slope class, and aspect class). How can the mapper deal with the various complexes?

In the case of multiple complex attribute groups within one polygon, it is a bigger issue if more than one of the complex attributes is of the composite component type. In bioterrain mapping, this is the case. Imagine a polygon with the following two attributes and values:

- *Surficial material/surface expression: 70% Mb, 30% Mv*
- *Soil drainage: 50% well, 30% moderately well, 20% rapidly*

With these two composite component complex attributes, the result of their combinations is:

- *35% well-Mb, 21% mod well-Mb, 14% rapidly-Mb, 15% well-Mv, 9% mod well-Mv, 6% rapidly-Mv*

As this is a mathematical relationship and not a spatial relationship, the location or likelihood of each of these combinations is unknown. With this in mind, the mapper has quite a few choices:

1. Use all the attributes for the polygon to determine the most likely ecosystem unit.

   Assign each of the attributes relationship values in the knowledge table. In the above example, Mb, Mv, well, moderately well and rapid would all have numeric values assigned to them indicating their relationship to the ecosystem units. When processing the polygon, all the numeric values for each of these would then be added up to determine the most likely ecosystem unit. In this case, only one ecosystem type would be assigned to a complex polygon but it could be difficult to code the knowledge table so that one ecosystem would result when there is such a mix of values for the two attributes.

2. Select one of the complex attribute groups for the polygon as the primary label source and compile the data for the other attribute groups.

   If the choice were soil drainage, as it was felt to have stronger relationships with the ecosystem types of the area, then the attribute dataset for the polygon would be:

   - *Well drained Mb/Mv over 50%, moderately well drained Mb/Mv over 30%, and rapidly drained Mb/Mv over 20%. The result for the polygon would be 50% ecosystem A, 30% ecosystem B, and 20% ecosystem C.*

3. Select the dominant value for each complex attribute group.

   In this case, the attribute description for the polygon would be well drained Mb. The result would be one ecosystem (i.e., the one with the most likely occurrence on well-drained Mb). This option has again simplified the issue but has also greatly simplified the predicted outcome and failed to identify some important differences within the polygon.

4. Use all possible attribute combinations, but select only the top three results.

   For our example polygon, the attribute set would be the six combinations of attribute values. Their potential resulting proportions are: 35% well-Mb, 21% mod well-Mb, 14% rapidly-Mb, 15% well-Mv, 9% mod well-Mv, 6% rapidly-Mv. Using this set of attribute values, it is possible that six ecosystems could be predicted. However, only the top 3 could be selected to represent the polygon, according to the RIC standards.

In this case, 71% of the polygon would be described well. If some of the predictions turn out the same, then the components could be combined. In that case, a reasonably high proportion of the polygon would be described adequately. The problem multiplies though if more than two composite component complex attributes are combined. Also, there is the problem that not all the combinations may even exist and some "important" combinations may not be in the top three. In our example, the Mb and well-drained attribute values dominate the resulting combinations.

5. Average the attribute values within a complex or select one and ignore the rest, in order to create more simple attributes.

   In our example, we could decide to use only surficial material, rather than surficial material and surface expression. Then we would have only 'M' (morainal) material. However, when two different surficial materials occur in a polygon we would have the same complex issue to deal with.

   With an attribute like drainage class, some attribute combinations could be used and others ignored. For example, adjacent drainage classes for a polygon could be considered as a class (i.e., well to rapid, imperfect to poor) in addition to the pure classes (e.g., well, poor). Polygon drainage values that are wide ranging (e.g., well to imperfect), could be ignored as they are too imprecise and the other polygon attributes used to determine the most likely ecosystem.

   On the other hand, the values could be averaged. In the case of drainage, intermediate classes might arise and the new values would have to be considered in the knowledge base. An attribute like surficial material is impossible to average.

6. Pair values from different attributes in order of proportions or in order of coding.

   This method would require linking the value of the first or highest proportion of one attribute complex to the value of the first or highest proportion of another, and then the second or second highest proportion, etc. This can be problematic, however, since there is often an uneven number of attributes coded between two attribute complexes. Further, it is possible that the order of the attribute values is not reflective of their association with the values of other attributes, as that is not a requirement in most inventories. In some cases, the order does not reflect dominance at all, for example, the slope classes in terrain mapping are listed in numerical order.

7. Predetermine the associations between values of attribute complexes that are most likely to occur.

   Create a table or matrix of acceptable ecological associations. For example, which drainage classes are associated with which surficial materials and surface expressions:
   Rapidly drained – Mv, Cv, R, FG
   Well drained – Mb, C, F, Mv, Cv, FG
   Moderately well drained – Mb, C, F
   Imperfectly drained – Mb, O, F

   Then only these 'acceptable ecological' associations would be applied to the polygon. In our example at the start of this section, of the six combinations mathematically possible, only the following four are ecologically acceptable: *well-Mb, mod well-Mb, well-Mv, and rapidly-Mv*. The two ecologically unacceptable combinations are *rapidly-Mb and mod well-Mv*. The proportions of the four acceptable associations could be standardized to 100% and then their most likely ecosystem determined by running through the knowledge base.

Although this will not always reduce the number of complex combinations to deal with, it will likely do so in many cases. This could provide a reasonable number of combinations to deal with in the knowledge tables and provide a reasonable description of most of the variation in a polygon. After processing, the proportions for the combinations that result in the same ecosystem unit could also be combined.

Other options could be combined with this one. One example would be that some combinations could be weighted as being more likely than others. Another example is to combine the attribute proportion relationships with the allowable combinations to try to narrow down the most likely combination(s). More complicated options are more difficult to program, however, the result may be worth the effort.

Any of the alternatives described above is possible to use, depending upon the attributes, their relationships to the ecosystem units to be predicted, and the required accuracy of the ecosystem map. The mapper needs to make these decisions—some will make more sense than others will in a particular mapping project.

Dealing with Complex Attributes during the Overlay Process

Some PEM projects have assumed that the bioterrain polygon linework and associated database should be nearly all that is required to develop a PEM map, since this is the foundation of TEM. However, from the discussion above, it is evident that this is not necessarily the case. Combining data digitally has different challenges from that of air photo interpretation of complex map entities. In a PEM, other digital data sources, like forest cover and attributes modeled from a DEM, are often required to make accurate ecosystem unit determinations.

This section explores some of the issues in combining complex attributes from different digital data sources. We have used the example of the combination of a bioterrain map with a forest cover map. We attempt to provide some options for combining these layers, leaving the decision as to which is best to the reader. The decision in any particular project depends upon project objectives, accuracy of various digital map sources, and the required accuracy of the final product.

The example is a bioterrain polygon with a surficial material and surface expression of 70% Mb – 30% Cv (using deciles, 7 Mb – 3 Cv). There are seven forest cover polygons (whole or partial) that overlap this bioterrain polygon: four are typical density lodgepole pine stands (average Pl), one is a spruce-leading stand (S(Pl)), one is a pine with minor spruce stand (Pl(S)), and one is a dry, open pine stand (open Pl). Together the last three polygons comprise about 40% of the polygon area. The polygon is from the dry plateau country around Princeton where spruce is a strong indicator of subhygric sites. We will only be illustrating the use of these few attributes. Four options for using the forest cover data with bioterrain are presented for this example:

A. Method A: Tally all forest cover type and bioterrain data within the bioterrain polygon. As described above, each attribute would be used individually (or in selected combinations) and assigned a value in the knowledge table as to the likely ecosystem unit that may be found on each feature (attribute). For surficial material and surface expression, the two attributes are combined and assigned relationship values in the knowledge table (e.g., Mb would likely receive a mesic weighting and Cv a submesic weighting). For the forest cover data, species, height and crown closure (if relevant) would be combined and assigned values—the average Pl would receive a mesic to submesic weighting, the Spruce would receive a subhygric weighting, and the short open Pl would receive a dry weighting. See Table 1 for an example of the knowledge table coding for this situation.

Table 1. Example of knowledge table ratings for four forest cover types

	Subxeric	Submesic	Mesic	Subhygric
Open Pl	3	1	0	0
“Average” Pl	1	2	2	0
Pl (S)	0	1	3	2
S (Pl)	0	0	1	3

Table 2. Example of knowledge table ratings for four forest cover types

	Subxeric	Submesic	Mesic	Subhygric
Cv	2	3	1	1
Mb	1	2	3	2

In the example of Method A in Figure 1, seven forest cover polygons with their attributes are presumed to influence the entire polygon and, therefore, each bioterrain type. The average Pl, wet S(Pl), moist Pl(S), and dry open Pl types all influence both the Mb and the Cv. When processed through the ecosystem knowledge table, the presence of such a wide range of forest cover types would result in the forest type have little influence on the outcome. That is, the weightings for each possible forest cover type for each ecosystem when tallied up would be about the same. In our example, however, the mesic ecosystem receives higher weighting (see Table 1) due to the overlap of knowledge base scores. This results in forest cover information weighting the outcome to mesic even with the addition of terrain data. The result is decile 7 for the mesic ecosystem – 3 for a submesic/mesic combination.

B. Method B (See Figure 2): Use the combined attributes as described in Method A, but now include the area of each forest cover polygon as a percent of the whole bioterrain polygon and assign forest cover to terrain types based upon proportion. This requires a mid-way step of sorting out which forest cover types should remain unique and which should be combined into a general type. Since the Spruce and dry Pine are unique and have ecological value in the determining ecosystem types, their information is not combined. As the average Pine is common and is not very differentiating, the information in these polygons is combined. The combined Pine types now constitute the greatest area of the bioterrain polygon, so its data is assigned to the dominant terrain type – the 7 Mb. Each of the Spruce and dry Pine polygons are a minor component of the bioterrain polygon, so each is assigned to the 3 Cv attribute.

The Pl/Mb portion of the polygon (see Table 3) would get the classification of decile 7 mesic ecosystem. The Pl+Pl(S)+S(Pl)/Cv portion ends up being a subhygric ecosystem, (see Table 4) even though there is considerable variation in forest type. The open Pl indicates a dry influence, the Pl(S) a moist influence, and the S(Pl) an even moister influence; however, in combination with Cv, the result is subhygric. The polygon becomes: 7 mesic – 3 subhygric.

Table 3. Example of knowledge table ratings for Mb and "average" Pl

	Subxeric	Submesic	Mesic	Subhygric
Mb	1	2	3	2
"Average" Pl	1	2	2	0
Totals	2	4	5	2

Table 4. Example of knowledge table ratings for Cv and combined forest cover types

	Subxeric	Submesic	Mesic	Subhygric
Cv	2	3	1	1
Open Pl	3	1	0	0
Pl (S)	0	1	3	2
S (Pl)	0	0	1	3
Totals:	5	5	5	6

C. Method C (See Figure 3): Use the combined attributes as described in Method A, and identify unique forest cover polygons as in Method B, but don't calculate percent area. Instead, use these unique forest cover polygons to divide the bioterrain polygon. Now the bioterrain polygon is divided into four sub-polygons all with the same bioterrain label, but now only one forest cover type in each.

After applying the ratings from Tables 1 and 2, the largest polygon of "average" Pl and 7 Mb-3 Cv (see Figure 3) would be labeled decile 7 mesic ecosystem – 3 submesic ecosystem. The Pl(S) polygon would lead to the classification of 7 mesic ecosystem – 3 submesic/mesic ecosystem. The spruce-leading polygon would be classified as entirely (decile 10) subhygric ecosystem. And lastly, the open Pl polygon would lead to a label of 10 subxeric ecosystem.

D. Method D (See Figure 4): Follow Method C for combining attributes and dividing the bioterrain polygon. Rather than applying the forest cover attributes to both terrain types within the resultant polygons, determine which of the terrain types the attributes are most likely associated with.

As described above in Item 7, a matrix of likely combinations can be prepared to eliminate unlikely or nonsense pairings. For our example, the following could be considered "acceptable" terrain types for each of the forest cover types:

Open Pl – Cv, FG
Average Pl – Mv, Mb, Cb, Cv
Pl (S) – Mb, Cb, F
S (Pl) – Mb, F, Cb

Using these "acceptable" combinations, two options are possible:

1. Attach the forest cover type to the terrain type when the matrix indicates that the pairing is "acceptable"; do not pair a forest cover type to a terrain type where the pairing is considered unacceptable. In this option, the combinations resulting from the overlay and the final ecosystem units are shown in Table 5 and Figure 4.
2. Attach the forest cover type to the terrain type where the matrix indicates that the pairing is "acceptable"; eliminate "unacceptable" pairings from consideration. Standardize proportions to 100% of resultant polygon. In this option, the combinations resulting from the overlay and the final ecosystem units are shown in Table 6 and Figure 5.

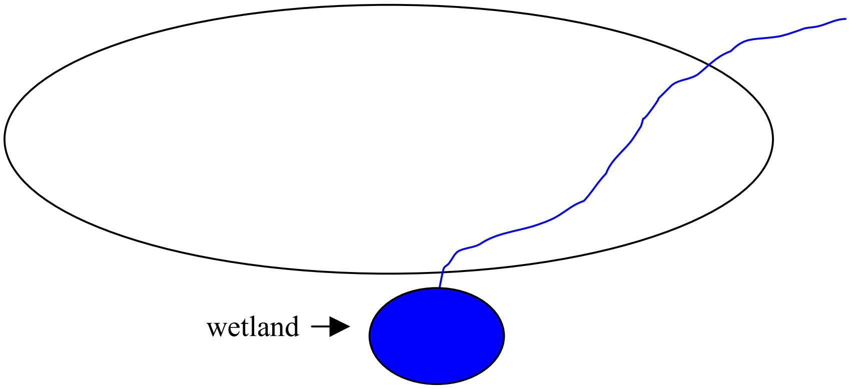
This method may not always reduce the number of options, but has the potential to result in a more accurate ecosystem prediction for the resulting forest cover/terrain polygons.

Table 5. Result of Using Method D – Option 1

Polygon overlay	Result after acceptability criteria applied	Ecosystem classification
Open Pl / 7Mb – 3Cv	7Mb / 3Cv-Open Pl	7 Mesic / 3 Subxeric
Average Pl / 7Mb – 3Cv	7Mb-Avg Pl / 3Cv-Avg Pl	7 Mesic / 3 Submesic
Pl (S) / 7Mb – 3Cv	7Mb-Pl(S) / 3 Cv	7 Mesic / 3 Submesic
S (Pl) / 7Mb – 3Cv	7Mb-S(Pl) / 3 Cv	7 Subhygric / 3 Submesic

Table 6. Result of Using Method D – Option 2

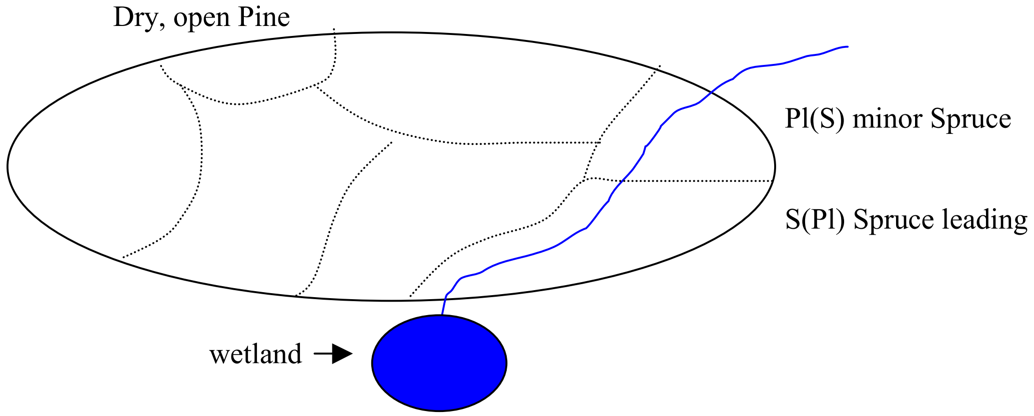
Polygon overlay	Result after acceptability criteria applied	Ecosystem classification
Open Pl / 7Mb – 3Cv	10Cv-Open Pl	10 Subxeric
Average Pl / 7Mb – 3Cv	7Mb-Avg Pl / 3Cv-Avg Pl	7 Mesic / 3 Submesic
Pl (S) / 7Mb – 3Cv	10Mb-Pl(S)	10 Mesic
S (Pl) / 7Mb – 3Cv	10Mb-S(Pl)	10 Subhygric



Method A: 7 Mesic – 3 Submesic/Mesic for whole polygon

Can we pull out where the riparian ecosystems are on this map?
No.

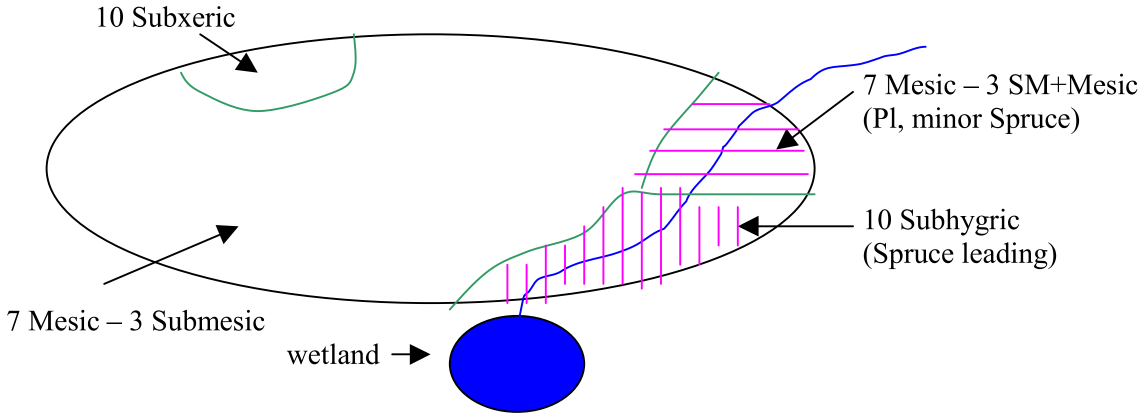
Figure 1. Method A.



Method B: 7 Mesic – 3 Subhygric for whole polygon

Can we pull out where the riparian ecosystems are on this map?
Forest cover polygon lines are not retained, only the F.C. data. Only a query of the original Forest Cover inventory would highlight the location of Spruce-component polygons.

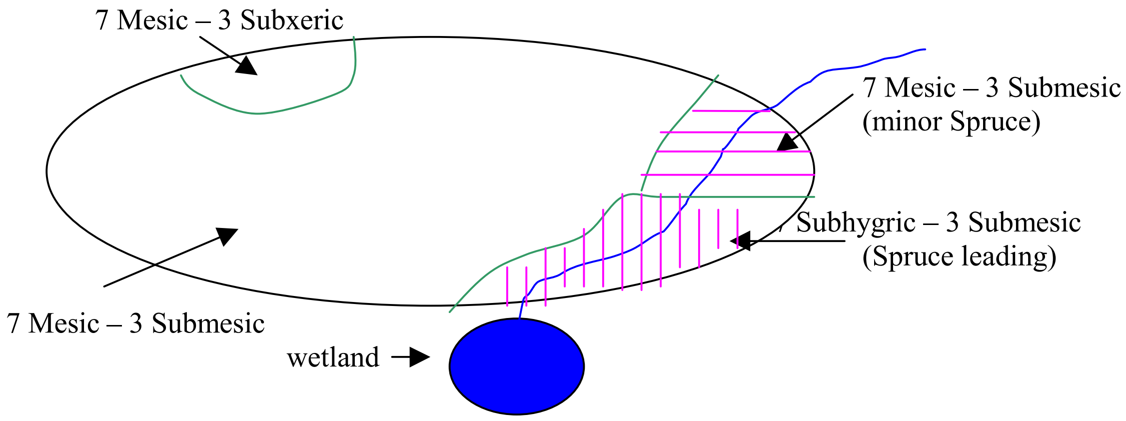
Figure 2. Method B.



Method C
(SM = Submesic)

Can we pull out where the riparian ecosystems are on this map?
Polygons are separate, so we can pull out greater than 50% spruce and display with vertical lines.
Pull out spruce present, but less than 50% spruce and display with horizontal lines.
These polygons are likely to have riparian ecosystems.

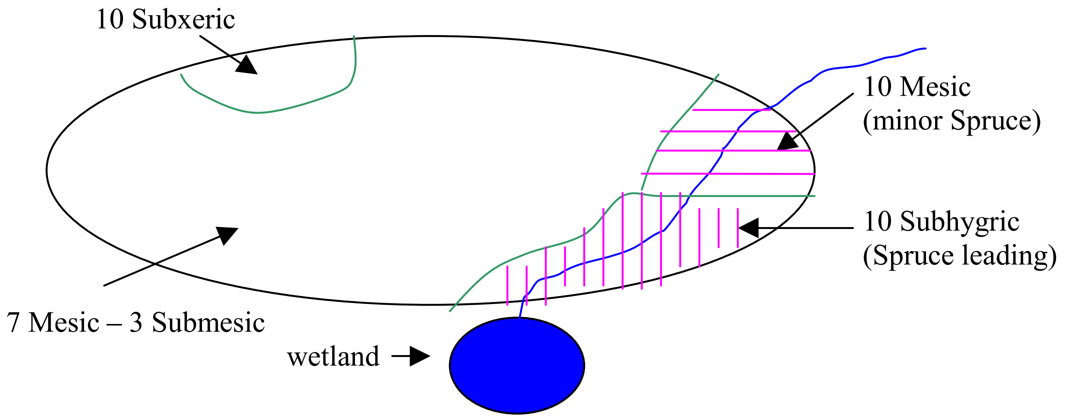
Figure 3: Method C.



Method D – Option #1

Can we pull out where the riparian ecosystems are on this map?
Polygons are separate, so we can pull out greater than 50% spruce and display with vertical lines.
Pull out spruce present, but less than 50% spruce and display with horizontal lines.
These polygons are likely to have riparian ecosystems.

Figure 4: Method D, Option 1



Method D – Option #2

Can we pull out where the riparian ecosystems are on this map?
Polygons are separate, so we can pull out greater than 50% spruce and display with vertical lines.
Pull out spruce present, but less than 50% spruce and display with horizontal lines.
These polygons are likely to have riparian ecosystems.

Figure 4: Method D, Option 2

Summary

Using multiple complex attribute groups requires the mapper to consider the characteristics of each and determine how they can best be combined. Solutions can be as simple as not using some attribute groups to emphasizing one for boundaries and the others for information or splitting polygons, to redesigning an input inventory.

If an attribute or set of attributes cannot add to the accuracy of the ecosystem predictions, due to redundancy, lack of correlation, inaccuracy, or being too complex, then they should not be used. In some cases, an attribute that is inaccurate or too complex in an inventory can be more accurately generated from a DEM—for example, slope and aspect classes in bioterrain.

When an inventory is being conducted for a PEM (e.g., bioterrain), then the issue of multiple complex attribute groups can be dealt with by modifying the design of the inventory. For example, slope can be generated from a DEM and used to create polygons homogeneous in slope class. Soil drainage could be directly linked to surficial material/surface expression so that one set of proportions in a polygon contains all pertinent information. Perhaps an attribute could be added that indicates whether the terrain types in a polygon are found in a predictable pattern throughout the polygon or if the types are confined to certain areas in a polygon. Delineation criteria could be modified to emphasize the importance of polygons that are simple rather than complex, and that have strong predictability for the ecosystems being predicted. The present bioterrain database structure works well for Terrestrial Ecosystem Mapping, but not for PEM.

This note has attempted to provide readers with an understanding of the issues that may be encountered when overlaying digital map data where complex polygons occur. Decisions on how to deal with overlays and the attributes associated with the various polygons have implications on the accuracy of the final map. Hopefully this note will allow PEM mappers to make informed decisions.

Acknowledgements

The authors acknowledge FRBC Research funding for this project, and Research Branch and Prince Rupert Regional Office of the BC Ministry of Forests for supporting EcoGen and the preparation of this extension note.

Reviewers providing comments were Marvin Eng of the BC Ministry of Forests and Dave Clark of the BC Ministry of Environment, Lands and Parks.