Region One
Vegetation Classification, Mapping,
Inventory and Analysis Report

Numbered Report 12-38 **April, 2012**

Mid-level and Base-level Databases of the R1 Existing Vegetation (VMap) Products

Jim Barber
Stephen R. Brown Jr.
Rob Ahl
Geospatial Group - Engineering Staff
USDA Forest Service, Region 1
200 E. Broadway, Missoula, MT 59807

1. Background

The VMap database is published by the Northern Region Geospatial Group, and is designed to be used for Forest and Landscape level analysis. This is consistent with the mission of the VMap program, which is to provide a Mid-level existing vegetation database for the Northern Region that follows the standards set forth in the *Existing Vegetation Classification and Mapping Technical Guide* (Brohman and Bryant, 2005). At the defined scales, spatial errors that are inherent to any remote sensing derived vegetation map are minimized, and the overall estimates of vegetation types match FIA inventory derived estimates. There may be a desire from field units to use these map products at the project, or base-level. At the base-level, however, there is a need for a higher level of spatial and thematic precision and a need to determine the reliability of the product for the specific area of interest.

The VMap database is designed so that it can be used at the project level given some additional refinement that leverages local knowledge of Forest and District specialists. In order to achieve the desired level of precision, and gain confidence in the product for project level use, a thorough understanding of the following elements is necessary:

1. The differences between mid- and base-level VMap database products
2. The R1 Existing Vegetation Classification System
3. VMap accuracy assessment procedures
4. VMap maintenance and update procedures

This document provides an overview of the above elements and references more detailed documents regarding each one. In addition, a number of steps are outlined that must be taken prior to using these databases at scales for which they were not published. These steps are guidelines and actual implementation may vary by project and analysis need.

2. VMap database construction; a short history

Two separate feature classes are delivered with the VMap product; mid-level and base-level polygons. The production of a mid-level database is necessary due to the computational limitations of today’s computers as the number of records and sheer size of the base-level database is such that complex forest-wide analysis can be nearly impossible to accomplish using the base-level database.

A brief description and history of the mid-level and base-level feature classes is presented here to aid in understanding the different uses of each. The basis for producing the VMap database is the image object, or polygon, which is derived via an image segmentation process. Collectively, these image objects are referred to as the “segmentation” and comprise the basic building blocks of the VMap database. The delivered mid-level segmentation is not, however, the segmentation used during the mapping process. It is the base-level segmentation that is the foundation of VMap. The mid-level feature class is derived either from the base-level features via a merging algorithm that is based on a combination of spectral and minimum map unit constraints or via a separate segmentation process, depending on the VMap version. In all cases, the mid-level attributes of dominance type, canopy cover, and tree size are computed from the base-level feature class through zonal majority functions.

The Regional VMap program has been constructing existing vegetation databases since 2000. Over the years the VMap process has evolved significantly because of improvements in technology, skills, and best available science. Although every effort has been made to migrate legacy databases to current versions, some differences in production methodology result in differences in spatial and thematic resolutions of published databases. Version numbers correspond to the year in which the data was published. For example, v11 is a database published in 2011. Although a database is published as later versions, this does not mean that comprehensive image data was used to 're-map' the forest. A brief overview of the three 'vintages' of VMap follows but more information can be found for each National Forest's VMap methodology on the Region 1 VMap web page (http://www.fs.usda.gov/goto/r1/VMap). As of the date of this document (April, 2012), there are two currently supported VMap database versions; v11 and v12.

The first VMap databases were constructed for the seven westside forests (IPNF, CLW, NEZ, KOOT, FLT, LOLO, BIT) and published in 2004. The Flathead National Forest (FNF) has since been re-mapped and is described further in subsequent sections.

Excluding the FNF, the westside databases contain a base-level feature class with no minimum map feature size. The mid-level feature class has a 5 acre minimum map feature. Input imagery was 15 meter spatial resolution and polygon smoothing (i.e., simplify) has only been applied to the mid-level features. Dominance Group 6040 was mapped using the base-level features and then thematically aggregated to produce the mid-level dominance plurality classes.

Four of the Eastside forests (HEL, LNC, GAL, CNF) were mapped using 2005/6 imagery and first published in 2009. Dominance Group 6040 was mapped using the base-level features and then thematically aggregated to produce the mid-level dominance plurality classes. The base-level feature class has a 1 acre minimum map feature and the mid-level feature class was derived from the base by merging and eliminating features less than 5 acres in size.

The Beaverhead-Deerlodge National Forest was mapped using 2009 imagery and first published in 2011. Mid-level dominance groups (DOM_MID_40, DOM_MID_60) were mapped independently and then crosswalked into Dominance Group 6040. Originally, the feature classes were constructed the same as for the other eastside forests described above. However, the version 12 database of the B-D includes a re-segmented mid-level feature class as is described below for the FNF.

The FNF was mapped using 2011 imagery and published in 2012. Mid-level dominance groups (DOM_MID_40, DOM_MID_60) were mapped independently using the base-level feature class and then crosswalked into Dominance Group 6040. The mid-level feature class was not created from the base, as was done previously, but through a separate image segmentation process using larger scale parameters and a classification-based segmentation. This is a new methodology but one that produces much better mid-level vegetation pattern delineation. While the base level does not have an applied minimum map feature, all objects less than 1 acre have been eliminated from the mid level.

3. Mid-level versus Base-level Segmentation and Feature Classes

The transition from base to mid-level feature classes, regardless of methodology used, creates generalizations in both polygon structure and associated map labels. This can result in slightly different totals of vegetation type composition across a Forest.

In the production of the FNF VMap product, a classification-based segmentation routine was used, and this yielded different polygon dimensions for the basic lifeform classes. Specifically, all polygons originally classified as water were merged into unified polygons so that while a lake may have been sub-dived into many small segments, all the lake's small polygons were merged into one unified lake polygon. This alone helps to reduce the overall number of polygons in the final database. Similarly, polygons originally classified as sparse vegetation were ultimately produced using a larger scale parameter than those classified as forest vegetation. Conversely, nonforest polygons were produced with a smaller scale parameter than forest vegetation to capture more small features in the grass and shrub lifeforms. Summary statistics of the various lifeforms and associated polygon dimensions are given below in Table 1.

Table 1. *Summary statistics of base and mid-level polygon configurations associated with nonforest, forest, sparse vegetation, and water lifeforms.*

Lifeform	**Min**	**Mean**	**Max**
Base-Level Grass/Shrub	0.05	4.1	70.4
Base-Level Forest	0.1	7.3	118.4
Base-Level Sparse Vegetation	0.05	3.3	62.3
Base-Level Water	0.1	53.2	122,538.0
Mid-Level Grass/Shrub	0.1	14.0	231.5
Mid-Level Forest	0.1	26.7	343.1
Mid-Level Sparse Vegetation	0.1	7.7	136.1
Mid-Level Water	0.2	77.4	122,581.0

Figure 1 illustrates differences in the overall base and mid-level polygon size distribution, using the FNF database as an example. The histograms of polygon size focus on the Tree lifefom, and generally show that the majority of base-level polygons are between three and eight acres, while most mid-level polygons range between ten and twenty five acres. In fact, the mean polygon size of base and mid-level polygons is 7, and 26 acres, respectively.

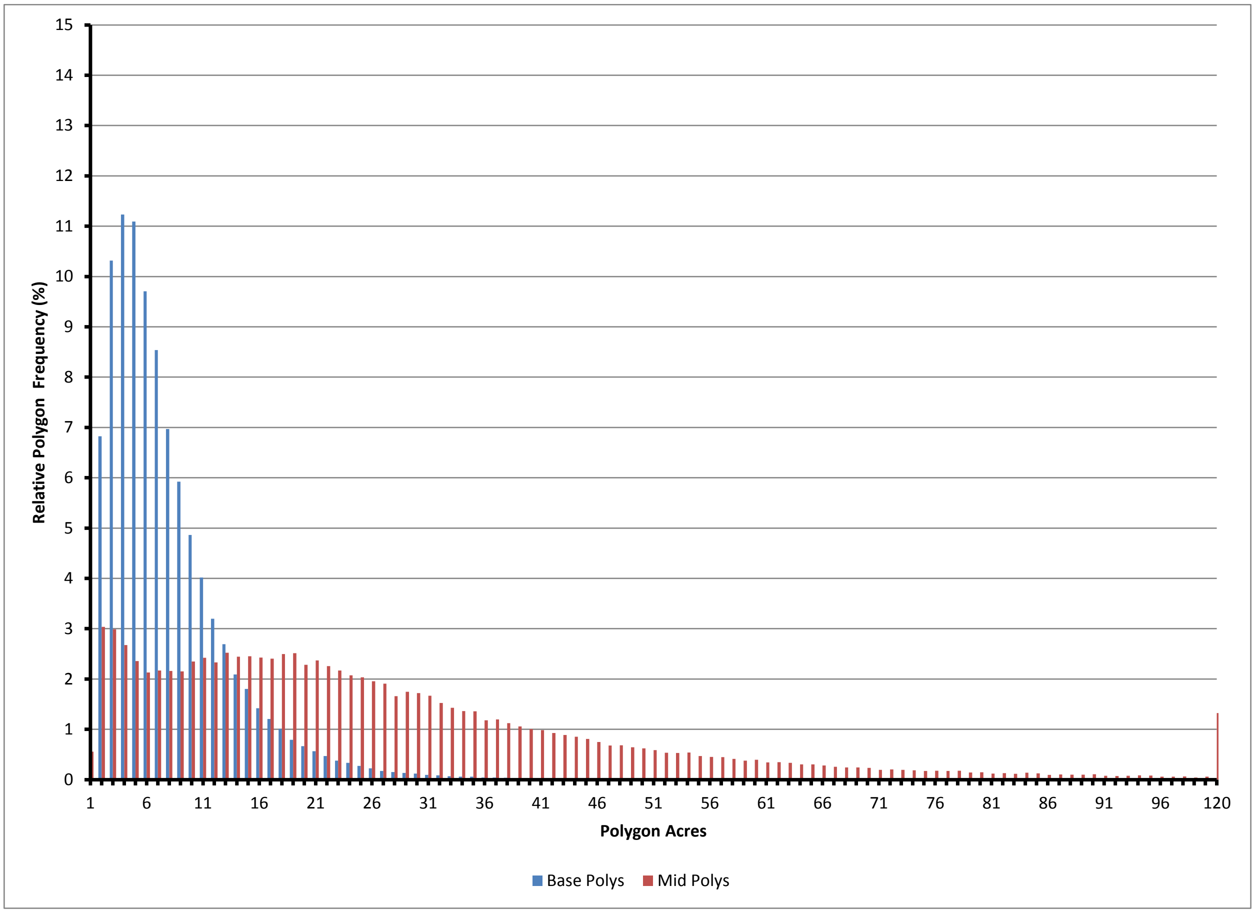

Figure1. Flathead v12 VMap mid and base-level database polygons

The generalization from the base to mid-level may render the mid-level VMap database unsuitable for project level work, or analysis of fine-scaled vegetation patterns, or uncommon types. One common phenomenon associated with the base to mid-level aggregation is that small or rare features tend to be consumed by the matrix of larger or more commonly occurring features. In simple terms, this means that vegetation that occurs in small patches such as aspen, for example, will tend to be represented by fewer acres in the mid-level database than in the base-level database. Tables 2, and 3, respectively show that although there is no statistical difference in the distribution of acres in DOM_MID40 and DOM_MID60 vegetation classes, there is 24% less land mapped as aspen in the mid-level database than in the base-level product. A similar pattern can be observed for many of the less common types, such as grand fir, whitebark pine, and cedar.

Table 2 *t-Test: Two-Sample t-Test Assuming Equal Variances that compares the distribution of acreages for DOM-MID40 vegetation classes in the Flathead Forest base and mid-level VMap*

	Variable 1	*Variable 2*
Mean	189,627	189,625
Variance	57,839,188,982	60,556,051,178
Observations	20	20
Pooled Variance	59,197,620,080	
Hypothesized Mean Difference	-	
df	38.000	
t Stat	0.000	
P(T<=t) one-tail	0.500	
t Critical one-tail	1.686	
P(T<=t) two-tail	1.000	
t Critical two-tail	2.024	

Table 3. *Comparison of base and mid-level acres of DOM_MID40 vegetation class in the Flathead Forest VMap, V12.*

DOM_MID_40	BASE ACRES	MID ACRES	RELATIVE DIFFERENCE (%)
HERB	466,252	467,675	0.3
SHRUB	247,714	232,242	6.2
WATER	195,506	194,939	0.3
SPVEG	216,317	207,706	4.0
MX-PIPO	60,688	57,566	5.1
MX-PSME	785,914	817,123	3.8
MX-ABGR	1,412	1,187	16.0
MX-LAOC	383,444	385,829	0.6
MX-PICO	522,767	529,890	1.3
MX-ABLA	599,736	608,375	1.4
MX-PIEN	226,520	208,922	7.8
MX-PIMO3	103	114	9.8
MX-THPL	2,512	2,315	7.8
MX-PIAL	17,708	14,291	19.3
MX-LALY	476	591	19.4
MX-POPUL	20,241	18,267	9.8
MX-POTR5	1,051	799	24.0
IMIX	17,240	18,770	8.1
TMIX	1,294	1,428	9.4
HMIX	25,640	24,473	4.5

Therefore, it is recommended that the base-level segmentation should be chosen for all project analyses. While the base-level database contains polygons that more precisely depict the vegetation patterns within a project area, the vegetation map labels (specifically the dominance type) are based on a mid-level vegetation classification system and may not contain the level of thematic precision necessary to support the project needs. See the *Region 1 Existing Vegetation Classification System and its Relationship to Inventory Data and the Region 1 Existing Vegetation Map Products* (Barber, et. al, 2011) for a detailed description of the multi-level vegetation classifications used in VMap. Also, the combination of spatial and thematic precision of the various vegetation types becomes of paramount importance and errors that may be acceptable at the mid-level are no longer tolerable at the project-level.

The VMap program produces and publishes a base-level database because end users want a multi-level database that can be used for a variety of different analysis needs at a variety of scales. It is critical to note, however, that the base-level VMap database is not ready to be used for project-level work 'right off the shelf'. It is necessary to perform an evaluation to determine the level of improvement needed to the database to ensure that the analysis needs will be met and managers can have the required level of confidence in the data needed to support decisions.

4. Process Steps for Using the VMap Mid-level Database for Project Planning and Analysis

Editing for project-level work should be should be done on a project by project basis as the data needs may vary by project. The base-level feature class should be clipped from the forest-wide feature class and renamed as appropriate. A three step process should then be undertaken to make the project-level VMap ready for use.

1. Evaluate
2. Edit
3. Validate

4.1 Evaluate

An evaluation of the base-level VMap for project work is performed simply to determine a level of confidence in how well the landscape of interest is characterized for the analysis question(s) at hand. VMap may not satisfactorily capture current vegetation pattern in the project area for a number of reasons. The date of the imagery used to produce the current version of VMap may be of a vintage such that landscape disturbances (e.g., insects, fire, management activities, etc.) have resulted in significant vegetation change since the VMap image acquisition date. The project level VMap database may also not capture current vegetation pattern because of mid-level mapping errors. The accuracy in map labels varies spatially and tends to be more accurate in areas with known training data locations that were used in the mid-level mapping. If there was no training data collected within the project area or in a similar adjacent landscape, the accuracy of the labeling of polygons within the project area may not be as good as those areas that do have training data in closer proximity.

Evaluation is not a measure of the map accuracy but rather an attempt to determine the amount, if any, of edits that need to be completed to make the VMap for the project area ready for use. The evaluation can be either quantitative or qualitative in nature and should be done with the analysis question in mind. For example, if the project and the management decision to be made deals primarily with tree canopy cover (e.g., elk hiding cover), then the evaluation of dominance type may not be necessary or warranted. Evaluation can be done remotely by using comprehensive high resolution imagery such as the National Agricultural Imagery Program (NAIP) data or by field visits to representative areas on the ground. Any evaluation must be done by specialists with a clear understanding of the *Region 1 Existing Vegetation Classification System and its Relationship to Inventory Data and the Region 1 Existing Vegetation Map Products* (Barber, et. al. 2011).

4.2 Edit

Following the evaluation, if it is decided to edit the project VMap database, there are several options available to choose from. These options can broadly be broken into two categories: manual and automated.

Manual editing can simply be described as a specialist manually changing the label of each polygon that is found to be in error. This can be done via field checking of the database or by comparison to high resolution imagery that is suitable to the job. In general NAIP data that is contemporary with the image classification is suitable for making edits on Lifeform, Tree Canopy Cover, and sometimes Dominance Type if the editor is familiar enough with the landscape to make a reliable determination from the NAIP imagery. Tree Size edits are difficult to accomplish via image interpretation. Also, certain of the more rare types, i.e. aspen, can be fairly easily identified using the NAIP imagery and can be accurately labeled in this process. Care must be taken, however, as aspen is easily confused with other deciduous species (e.g., willow, cottonwood) and if the editor is unfamiliar with the landscape in question these other types may be erroneously labeled as aspen.

Automated editing of the database is accomplished through the collection and use of training data specific to the project area. This training data is used to reconstruct the map labeling algorithms so that VMap will more accurately depict the vegetation within the project area. The training data can come from either field collection or image interpretation, similar to the manual editing process. Generally, the automated process is chosen when the Project area is either very large or the VMap labels do not contain the thematic resolution necessary to support management decisions. Re-running of the mapping algorithms might also be necessary if a given type is lacking within the project area. For example, it can be difficult to obtain a sufficient number of whitebark pine samples due to the general inaccessibility of its habitat and the resulting map may under represent the true whitebark pine distribution within the Project area. By collecting whitebark pine samples within a Project area a new map can be built that will more accurately reflect the species' distribution.

Although manual editing will be employed in a project area more often than automated editing, either, or both, of these maintenance options can easily be accomplished within the timelines of a given project as long as they are accounted for in the project planning phase. The Region 1

Geospatial Group is willing to assist in the planning and implementation of either of these maintenance options and will provide the training and support necessary to accomplish these tasks.

4.3 Validate

Validation of the project VMap database involves conducting a quantitative assessment of the map product(s). The Map Class Accuracy Assessment that has been provided with the mid-level database (e.g., *Eastside R1-VMap Accuracy Assessment*, Vanderzanden, et.al., 2010) is suitable and consistent with forest-wide analysis needs. It is not, however, directly transferrable to the project-level and cannot be used to support project-level decisions. An accuracy assessment of the map product(s) within the Project area, regardless of whether the map has been edited, is needed if management decisions will be made based on map data alone.

Map accuracy, however, is not a state variable. It is very important to evaluate the results of any accuracy assessment in the context of the intended analysis application and the management decision the data and analyses are intended to support. This evaluation needs to balance the desired level of precision (i.e., the level of thematic detail) with the desired level of accuracy (i.e., spatial location of a given attribute). For many analyses, detailed thematic classes are aggregated to produce more generalized classes that will typically increase the accuracy of a given map. It is appropriate in these instances to assess the accuracy of the aggregated classes rather than characterize the aggregations with the detailed assessment. It may even be appropriate to aggregate some classes based on the structure of the error, provided that the aggregations meet the analysis objectives. It is also important to determine the level of uncertainty that is acceptable to support a particular management decision.

Quantitative accuracy assessment depends upon the collection of reference data with which to compare the map product in question. It is therefore assumed that the reference data is “truth”, that is 100% correct. Reference data can be obtained via field site visits, photo-interpretation, existing plot data, or a combination of these methods. Statistical validity of the sample, however, is most easily maintained through a random selection of sites which can make the acquisition of reference data both cost and time prohibitive. Following the recommendations of Stehman and Czaplewski (1998), a stratified random sample design should be constructed based on the classes of interest. For example, if an evaluation of canopy cover is desired then five strata should be constructed. One strata for Non-Forest and one strata each for the four canopy cover classes. Thirty samples within each strata should then be selected for evaluation.

The agreement between the evaluated sites and map classes can be displayed in various ways and the most common way to analyze agreement is via the error matrix, where the rows represent mapped classes and columns represent reference classes. Via the error matrix, several useful measures of map accuracy can be computed; including overall, producer, and user metrics. Overall accuracy is a common metric that describes how well the map compares to a reference dataset as a whole. Producer’s accuracy focuses on errors of exclusion and is the probability of a reference site being correctly classified. A high level of Producer’s accuracy indicates that the feature in question will more often be found in the correct class and will not be found mislabeled as another class. User’s accuracy, on the other hand, is based on errors of inclusion and reflects the probability that a feature on the map actually represents that category on the ground.

The Region 1 Geospatial Group is available to help in accuracy assessment design and interpretation for project-level VMap products. However, adequate time must be scheduled within the project timeline for data collection and analysis to accomplish this. The amount of time may be quite variable depending on the specific accuracy assessment requirements and strata to be addressed.

5. Upward Reporting of Base-level Edits to the Mid-level Database

There may be a desire to update the forest-wide mid-level database with edits from project VMap databases. The Region 1 Geospatial Group urges caution in doing this. Although the desire to have the ‘best’ available data for mid-level analysis, one potential problem is that once edits have been made to the mid-level database, a new version is now created and the published accuracy assessment is no longer valid for those mid-level map products. An argument can be made, however, that project edits only increase the accuracy and the published accuracy assessment results are now a minimum. However, it is unclear how this interpretation can be validated. The accuracy assessment and error matrix construction can be re-run following mid-level updates but this is a time consuming process that should only be done on an infrequent basis as needed. A document that explains recommended mid-level VMap maintenance and analytical procedures is in preparation (Barber and Brown, In Prep).

Literature Cited

Barber, J., D. Berglund, R. Bush. 2011. Region 1 Existing Vegetation Classification System and its Relationship to Inventory Data and the Region 1 Existing Vegetation Map Products. USDA Forest Service, Northern Region, CMIA Numbered Report 11-08. http://fsweb.r1.fs.fed.us/forest/inv/classify/r1_ex_veg_6_11.pdf

Barber, J. and S. Brown. In Prep. Maintaining VMap Mid-level Map Databases

Brohman, R. and L. Bryant. 2005. Existing vegetation classification and mapping technical guide. U.S. Department of Agriculture, Forest Service, Washington Office, Ecosystem Management Coordination Staff. http://www.fs.fed.us/emc/rig/documents/integrated_inventory/FS_ExistingVEG_classif_mapping_TG_05.pdf

Stehman, S.V. and Czaplewski, R.L. (1998), Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles. *Remote Sensing of Environment* 64:331-344

Vanderzanden, D., S. Brown, R. Ahl, J. Barber. 2010. Eastside R1-VMap Accuracy Assessment (Lewis and Clark, Helena, Custer and Gallatin National Forests). USDA Forest Service,

Northern Region, CMIA Numbered Report 10-6.
http://www.fs.fed.us/r1/gis/vmap/r1VMap-eastside_assessment.pdf