

Vegetation Resources Inventory

VRI Sample Data Analysis Procedures and Standards

Prepared by
Ministry of Forests, Lands and Natural Resource Operations
Forest Analysis and Inventory Branch

Version 1.0
June 2011

Table of Contents

1	INTRODUCTION	1
1.1	Overview of Vegetation Resources Inventory	1
1.2	The Role of Attribute Analysis	2
1.3	Roles and Responsibilities	3
1.4	Principles of the VRI Sample Data Analysis	3
2	SAMPLE SELECTION	4
2.1	PPSWR sample selection.....	4
2.2	Grid Based Sample Selection	5
3	DATA PREPARATION AND ASSEMBLY	6
3.1	Merging the Phase I and Phase II data.....	6
3.2	Projecting Phase I data to year of ground sampling	7
3.3	Matching height and age data.....	8
3.4	Potential Error Sources and Data Screening	10
4	DATA STRATIFICATION.....	13
4.1	Purpose of Stratification.....	13
4.2	Pre-Stratification	13
4.3	Post-Stratification	13
4.4	Rules on stratification.....	14
4.5	Sample weights	14
5	VRI ANALYSIS ATTRIBUTES	15
5.1	List of Attributes.....	15
5.2	Analysis process	15

6	DERIVATION OF RATIOS	16
6.1	General background	16
6.2	Procedure.....	16
6.3	Summary statistics from analyses	18
7	ANALYSIS REPORT CONTENTS	20
7.1	Executive summary.....	20
7.2	Description of the inventory unit.....	20
7.3	Description of Phase I and Phase II issues.....	20
7.4	Description of data screening process	21
7.5	Presentation of results	21
7.6	Discussion of planned and attained Sampling error.....	21
7.7	Summary and conclusion	21
7.8	Sampling plan.....	21
7.9	Check-list for VRI Analysis deliverables	22
8	REFERENCES	23
9	COMPUTATION OF STATISTICS	25
9.1	Estimation of population total and variance of the population total	25
9.2	Estimation of sampling error and confidence intervals for population total	29
	APPENDIX A.....	24
	APPENDIX B.....	31
10	ANALYSIS EXAMPLES	32
10.1	AGE and HEIGHT	32
10.2	BASAL AREA & TREES PERER HECTARE	34
10.3	LOREY HEIGHT and Decay, waste & breakage volume	35

10.4	Species Composition Comparison	36
	APPENDIX C	37
11	INTERACTIVE ANALYSIS TEMPLATE	38

1 Introduction

Purpose of this Document

The purpose of this document is to define the standard process that is recommended for use in Vegetation resources Inventory (VRI) sample data analysis. The larger part of the document describes the data screening and preparation which are essential to the analysis.

The anticipated audience for this document includes biometricians, forestry professionals, technicians and consultants involved in VRI sample data analysis. A basic understanding of biometrics is required to use the document. It is, however, recommended that analysts who may wish to perform the analysis described herein, consult a qualified professional biometrician to assist in extrapolating the analysis results to the populations that are targeted.

Organization of this Document

This document has nine sections:

- Section 1** Provides a broad overview of the VRI and the role of sample data analysis.
- Section 2** Provides a brief overview of the sample selection process.
- Section 3** Describes the steps in the VRI data preparation and assembly
- Section 4** Describes data stratification and weighting
- Section 5** Describes VRI analysis attributes.
- Section 6** Describes computation of ratios of ground attribute averages to photo interpreted averages
- Section 7** Describes contents of VRI analysis report
- Section 8** Provides standards for the minimum content required for the VRI sample data analysis and documentation.

1.1 Overview of Vegetation Resources Inventory

The Vegetation Resources Inventory is the current standard for vegetation inventories in the Province of British Columbia for all Crown-managed Timber Supply Areas (TSAs) and Tree Farm License (TFLs) lands.

The VRI can be considered as an inventory toolbox that can be used to assess the quantity and quality of BC's timber and other vegetation resources. The VRI provides the means and methods to estimate current overall population totals and averages. This information is used to support:

- sustainable forest management planning;
- day-to-day forest management;
- provincial inventory reporting; and
- strategic land use planning.

The VRI process involves photo estimation (commonly called Phase I), ground sampling (commonly called Phase II), Net Volume Adjustment Factor (NVAF) sampling, and data compilation. VRI sample data analysis is the logical conclusion to the data collection, and it provides feedback to the inventory process indicating the state of accuracy of the photo interpreted attributes.

The key phases are:

Photo Estimation – forest and vegetation characteristics are identified, quantified and mapped on aerial photographs. The photo-based quantification constitutes initial attribute estimates. This information indicates where the vegetation resources are located and how much there is. Phase I estimates could come either from existing forest inventories (pre-VRI) or from new delineation and photo estimation of vegetation characteristics completed to VRI standards.

Ground Sampling – accurate measurement of selected vegetation characteristics, based on a sample. These measurements is used to assess how much of a given vegetation characteristic is present within a given inventory area. This may involve ground measurement of several tree, ecological, and range variables. The sampling design ensures that ground estimates are unbiased. Other ground sampling supporting activities include:

- i) net volume adjustment factor (NVAF) sampling and,
- ii) within-polygon variation (WPV) sampling.

VRI Sample Data Analysis –creating data summaries from the compiled ground sample data, analyzing the relationship between photo estimation and ground sampling data, developing ratios, and extrapolating sample data to make population inferences. The analysis procedures identified in this document are the minimum standard for analysis of VRI ground sample data, collected after 2009. These standards do not apply to data collected before that date.

Other Supporting Activities – include graphic and attribute data storage, database management, growth and yield predictions, and inventory update.

1.2 The Role of VRI Sample Data Analysis

VRI sample data analysis is the process of evaluating the accuracy of either new photo interpretation or existing inventory data using ground sample data as the benchmark for assessment. The objectives of the analysis are:

1. To compute ratios of ground averages of key attributes to the corresponding photo or inventory attribute averages.
2. To obtain overall totals for an inventory unit that are statistically unbiased and respect the biological and dynamic nature of vegetation types, and

3. To assess the state of the inventory in relation to what is indicated on the ground.

1.3 Roles and Responsibilities

The responsibility for completing VRI sample data analysis is now diversified to include a role for licensees, the Ministry of Forests and Range (MFR) and other related agencies. The MFR is responsible for providing the minimum standards and ensuring that required analysis is completed. MFR also provides a mentoring and advisory role to both industry and other government agencies around all phases and products of the VRI. The business components of VRI project planning and data analysis are reviewed and approved by the MFR with concurrence from district, regional and headquarters inventory program staff.

1.4 Principles of the VRI Sample Data Analysis

The photo-estimation and ground sampling processes are intended to address the following two questions:

1. **How Much? - Statistical estimation of overall values.** These values include totals and averages for the three target timber attributes (age, height and volume) for the inventory unit.
2. **Where is it? - Distribution of the totals over the polygons,** which is the process of assigning values to individual polygons. In this analysis protocol, the ground sampling will facilitate the assessment of how much of an attribute is in an area, but it will not be suitable to extrapolate the assessed impact to a polygon level due to small sample size.

In addressing the two questions above, the VRI sample data analysis rests on the following principles:

1. Ground sample data are unbiased and better reflect the overall population totals than the existing photo-estimated attributes.
2. Photo-interpreted attributes are initial estimates that may be improved by ground based adjustments. But attribute adjustment will not be implemented for VRI projects implemented after 2009.
3. During the analysis, the sample data may be grouped by logical strata.
4. Relationships between the ground data and the photo-interpreted information should be used to make inferences on the state of the inventory within specified strata.
5. The sampling error of the total volume or the ratio of total volumes should be used to interpret the risk and uncertainty of the sampling process.
6. The methods used to compute the sample statistics should be consistent with the methods used to select the samples.

2 Sample Selection

Both sampling with Probability Proportional to Size With Replacement (PPSWR) and sampling based on a grid system are valid and acceptable for VRI. The two methods are briefly described in sections 2.1 and 2.2.

The grid-based sampling should be considered as a selection of a point within a polygon because the data collected at a grid intersection will be compared against corresponding information for a polygon within which the grid intersection falls. Ultimately, both sample selection methods produce equal probability samples, where by every sample within a stratum has equal weight.

The implications with regard to sample weighting are that the weights should depend on whether sample allocation to the different strata is proportional or not. In other words weighting is necessary if sample allocation to the different strata in the population is disproportional.

2.1 PPSWR sample selection

The current standard for selecting polygons for a VRI ground sample inventory uses the *probability proportional to size with replacement sampling methodology* (PPSWR). This standard was set for projects occurring during 2001 and beyond.

PPSWR sampling requires that polygons be pre-stratified based on criteria identified in the VRI ground sample project plan (VPIP). The most common stratification criteria are vegetated treed stands stratified by major leading species groups that are represented in the analysis unit yield tables used in the timber supply analysis process.

For the PPSWR sampling design, all stands within a given species stratum are commonly grouped into three or more volume classes. This component of the pre-stratification must be taken into account in subsequent calculations of variance. A sample of polygons is selected from the stands within a volume class with probability proportional to the total area of the class within a stratum. Typically, the samples are allocated proportionally to the leading species strata and volume classes. If required, one or more strata with special business interest may be allocated more samples than are required based on proportional allocation. Such cases should be documented and tracked to ensure that any disproportional allocation of samples is recognized via weighting in the sample data analysis.

Ideally, the strata identified at the time of sample selection should be maintained during data analysis.

Post-stratification of sample observations may be conducted; however, this exercise may complicate the analysis. The MFR should be consulted if post stratification is contemplated.

In the stratified PPSWR methodology, statistical summaries can be generated for each stratum; however, it is recommended that overall statistics (sampling error of the ratio)

are also generated for all strata combined. This is particularly important for expressing the sampling error of an inventory project. Overall population statistics are one of the criteria used in the timber supply analysis process to determine overall risk and uncertainty associated with the estimates of standing volume are useful to determine if the original project objectives (level of precision or sampling error at the 95% level of probability) were achieved.

The pre-stratification used in PPSWR sampling is more efficient than PPSWR sample selection without stratification. Stratification provides stakeholders with known sample sizes by major leading species (or species groups) and significantly simplifies the immediate and any future analysis. However, pre-stratification may have the drawbacks of potentially increasing sample size (and cost) for a given level of precision.

2.2 Grid Based Sample Selection

In the grid-based sample selection process, a standard grid interval (or system) should be specified. This system will generally generate more potential sample locations than are required for an audit type sample within a given management area. A numbered list of grid points, their coordinates and the associated forest cover information from the polygon where a grid point is located should be compiled.

A required number of sample locations that will serve the audit purpose should be selected randomly from the grid list. The forest cover attribute information linked to each grid point can be used to pre- or post-stratify the points. Sample sizes can be allocated to the different strata in the same way as is done for PPSWR.

If no stratification of grid points is undertaken, all grid points in a management area will have equal weight. If stratification occurs, and sample allocation to strata is not proportional, then appropriate weighting of grid point data is required in the computation of overall sample statistics for all strata combined.

The following section focuses on describing the analysis procedures for a VRI ground sample inventory selected using PPSWR (or Grid based) methodology.

3 Data Preparation and Assembly

3.1 Merging the Phase I and Phase II data

Data preparation for the VRI sample data analysis process starts with assembling the data from different sources. In almost all cases, the ground sample (Phase II) data do not reside in the same database as the photo interpreted (Phase I) data. The ground sample data come from the VRI data compilation process and contain a wide variety of attributes that may not be directly used in the sample data analysis process. The photo interpreted inventory attributes may be in one of three different data source formats including:

- new VRI photo-interpretation in Vegetation Inventory File (VIF) format;
- traditional Forest Inventory Planning (FIP) files that have been converted to the new VRI format; or
- “incomplete” or non-standard inventory formats.

Normally, the same version of the Phase I inventory data used for sample selection should also be used for calculating ratios. However, because a time delay often exists between sample selection and ground sampling (usually more than a year), there may be changes in the original inventory data caused by man-made or natural disturbances. Ideally, the inventory data should be projected to the same date as the year of ground sampling. Differences of less than two years are generally not considered an issue.

If significant changes (e.g., new photo estimation or updates) have occurred since the time of the initial sample selection, the impacts on the analysis should be investigated and any issues that arise should be documented. For example, if polygon re-delineation or renumbering has occurred, the polygon in which the Integrated Plot Centre (IPC) is located may need to be confirmed using the UTM coordinates of the IPC. In other cases, recent logging may have occurred in polygons selected for ground sample establishment. If the IPC falls in an area of recent logging that has not yet been captured by the inventory update process, the sample will usually be screened out and excluded from the sample data analysis process.

For the attributes that are targeted for sample data analysis (see section 3.3), it is necessary to merge the ground attributes with corresponding photo interpreted equivalents. The following steps should be followed in assembling this data:

- i) Obtain the ground sample data after compilation.
- ii) Verify if the photo interpretation data (polygon boundaries, areas and attribute estimates) were altered between the time of sample selection and data analysis.
- iii) Where new Phase I data (i.e. since time of sample selection) have become available, identify the target polygons that were sampled on the ground. This is a GIS exercise in which the location of the Integrated Plot Centers (IPCs) are mapped onto the new polygons. The location of the IPC determines the target polygon. If an IPC falls close to a polygon edge, it may be necessary to

identify auxiliary plots that fall outside the target polygon and eliminate them from the compilation.

- iv) If there is no new Phase I data (polygon boundaries and polygon numbering have remained the same), establish if a depletion update occurred after the sample selection process. The more up-to-date data should be used in the analysis rather than the version used at the sample selection stage.
- v) If neither new Phase I nor depletion update occurred after sample selection, then obtain the Phase I data used in the sample selection process.
- vi) Extract a copy of relevant attribute data for the polygons visited on the ground from the population of polygons in the inventory unit in question.
- vii) Project the Phase I attribute data to the year of ground sampling (see Section 4.2 below). Where ground sampling was conducted over multiple years, it is suggested that the Phase I data be projected to the year when the majority of the sampling was done¹.
- viii) Merge the ground data (Phase II) with the corresponding photo interpreted data (Phase I).
- ix) Check to ensure that the Phase I and Phase II data are merged correctly by map-sheet² and polygon number.
- x) Establish the actual number of samples completed and compare that to the planned sample size. Compute sample observation weights (see Section 5.5) using the actual number of samples completed, polygon areas and stratum areas.
- xi) Document where samples were inaccessible and/or replaced, where changes were made to polygon numbers, etc.

3.2 Projecting Phase I data to year of ground sampling

Once the appropriate data have been obtained and the Phase I and Phase II data for the samples have been merged, the Phase I data must be projected to the year of ground sampling before the analysis can begin. If the Phase I inventory is identified as being a “V-standard”, VDYP7 will invoke an internal module called VRISTART to perform the projection. If the inventory is identified as being an “F-standard”, an internal module of VDYP7 called FIPSTART will automatically be run to perform the projection and to

¹ Projection to the year of ground sampling is considered to be a refinement to the analysis process. If an alternative year is chosen for the projection then this date must be documented.

² The recording of map-sheets and polygon numbers in VRIMS has changed over the past 5 years. It may be necessary to merge several data sets before obtaining at the information required to select sample polygons.

generate additional variables that are required to run VDYP7 (i.e., basal area and trees/hectare).

Where ground sampling was conducted over multiple years, it is suggested that the Phase I data be projected to the year when the majority of the sampling was done. This year must be documented. Projecting the individual samples to their specific year of ground sampling is considered a refinement to the analysis process, and is not essential.

3.3 Matching height and age data

During the ground data collection, both height and age are measured on various categories of trees. These categories include:

- a) Top height tree (T trees), measured only at the integrated plot center (IPC).
- b) Site height trees, measured at up to 5 plots in the VRI ground sample cluster (L trees).
- c) Second species site height trees, measured only at the IPC (S trees).
- d) If the “L” tree is not suitable as defined in the ground sampling procedures guide, the second largest diameter tree (within the 5.64m plot) of the same species as the “L” tree is measure and is described as the “O” tree.
- e) If the “O” tree is not suitable as defined in the ground sampling procedures guide, a tree outside the 5.64m plot on the same species as the “L” tree and being approximately the same diameter (DBH) as the “L” tree is measured and is described as the “X” tree.

The compiled height and age data from these T, L, X and O trees are all eligible for use in the analysis of the Phase I age and height attributes. In order to determine the analysis of height and age values in a systematic and consistent manner, a standard process has been developed which allows appropriate matching of Phase I and Phase II species in order to determine appropriate height and age to use in deriving ratios for these attributes. The data matching process is based on a number of assumptions:

1. It is assumed that the Phase II leading species is the correct leading species for the polygon. The use of a single Phase II VRI plot cluster for a polygon represents an extremely small sample that could sometimes be an inadequate sample to identify the polygon leading species, but in general, the ground measurements are felt to be indicative of the true leading species.
2. It is assumed that the relationship (e.g. ratio of means) between the average Phase II leading species height and the averages of both the Phase I leading and second species heights are the same.
3. It is assumed that both Phase I (photo-estimation) and Phase II (ground data) define leading species based on the same reference utilization level (≥ 4.0 cm dbh).
4. It is assumed that the rank 1 layer height and age for Phase I are matched with equivalent ground sample (Phase II) data, even though the ground data are not apportioned by layer.

Five possible cases for determining the most appropriate species match to be used in developing the height and age ratios are as follows:

Case 1. If the Phase II leading species (at 4cm+ dbh utilization) is the same as the Phase I leading species, match the Phase I leading species data with the Phase II leading species data.

Case 2. If the Phase II leading species is the same as the Phase I second species, match the Phase I second species data with the Phase II leading species data.

Case 3. If a match cannot be found using Cases 1 or 2, match the Phase II leading species data with the Phase I leading species data regardless of species, providing both are hardwood or both are conifers.

Case 4. If a match cannot be made based on Case 3, then match the Phase II leading species data with the Phase I second species data regardless of species, providing both are hardwood or both are conifers.

Case 5. If no other case applies, the polygon shall be dropped from the age/height analysis. Case 5 will include the polygons where the Phase II leading species is hardwood and the Phase I leading and second species are both conifer, or vice versa. The data for that polygon is considered incompatible for the analysis process. The age/height attributes for that polygon only shall be dropped from the age/ height analysis. The polygon may still be useful for analysis of other timber attributes such as volume and basal area.

The five cases described above are also explained in the form of a decision tree, which is provided below.

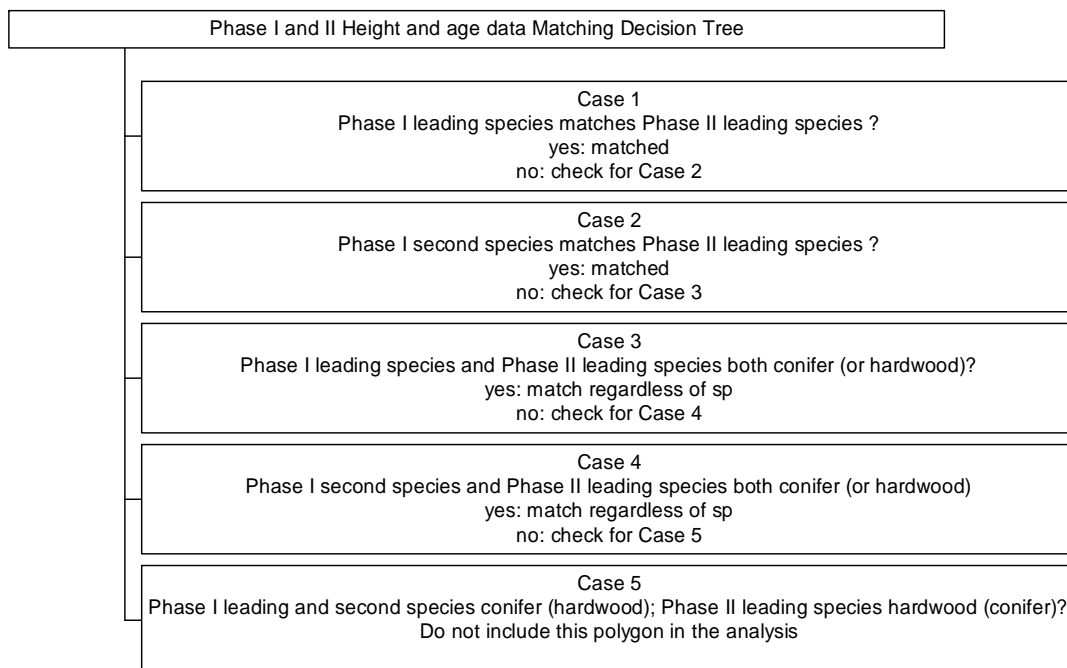


Figure 1: Phase I and II height and age data matching

There are a number of important additional criteria for handling incomplete Phase I or II data³:

- If both age and height data are missing for the matching species, do not include the polygon in the height and age analysis. However these observations should be used in the volume analysis.
- If only one of age or height data (Phase II) is missing, use the data that is present and exclude the data that is missing when computing the averages of age or height. For example, if age is missing for a given polygon but height is available, this polygon would contribute to the computation of the height ratio but it would not contribute to the computation of the age ratio.
- Where leading and second species composition (either Phase I or II) is equal and one or both of the age and height data are missing, the second species may be taken to provide complete data or a match.

Note that if the Phase I second species age/height is selected, care must be taken to ensure that the second species information is projected to the same year as the first species information (i.e. year of ground sampling). Second species data is not automatically projected in the current inventory file projection process and generally must be completed manually.

3.4 Potential Error Sources and Data Screening

All Phase I and Phase II sample records should be examined concurrently (in pairs) for each sample observation. In spite of the rigorous data checking that occurs prior to VRI tree data compilation, additional checking after compilation is essential to provide clean data for analysis. Without this additional effort, there is no guarantee that the results will be error free.

Generally, the data checking process is aimed at detecting and eliminating data entry errors. It also checks for validity of some observations that exhibit outlier characteristics. All attributes should be checked thoroughly prior to final analysis. This sometimes involves the execution of preliminary analyses to check for inconsistencies.

Potential errors may fall into four categories:

- 1) Incorrect or missing Phase I data.

³ For both Phase I and II incomplete age and height data (non-responses) should be avoided as much as possible since these variables are very important in the TSR process. Data collection standards should be reviewed with the aim of minimizing missing data for these variables. This could include extrapolating ages when decay is a problem and/or looking beyond the limits of the main IPC plot for suitable trees. In such cases, the procedures used and the assumptions made must be thoroughly documented.

- 2) Inappropriate matching of Phase I and Phase II data (including incorrect Phase I polygon identification).
- 3) Incorrect ground data (Phase II) compilation outputs.
- 4) Incomplete or missing data.

Following the procedures identified in sections 3.1 (Merging Phase I and Phase II Data) and 3.3 (Matching Height and Age) will significantly reduce the number of potential errors in the analysis. However, specific screening of the main Phase I and Phase II attributes from which ratios are calculated will also help to identify potential issues and inconsistencies with the data.

The data comparisons or data screening is largely a manual process. It involves the examination of individual sample observations to ensure that they are realistic. Graphical analysis is perhaps the most valuable tool in this process. Analysts with access to SAS software will also have handy tools such as “Proc Univar” for checking potential problem observations. Cross-validation can be checked for two or more attributes.

Other software packages such as Excel and MS Access can also be used in data validation. The initial general screening determines if there are obvious records which do not make sense. For instance, a sample record where observed ground height is claimed to be 70 meters should be verified by examining actual field data records. Such values may be the result of data entry errors.

It is more difficult to validate the Phase I records for the entire population. This is due to the large number of records, which may range between 10,000 and 100,000 or more, depending on the size of the inventory unit. Fortunately, the Phase I data capture process includes a quality assurance component, and thus reduces the possibility of encountering excessive data error issues. However, to prevent unexpected results when running the VDYP7 yield model it is recommended that high level data screening⁴ of the Phase I population be done for attributes such as basal area, height, age, trees per ha, and quadratic mean diameter (QMD). If unexpected problems are encountered, the regional inventory coordinator or database coordinator should be contacted to help resolve the issues.

3.4.1 Screening height and age

The objective of the height and age matching procedure outlined in Section 3.3 is to ensure a consistent choice of the most appropriate pairs of Phase I and II height and age data. However, even after following the procedures, it is essential that additional checks on the height and age data be carried out. The focus of such screening is detection of inconsistent values. Graphical analysis is the quickest and most efficient means of screening the height and age data. When inconsistent data is observed, careful investigation of the Phase II plot data and the assumptions made for the data matching should be conducted.

⁴ For example, upper and lower bounds, missing values checks, etc.

Often times, this step may reveal errors in the Phase I and II data merge (ground sample matched with incorrect Phase I polygon), recent logging, incorrect rank/layer for the Phase I data, etc. It may also reveal data entry errors for the Phase II ground sample data or ground sample height or age data that is inappropriate for ratio development⁵.

3.4.2 Screening basal area & trees per hectare

For inventory projects where the photo interpreted information is collected according to the VRI standards, the basal area (BA) and trees per hectare (TPH) attributes are interpreted directly from the photos. For projects based on Forest Inventory Planning files (FIP), BA and TPH may not be available. For these projects, Phase I BA and TPH are automatically generated by running VDYP7, which uses an internal module called FIPSTART to accomplish this.

The main focus of the data screening for these attributes should be the detection of inconsistent values. One of the situations that arise often is the occurrence of records with a species composition, but the indicated BA or TPH is zero or missing. The analyst should check such situations to determine if the zero or missing values are due to the trees in the polygons in question being too young to show any values, or whether the method used to determine them could not provide accurate information for the situation. Findings for such occurrences should be documented.

3.4.3 Screening volume

Volume has no Phase I values interpreted directly from photographs. The Phase I equivalent is derived indirectly throughout various forms of prediction models in VDYP7's internal module called VRISTART. The screening process for volume may point out samples with recent logging or incorrect rank/layer for the Phase I data, etc. Treatment of such samples and/or resolution to associated data issues should be documented.

⁵ Consider the following hypothetical case: The Phase I age for a sample is 120 years but the Phase II age is 700 years. Only one sample tree was collected to produce the Phase II ground sample age. It is suspected that the Phase II age tree may have been a veteran. Closer inspection of the Phase II sample tree data showed that the tree in question had significant rot and that the age had been prorated based on a relatively small counted age. This sample had a significant impact on the value of the ratio for age. After carefully documenting these factors, it was decided to exclude this sample from the derivation of the age ratio.

4 Data Stratification

4.1 Purpose of Stratification

Generally, pre- or post-stratification of VRI inventory data is carried out for several reasons, which include:

- a) assembling polygons with similar forest cover characteristics so as to localize the sample data analysis within a homogeneous grouping. These strata usually tend to be similar to those used in Timber Supply Analysis, but they are not identical.
- b) improving the precision of attribute estimation for the individual strata, and overall totals.
- c) reducing the impact of one individual stratum, on the overall, or specific analysis relevant to other strata.

4.2 Pre-Stratification

In most cases, the target population will have been pre-stratified based on leading species or species groups and then each of these will have been further pre-stratified into three or four volume classes. Both the species and volume class pre-stratification must be accounted for in the computation of the ratio statistics.

4.3 Post-Stratification

Post-stratification occurs after ground data collection. For samples collected using the sorted list approach, this is the avenue for achieving stratification. The advantages of post-stratification include:

- a) No assumptions are made regarding misclassification in the attributes used to stratify prior to sample selection. However, the classification errors are just as disadvantageous in post-stratification as they are in pre-stratification.
- b) The sample selection procedure is easier. But with the advent of advanced computers, this is a non-issue.

The disadvantages of post-stratification include:

- a) The sample selection process fails to anticipate data needs for key inventory information users.
- b) The process does not pre-plan for important strata which may have low representation in the population.
- c) The process may be inconvenient for data analysis, if it is determined that more data is require to augment an important stratum.

4.4 Rules on stratification

It is recommended that pre- or post-stratification be based on leading species and/or logical age breaks. Both of these criteria for stratification have relevance to Timber Supply Analysis.

It is also recommended that the same stratification criteria be applied to all attributes that are being analysed for a given inventory unit. Practices such as using the leading species stratification for height and age analysis, but using BEC for volume analysis are not allowed.

4.5 Sample weights

Many different types of weights can be computed for the VRI sample observations. The weights relevant to polygon selection include:

- Sampling unit weights, which depend on the sampling design and are related to the probability of an individual sampling unit being included in the sample for a VRI project. This weight is the inverse of the probability of selecting a polygon using the PPSWR sampling design. The polygon selection probability is computed as: a_i/A , where a_i = the area of the polygon i , and A is the total area of polygons in the target population being sampled. The sampling weight is: A_i/a_i .
- Weights reflecting the intensity of sampling within a stratum, i.e., the number of hectares represented by one sample within a stratum. These weights are computed as A_h/n_h , where A_h = total area of polygons in stratum h , and n_h = number of sample observations in stratum h .
- Strata proportion weights. These weights are computed as; A_h/A , where A_h = total area of polygons in stratum h , and A = total area of polygons in all strata combined (or target population).

The sampling unit weight and the strata proportion weights are most commonly used (see equations in Appendix A). If samples are selected with different sampling intensities within an analysis stratum, then an additional weight must be applied. A biometrician should be consulted to ensure that the correct weights are applied in such circumstances.

5 VRI ANALYSIS ATTRIBUTES

5.1 List of Attributes

There are 6 timber attributes that are currently considered for VRI sample data analysis.

This list, which includes some of the key attributes for projecting volume, is as follows:

1. Height of the first species
2. Age of the first species
3. Number of trees per hectare at 7.5cm+ dbh utilization level
4. Basal area at 7.5 cm+ dbh utilization level
5. Lorey height at 7.5 cm+ dbh utilization level
6. Volume net top, stump, decay, waste and breakage at 12.5 cm+ dbh utilization level.

5.2 Analysis process

Ratios for the 6 attributes listed above are developed in one cycle. All ratios are calculated after projecting raw photo interpreted data for these attributes to the date of ground sampling.

All analysis data used to calculate attribute ratios is based on the rank 1 layer of the photo-interpreted inventory.

The process is outlined in the steps that follow:

Step 1: Run the Phase I inventory data through VDYP7 to project the inventory to the year of ground sampling. If the Phase I inventory is FIP-based, VDYP7 will automatically invoke the FIPSTART module to generate Phase I values for basal area (BA) and trees/ha (TPH). If the inventory is VIF-based, VDYP7 will automatically invoke the VRISTART module. This step can be considered part of the data preparation stage of the analysis but is reiterated here because of its role in ensuring that Phase I values for BA and TPH are available for all samples.

Step 2: Calculate ratios, on a strata basis, for each of the 6 analysis attributes (see Appendix A for detailed calculations)

Step 3: Produce sample statistics. The key statistics are the sampling errors of the ratios for the different attributes by strata. The overall sampling error for all strata combined should also be computed.

Step 4: Write a brief report. The report is a documentation of the sampling process and the outcome of the sampling. It should include conclusions that explain the implications of the ground assessment of the photo interpreted information on the inventory.

6 DERIVATION OF RATIOS

6.1 General background (PPSWR)

There are many methods of deriving relationships between ground and photo based attributes for the 6 attributes listed in section 5.1. The principles described in this section are the standard for VRI sample data analysis. The approach proposed in this document is based on using a ratio of means estimator as the primary relationship. The ratios are determined based on the relationship of the ground attributes to the inventory attributes. If the data has been pre-, or post-stratified, a ratio will be determined for each stratum. The process of deriving ratios for individual strata within an inventory unit for each attribute can be summarized as follows:

- Compute the estimated strata totals (\hat{Y}_h) based on the sample ground values for each attribute.
- Compute the estimated strata totals (\hat{X}_h) based on the sample photo interpreted values for each attribute.
- Compute the ratios of the estimated ground totals (\hat{Y}_h) to the estimated photo interpreted totals (\hat{X}_h), as follows:

$$R_h = \frac{\hat{Y}_h}{\hat{X}_h} \quad [1]$$

The computation of the ratio and the variance of the ratio is the same for both PPSWR and grid-based sample selection. For either sample selection methodology, however, it is important to recognize sub-strata and/or strata weights appropriately in the computations (see Sections 2 and 4).

6.2 Procedure

The basic procedure to derive the ratios involves plotting the data, computing the properly weighted ratio of means, and using analysis of the residuals to evaluate the ratios, as described below:

1. Plot the data (Phase II – ground vs. Phase I – inventory) to visually analyze the relationship.
2. Calculate the ratio of means based on the PPSWR (or grid-based) sample selection method, as follows:

$$R_h = \frac{\hat{Y}_h}{\hat{X}_h} = \frac{(1/n_h) \sum Y_{hi} / p_{ki}}{(1/n_h) \sum X_{hi} / p_{ki}} = \frac{(1/n_h) [\sum (a_{hi} y_{hi}) / p_{hi}]}{(1/n_h) [\sum (a_{hi} x_{hi}) / p_{hi}]} \quad [2]$$

where:

R_h = Ratio of means for stratum h

p_{hi} = Probability of selecting sample polygon i in stratum h (i.e., $p_{hi} = a_{hi}/A_h$, where a_{hi} = polygon area; A_h = total area for stratum h)⁶

n_h = number of sample observations in stratum h

Y_{hi} = Ground measured (Phase II) value of an attribute y_{hi} times polygon area a_i

X_{hi} = Photo estimated (Phase I) value of an attribute x_{hi} times polygon area a_i

\hat{Y}_h = Ratio-estimated value of an attribute in stratum h using the ratio estimator R_h

$$= R_h \times X_{hi}$$

3. Compute the residuals $(Y_{hi} - \hat{Y}_{hi})$ where Y_{hi} and \hat{Y}_{hi} are the ground measured values and the ratio-estimated values of an attribute, respectively. A plot of the computed residuals against the \hat{Y}_{hi} or the X_{hi} values provides information on how well the ratio defines the relationship between ground and photo interpreted or inventory values of an attribute. For an ideal fit, the residuals should appear as a “band” of points distributed uniformly on both the positive and negative sides of the horizontal zero line. Occasionally, the points will be arranged to form a trend cutting across the zero line. Such occurrences may be symptoms of an inappropriate ratio relationship. A biometrician should be consulted to determine the course of action to take in such circumstances.

If the residual plots are not satisfactory⁷, alternative post stratification or alternative models should be investigated.

6.3 Grid-based ratio computation & analysis

If the establishment of VRI sample locations is based on a system of grid points, and if the distribution of sample size to strata is proportional to strata area within the target population, each sample point has equal weight. In this situation, simple random sampling equations are all that is required to compute sample means, ratios and variances.

If the distribution of sample sizes is not proportional to strata areas, the data from the different strata should be weighted appropriately during the computation of overall (combined strata) means, ratios and variances.

⁶ $p_{hi} = 1/w_{hi}$ where w_{hi} is the weight associated with sample i in stratum h . If samples are selected with different intensities within a stratum, then an additional weight must be applied (see Section 4).

⁷ i.e. indicative of potential bias or lack of fit

6.4 Summary Statistics from Analyses

After completion of the analysis and derivation of the ratios for the six attributes as required in these procedures, the basic statistics generated should be summarized in table form. Such a table will be useful as a reference during the implementation of the VRI sample data analysis. An example of such a summary table is provided in Table 1. Note that the methodology for computing sampling error is provided in Appendix A.

Table 1: An example of summary statistics and ratios that are required for the six baseline attributes that must be included in the current VRI sample data analysis procedures.

<i>Attribute</i>	<i>Stratum</i>	<i>n</i>	<i>Mean weighted Phase II value, by stratum (\bar{Y}_h)</i>	<i>Mean weighted Phase I value, by stratum (\bar{X}_h)</i>	<i>Ratio of means</i>	<i>Sampling error % at 95% confidence level</i>
Height of 1 st sp	Fd	24	21.5	22.5	0.96	10.0%
	P+A	24	22.3	21.9	1.02	8.3%
	S+B	18	26.1	25.7	1.02	11.0%
Age of 1 st sp	Fd	24	133	123	1.08	16.9%
	P+A	25	127	121	1.05	10.7%
	S+B	18	177	173	1.02	16.4%
Trees/ha @7.5cm+ dbh	Fd	27	507	992	0.51	29.0%
	P+A	29	1402	1788	0.78	31.1%
	S+B	21	1165	1770	0.66	48.0%
Basal area/ha @7.5cm+ dbh	Fd	27	26.0	29	0.90	20.9%
	P+A	29	38.9	49.0	0.79	13.6%
	S+B	21	46.0	44.4	1.04	22.3%
Lorey height @7.5cm+ dbh	Fd	27	20.1	20.5	0.98	8.2%
	P+A	29	18.	18.8	0.96	7.5%
	S+B	21	23.1	22.9	1.01	10.0%
Volume/ha net top, stump, decay & waste @12.5cm+ dbh	Fd	27	182.6	187.9	0.97	24.6%
	P+A	29	251.9	262.4	0.96	17.3%
	S+B	21	356.3	354.5	1.01	26.6%

The statistics in Table 1 are generated from the sample. The overall ratio-estimated total volume inferred by the ground sampling must also be computed (see Table 2). In addition, the sampling error for the ratio-estimated population total volume must be computed.

Table 2: An example of ratio-estimated overall total volume, the ratio estimator and sampling error for the total ratio-estimated volume, based on the sample.

<i>Merchantable volume net dwb @ 12.5cm+ dbh</i>	<i>n</i>	<i>Total area (ha)</i>	<i>Overall ratio-estimated total inventory volume (m³)</i>	<i>Overall estimated total VDYP7 inventory volume (m³)</i>	<i>Overall volume Ratio</i>	<i>Sampling error % for total ratio-estimated volume (at 95% confidence level)</i>
<i>Overall</i>	<i>77</i>	<i>1,297,186</i>	<i>323,736,512</i>	<i>321,462,023</i>	<i>1.007</i>	<i>12.53%</i>

7 Analysis Report Contents

The analysis report should provide complete documentation of the analysis procedure and results. Information should be provided to allow the ratio calculation to be replicated, if required. All assumptions and non-standard procedures must be carefully noted. A list of the samples, their Phase I and Phase II attributes, stratum assignment and weights should also be provided. The main sections to be included in the report are highlighted below.

7.1 Executive summary

The executive summary should outline what the objectives of the study were, and indicate whether they were met satisfactorily. It should be a brief description in less than 500 words.

The summary should report the sampling error for net volume (net decay, waste and breakage) at a utilization level that is commonly used in the area where the inventory unit is located.

7.2 Description of the inventory unit

A description of the inventory unit should indicate where it is located within the province of British Columbia. It should ideally describe the Bio-geo-climatic zones that exist in the inventory unit and should also describe physiographic factors that influence growth in the unit.

The specific criteria defining the population of interest for which the VRI sample data analysis will be extrapolated must also be specified.

The objectives of the VRI and some of the key issues influencing the quality of the original inventory could also be described here.

7.3 Description of Phase I and Phase II issues

A description of any new photo interpretation related to the VRI should be provided. In some inventory units a portion of the unit may have new photo interpretation, while the remainder may have older photography. Other units may have older photo interpreted inventories which have been retrofitted. Some units may contain several vintages of photo interpretation. All these photo interpretation situations should be explained. The explanations will provide useful background to the new inventory.

The year(s) when the photo interpretation was acquired is an important reference point for the VRI sample data analysis. It is the year from which all inventory projections are based, and it is an important bench-mark for comparing different projection systems. It should be provided in all analysis reports.

Specific issues related to the Phase II ground sample collection should also be documented in the report. This includes a discussion of sample selection issues (including sample selection pre-stratification), the year(s) in which the ground samples were

established, and non-standard data collection procedures. Samples that were dropped and/or replaced should also be documented in this section.

The year(s) to which the Phase I data is projected, for the purpose of calculating ratios, must also be documented.

7.4 Description of data screening process

A brief description of the data checking and screening should be provided. Any outstanding issues that were discovered during the screening should be explained. If any sample observations are dropped during the data checking process, they should be identified, and an explanation should be provided indicating why it occurred.

7.5 Presentation of results

The presentation of the results is the most important component of the analysis report. This section should explain which analysis methodology was used (i.e. ratio-of-means). If a methodology other than ratio-of-means is used, a rationale should be provided.

This section should contain a number of tables showing various results, which include summary statistics on various attributes of interest, ratios and the sampling error associated with them by strata, and overall statistics.

The results section should contain a discussion on the implication of the analysis in relation to the target population. It should also provide possible explanations for unexpected results.

7.6 Discussion of planned and attained Sampling error

A section on sampling error should be provided. This is the section provides pertinent information on the risks that are associated with key statistics in the inventory. It also provides an indication on how well the planned objectives outlined in the sampling plan are met by the ground data collection process.

The most important sampling errors are those associated with volume, net decay, waste and breakage for the relevant utilization level within the inventory unit. These sampling errors should be provided for all samples combined and by strata.

7.7 Summary and conclusion

The summary and conclusion section put emphasis on the inferences drawn from the VRI sample data analysis exercise. It should also contain recommendations on what further work may need to be done to improve the inventory in the unit in question.

7.8 Sampling plan

A sampling plan is a key reference document. It usually contains details on the ground sampling objectives. It also contains a specification of target sampling error. It should

appear as an appendix to the analysis report. It provides completeness to the VRI. In this format all of the information relevant to the analysis will be located in the same place.

7.9 Check-list for VRI Analysis deliverables

The following is a list of deliverables that should be provided to Forest Analysis and Inventory Branch upon completion of an analysis.

1. Planning documents
2. Original sample list
3. Population list used to select sample
5. Combined Phase I & II data used in calculating ratios
5. Analysis report containing the following:
 - a) a description of the issues raised by stakeholders
 - b) a description of the target population area
 - c) a description of the NVAF data collection & analysis
 - d) a description of the phase II data collection, analysis issues and the analysis estimation.
 - e) a presentation of ratios and their sampling errors
 - f) a statement of the risk and uncertainty associated with volume estimation.
 - g) recommendations for the application of the analysis results
6. A project approval form signed off by the proponents & MFR approvers.

8 References

Iles, K. 1995. A compedium of background papers prepared for the Vegetation Inventory Working Group. Ministry of Forests and Range Forest Analysis and Inventory Branch, Victoria, BC.

Vegetation Inventory Working Group. 1994. A final report of the Vegetation Inventory Working Group. Ministry of Forest and Range Forest Analysis and Inventory Branch, Victoria, BC.

Penner, Margaret. 2009a. The implementation of the Inventory Adjustment Strategy – Kalum TSA. Ministry of Forests and Range Forest Analysis and Inventory Branch, Victoria, BC.

Penner, Margaret. 2009b. Adjustment Issues – Prince George. Ministry of Forests and Range Forest Analysis and Inventory Branch, Victoria, BC.

APPENDIX A

9 COMPUTATION OF STATISTICS

The estimators discussed in this section are consistent with the probability proportional to size with replacement (PPSWR) sampling design, which is the current standard for VRI sample selection. Generally, the unique feature of PPSWR is that the sample observations for any attribute must be weighted by the reciprocal of the polygon inclusion probability during the computation of all statistics. If different sample selection schemes are used, then appropriate formulae should be used.

9.1 Estimation of population total and variance of the population total

Currently, only the variance of the volume ratio is required for purposes of determining the sampling error. It is, however, good practice to compute the variances and standard errors of all continuous attributes and provide them to potential inventory users.

9.1.1 The Separate & Combined Ratio estimators for PPSWR⁸

In this discussion, the sampling unit is the polygon.

9.1.1.1 Notation for the equations

\hat{Y}_{hji}	= $y_{hji} \cdot a_{hji}$ = ground measured attribute times area for polygon i in sub-stratum j of stratum h .
X_{hji}	= $x_{hji} \cdot a_{hji}$ = photo interpreted or unadjusted inventory attribute times area for polygon i in sub-stratum j of stratum h
p_{hji}	= selection probability of polygon i in sub-stratum j of stratum h
a_{hji}	= area of polygon i in sub-stratum j of stratum h
A_{hj}	= total area of sub-stratum j in stratum h
A_h	= total area of stratum h
n_{hj}	= number of sample clusters in sub-stratum j in stratum h
n_h	= number of sample clusters in stratum h
N_j	= number of polygons in sub-stratum j
R_{shj}	= computed ratio for sub-stratum j in stratum h
R_{ch}	= computed (combined) ratio for stratum h
e_{hji}	= residuals computed using either separate or combined ratios
Y	= computed (adjusted) population total for all strata combined

⁸ Equations 4-13 were initially provided by Dr. Bill Warren. Some of the equations were modified by MFR to be consistent with sample selection sub-stratification, which had not been taken into account.

- X = area weighted total of inventory or photo interpreted (unadjusted) attribute for the population
- X_{hj} = photo-based totals of sub-strata (j) within strata (h)
- N_{hj} = number of polygons in sub-stratum j of stratum h
- Y_s = computed population total based on the separate ratio estimator
- Y_{shj} = Adjusted attribute totals for sub-stratum j in stratum h
- X_{shj} = Unadjusted attribute totals for sub-stratum j in stratum h
- Y_{ch} = computed strata totals based the combined estimator at the stratum level
- Y_c = computed population total based on the combined ratio estimator
- J = number of sub-strata in a stratum.
- H = number of strata in sampled population
- $v(\hat{R}_{shj})$ = variances of the estimated ratios at the sub-stratum level
- $v(\hat{R}_{ch})$ = variances of combined ratio estimator at the stratum level
- $v(\hat{Y}_{shj})$ = variances of estimated sub-strata totals
- $v(\hat{Y}_{ch})$ = variances of estimated strata totals
- $v(\hat{Y}_s)$ = sum of variances of sub-strata totals (based on separate ratio estimates)

9.1.1.2 General description of the ratio estimators

For the j^{th} sub-stratum in the h^{th} stratum we take:

$$\hat{R}_{shj} = \frac{\sum_{i=1}^{n_{hj}} \hat{Y}_{hji} / p_{hji}}{\sum_{i=1}^{n_{hj}} X_{hji} / p_{hji}} = \frac{\sum_{i=1}^{n_{hj}} \hat{Y}_{hji} / a_{hji}}{\sum_{i=1}^{n_{hj}} X_{hji} / a_{hji}}, \quad p_{hji} = \frac{a_{hji}}{A_{hj}} \quad [4]$$

The recommended variance estimate is:

$$v(\hat{R}_{shj}) = \frac{1}{X_{hj}^2 n_{hj} (n_{hj} - 1)} \sum_{i=1}^{n_{hj}} \left(\frac{e_{hji}}{p_{hji}} - \frac{1}{n_{hj}} \sum_{i=1}^{n_{hj}} \frac{e_{hji}}{p_{hji}} \right)^2 \quad [5]$$

$$\text{where: } e_{hji} = \hat{Y}_{hji} - \hat{R}_{hjs} X_{hji}, \quad X_{hj} = \sum_{j=1}^{N_{hj}} X_{hji}, \quad j = 1, 2, \dots, k; \quad i = 1, 2, \dots, N_j$$

This variance computation can also be expressed as:

$$v(\hat{R}_{shj}) = \frac{A_{hj}^2}{X_{hj}^2 n_{hj} (n_{hj} - 1)} \sum_{j=1}^{m_j} \sum_{i=1}^{n_{hj}} \left(\frac{e_{hji}}{a_{hji}} - \frac{1}{n_{hj}} \sum_{i=1}^{n_{hj}} \frac{e_{hji}}{a_{hji}} \right)^2 \quad [6a]$$

Under some circumstances where the distribution of samples to the sub-strata is proportional to the relative sub-strata areas within the strata, equation [6a] simplifies⁹ to:

$$v(\hat{R}_{shj}) = \sum_{k=1}^{h_k} \sum_{j=1}^{m_j} \frac{A_{hj}^2}{X_{hj}^2 n_{hj} (n_{hj} - 1)} \sum_{k=1}^{h_k} \sum_{j=1}^{m_j} \sum_{i=1}^{n_{hj}} \left(\frac{e_{hji}}{a_{hji}} \right)^2 \quad [6b]$$

For the **sub-strata totals** we have:

$$\hat{Y}_{shj} = X_{hj} \hat{R}_{shj} \quad [7]$$

$$v(\hat{Y}_{shj}) = X_{hj}^2 v(\hat{R}_{shj}) = \frac{A_{hj}^2}{n_{hj} (n_{hj} - 1)} \sum_{i=1}^{n_{hj}} \left(\frac{e_{hji}}{a_{hji}} - \frac{1}{n_{hj}} \sum_{i=1}^{n_{hj}} \frac{e_{hji}}{a_{hji}} \right)^2$$

4.9.2.2 Separate estimator using sub-strata values

For the total over all strata, the separate estimator is:

$$\hat{Y}_s = \sum_{h=1}^H \sum_{j=1}^J \hat{Y}_{shj} = \sum_{h=1}^H \sum_{j=1}^J X_{hj} \hat{R}_{shj} \quad [8]$$

$$v(\hat{Y}_s) = \sum_{h=1}^H \sum_{j=1}^J X_{hj}^2 v(\hat{R}_{shj})$$

4.9.2.3 Combined estimators at the stratum level

For the combined estimator at the stratum level we have:

$$\hat{R}_{ch} = \frac{\sum_{j=1}^J A_{hj} / n_{hj} \left[\sum_{i=1}^{n_{hj}} \hat{Y}_{hji} / a_{hji} \right]}{\sum_{j=1}^J A_{hj} / n_{hj} \left[\sum_{i=1}^{n_{hj}} X_{hji} / a_{hji} \right]} \quad [9]$$

⁹ This simplification also applies to Equation 5.

The variance of \hat{R}_{ch} is:

$$v(\hat{R}_{ch}) = \frac{1}{X_h^2} \sum_{j=1}^J \left[\frac{A_{hj}^2}{n_{hj}(n_{hj}-1)} \sum_{i=1}^{n_{hj}} \left(\frac{e_{hji}}{a_{hji}} - \frac{1}{n_{hj}} \sum_{i=1}^{n_{hj}} \frac{e_{hji}}{a_{hji}} \right)^2 \right], \quad X_h = \sum_{j=1}^J X_{hj} \quad [10]$$

$$\text{where now: } e_{hji} = \hat{Y}_{hji} - \hat{R}_{ch} X_{hji}$$

In equation [10], the sum $\sum e_{hji}/a_{hji}$, may not necessarily equal 0.0; therefore, it is important to compute the sum and divide it by n_{hj} , and use the result to compute the sum of squares for the residuals. If this is not done, incorrect variances will be obtained for strata.

For the total strata values for the combined ratio estimator, we have:

$$\hat{Y}_{ch} = X_h \hat{R}_{ch}, \quad \text{and} \quad [11]$$

$$v(\hat{Y}_{ch}) = X_h^2 v(\hat{R}_{ch})$$

Due to the expected small sample sizes at the sub-stratum level, it is recommended that the combined estimators described above be used to compute totals and variances at the stratum level.

9.1.1.3 Comments

“...Unless R_h is constant from stratum to stratum, the use of a separate ratio estimate in each stratum is likely to be more precise. This ... assumes, however, that the sample in each stratum is large enough so that the appropriate formula for $v(Y_{hs})$ is valid. With only a small sample in each stratum, the combined estimate is to be recommended, unless there is good evidence to the contrary...The separate estimate is preferable if n_h is large in each stratum and the true ratio is likely to vary from stratum to stratum.” (Cochran, 1968)¹⁰

It is reasonable to expect, and empirical evidence seems to confirm differences between the R_h , but n_h may be relatively small. This suggests a compromise. For example, suppose there were five strata, with two taking up 40% of the area each, with the remaining three covering 20% of the area in total. This suggests using the separate

¹⁰ William G. Cochran. 1963. Sampling Techniques. John Wiley and Sons.

estimator for the two “major” strata and the combined estimator for the three “minor” strata.

If $h = 1, 2$ define the major strata and $h = 3, 4, 5$ the minor;

$$\hat{Y} = X_1 \hat{R}_{1s} + X_2 \hat{R}_{2s} + (X_3 + X_4 + X_5) \hat{R}_c \quad [12]$$

$$\text{where in: } \hat{R}_c, \quad X = X_3 + X_4 + X_5, \quad \text{and} \quad \sum_h^H = \sum_{h=3}^5$$

In this compromise situation, variance of the estimated total is computed as:

$$v(\hat{Y}) = X_1^2 v(\hat{R}_{1s}) + X_2^2 v(\hat{R}_{2s}) + (X_3 + X_4 + X_5)^2 v(\hat{R}_c) \quad [13]$$

(The guidance of a qualified biometrician should be sought in executing these equations.)

9.1.1.4 Pooled variance for all strata

For the VRI, the sampling error standards are not specified for strata. Instead, the standards are specified for the pooled strata. As far as pooled variance computation is concerned, a combination of separate and combined ratio estimators can be used as suggested in equation 13.

In most cases, the variance at the stratum level will be computed using the combined ratio estimators, but for overall sample, the individual strata variances will be summed to obtain sample variance - to be consistent with the separate ratio estimators.

Strata are created with the expectation that the computed strata ratios will differ from stratum to stratum. When this is true, the separate ratio estimators should be used, unless evidence suggests otherwise.

9.1.1.5 Least Squares (LS) and Geometric Mean Regression (GMR) Estimators

The LS and GMR analysis methodologies are not standard. If you wish to use either of them, please contact the MFR first. Each situation will be examined on a case-by-case basis.

9.2 Estimation of sampling error and confidence intervals for population total

Sampling error is the basis for assessing the risk and uncertainty of the estimated total net volume (net top, stump, decay, waste & breakage). The estimated total volume is a point estimate. It is unlikely that this estimate will be exactly equal to the true population value, which is unknown. For this reason, sampling error is used to determine the lower and

upper bounds within which the true population value is expected to occur with a specified level of confidence.

In the VRI, sampling error is determined at the 95% level of confidence, which means the true population total net volume will be within the bounds of the confidence intervals 19 out of 20 times. There is a 1 in 20 chance that it might not.

Currently, sampling error should be in the range of 10 to 15% at the 95% level of probability in order for the adjusted net volumes to be acceptable for timber supply review purposes. Sampling errors larger than 15% may raise flags to the decision makers who use the adjusted net volumes. In such cases, the decision makers may chose to reject the adjusted net volume data, if the they consider the uncertainty in the data is too high for a given management unit.

The decision makers who use the adjusted net volumes should use the sampling error generated from the VRI to conduct sensitivity analysis regarding the total net volumes.

Sampling error is computed as follows:

$$SE\%(\hat{Y}) = t_{\alpha/2, n-1} \frac{\sqrt{v(\hat{Y})}}{\hat{Y}_k} \times 100$$

where α is a stated probability level of a Type I error (usually $\alpha = 0.05$, corresponding to 95% probability level), $t_{\alpha/2}$ is a value from the t -distribution with $n-1$ degrees of freedom, and $n =$ number of VRI samples.

The confidence interval around the population total is expressed as:

$$\hat{Y} \pm SE(\hat{Y})$$

APPENDIX B

10 ANALYSIS EXAMPLES

As part of the VRI sample data analysis simplification, the process will require four mandatory tables for attribute analysis. Table 1 will present the results for the Height and Age analysis. Table 2 will present the Basal area and trees per hectare results. Tables 3 and 4 will present the Lorey Ht/ Volume results and the first species interpretation results, respectively. Samples of the four tables are provided below. The results should be presented by strata, and for all strata combined.

These tables are created after merging photo interpreted and ground, attribute projection to date of ground sampling, height/age matching and screening to eliminate legitimate outliers or bad data (section 3.1 to 3.4.3). To maintain easy access, and documentation, the original ground sample data, merged with the corresponding photo interpreted attributes should be kept in one file. The single file should be linked to the four tables with feature id, and polygon number as link keys.

10.1 AGE and HEIGHT

The sample ground heights and ages should be determined after applying the matching process described in section 3.3. A documentation of the matching option should be provided. The photo interpreted heights should be the estimated heights of the leading species.

Table 1: Height and age data for TSA " XYZ "

Cluster_id	Strata	Sub_Strata	Weight	#_plots	Grnd_Ht	Phto_Ht	Grnd_age	Phto_Age
0031-0001-QO1	All_Other	Imm		5	10.8	7.5	69	63
0031-0002-QO1	All_Other	Imm		5	15.1	15.5	58	63
0031-0004-QO1	All_Other	Mat		3	5.8	14.3	46	102
0031-0005-QO1	All_Other	Mat		5	24.3	31.2	99	182
0031-0006-QO1	All_Other	Mat		5	22.4	26.2	64	218
0031-0007-QO1	All_Other	Mat		5	27.8	26.1	282	212
0031-0009-QO1	All_Other	Mat		5	32.5	22.2	125	122
0031-0010-QO1	All_Other	Mat		5	27.0	25.1	134	202
0031-0011-QO1	All_Other	Mat		5	30.4	30.1	156	212
0031-0012-QO1	All_Other	Mat		5	35.0	30.2	149	142
0031-0013-QO1	All_Other	Mat		5	38.1	32.1	297	243
0031-0015-QO1	All_Other	Old		5	14.9	20.1	125	272
0031-0016-QO1	All_Other	Old		5	29.4	32.1	167	253
0031-0018-QO1	All_Other	Old		5	30.7	32.1	282	272
0031-0019-QO1	All_Other	Old		5	23.3	26.1	284	272
0031-0020-QO1	All_Other	Old		5	33.1	30.1	246	313
Means					25.0	25.1	161.4	196
Ratios					0.9984		0.8218	
SE_ratios								
Sampling errors								

10.2 BASAL AREA & TREES PERER HECTARE

For the “V- and I-standard” inventories, the photo-based estimates of basal area (Ba) and trees per hectare (TPH) should be at the 7.5cm utilization level, and should be the result of projecting the photo interpreted values to date of ground sampling.

For “F-standard” inventories, the photo-based basal area and trees per hectare should be the result of applying FIPSTART after height, age and crown closure and projected to date of ground sampling.

The ground Ba and TPH should be based on the 7.5cm utilization level compilation.

Table 2: Basal area and Trees per hectare data for TSA " XYZ ", at the 7.5cm utilization level

Cluster_id	Strata	Sub_Strata	Weight	#_plots	Grnd_Ba	Photo_Ba	Grnd_TPH	Photo_TPH
0031-0001-QO1	All_Other	Imm		5	17.0	2.2	516	262
0031-0002-QO1	All_Other	Imm		5	14.0	21.1	297	1308
0031-0004-QO1	All_Other	Mat		3		3.3		96
0031-0005-QO1	All_Other	Mat		5	41.4	21.0	1042	125
0031-0006-QO1	All_Other	Mat		5	36.0	29.0	995	284
0031-0007-QO1	All_Other	Mat		5	53.9	44.0	1409	507
0031-0009-QO1	All_Other	Mat		5	73.5	46.4	1490	1087
0031-0010-QO1	All_Other	Mat		5	54.0	50.0	985	1047
0031-0011-QO1	All_Other	Mat		5	46.6	50.0	641	374
0031-0012-QO1	All_Other	Mat		5	36.0	53.2	245	444
0031-0013-QO1	All_Other	Mat		5	64.0	70.0	742	472
0031-0015-QO1	All_Other	Old		5	32.4	10.0	594	115
0031-0016-QO1	All_Other	Old		5	16.0	12.0	57	35
0031-0018-QO1	All_Other	Old		5	73.6	48.0	1083	1030
0031-0019-QO1	All_Other	Old		5	30.6	53.0	1270	575
0031-0020-QO1	All_Other	Old		5	50.4	53.0	260	450
Means					42.6	35.4	775.1	513
Ratios					1.2045		1.5101	
SE_ratios								
Sampling errors								

10.3 LOREY HEIGHT and Decay, waste & breakage volume

The analysis of Lorey height and volume (net decay, waste and breakage) is combined in one table because they are derived from after feeding photo interpreted attributes into VDYP7.

The photo-based data for these attributes should be obtained after projecting the inventory to the date of ground sampling. The corresponding ground data should be derived from ground measurement of tree heights in the case of lorey height; and as a result of applying taper functions to obtain volume.

Table 3: Lorey Height and Decay, waste and breakage net volume data for TSA " XYZ ", at the 12.5cm utilization level

Cluster_id	Strata	Sub_Strata	Weight	#_plots	Grnd_LHt	Phto_LHt	Grnd_V_dwb	Phto_V_dwb
							(???)	
0031-0001-Q01	All_Other	Imm		5	10.8	7.5	91.1	1.9
0031-0002-Q01	All_Other	Imm		5	15.1	15.5	97.0	78.8
0031-0004-Q01	All_Other	Mat		3	5.8	14.3		12.6
0031-0005-Q01	All_Other	Mat		5	24.3	31.2	306.8	145.2
0031-0006-Q01	All_Other	Mat		5	22.4	26.2	301.3	180.0
0031-0007-Q01	All_Other	Mat		5	27.8	26.1	448.8	236.8
0031-0009-Q01	All_Other	Mat		5	32.5	22.2	646.1	293.8
0031-0010-Q01	All_Other	Mat		5	27.0	25.1	427.3	342.2
0031-0011-Q01	All_Other	Mat		5	30.4	30.1	440.3	378.5
0031-0012-Q01	All_Other	Mat		5	35.0	30.2	363.7	384.0
0031-0013-Q01	All_Other	Mat		5	38.1	32.1	793.9	429.0
0031-0015-Q01	All_Other	Old		5	14.9	20.1	231.4	56.9
0031-0016-Q01	All_Other	Old		5	29.4	32.1	253.3	84.9
0031-0018-Q01	All_Other	Old		5	30.7	32.1	737.0	293.9
0031-0019-Q01	All_Other	Old		5	23.3	26.1	168.2	401.5
0031-0020-Q01	All_Other	Old		5	33.1	30.1	623.4	460.7
Means					25.0	25.1	395.3	236
Ratios					0.9984		1.6729	
SE_ratios								
Sampling errors								

10.3 Species Composition Comparison

The photo interpreted species composition should be used directly as interpreted, but the 7.5cm utilization level species composition should be used. The reason for using the 7.5cm utilization level is that VDYP7 uses this utilization level as the benchmark any subsequent volume analysis.

Table 4: Ground and photo species compositions at the 7.5cm (?) utilization level

Cluster_id	Strata	Sub_Strata	Weight	#_plots	Ground Species Comp	Photo Species Comp
0031-0001-Q01	All_Other	Imm		5	Hw 47 Bl 40 Ba 07 S 06	Bl 60 Hw 30 Sx 10
0031-0002-Q01	All_Other	Imm		5	Bl 50 Pl 36 Hw 07 S 07	Pl 60 Bl 20 Sx 12 Hw 8
0031-0004-Q01	All_Other	Mat		3		Bl 100
0031-0005-Q01	All_Other	Mat		5	Bl 65 Pl 15 Ac 15 S 05	Sx 40 Bl 20 Ac 20 At 20
0031-0006-Q01	All_Other	Mat		5	S 64 Ac 21 Bl 15	Bl 50 Sx 30 At 20
0031-0007-Q01	All_Other	Mat		5	Hw 60 Ba 20 Bl 15 S 05	Hw 80 Bl 20
0031-0009-Q01	All_Other	Mat		5	S 45 Bl 31 Hw 21 Pl 03	Bl 40 Pl 20 Sx 20 Hw 20
0031-0010-Q01	All_Other	Mat		5	Hw 52 Cw 38 S 07 Bl 03	Hw 40 Bl 30 Sx 30
0031-0011-Q01	All_Other	Mat		5	Hw 83 S 11 Bl 06	Sx 40 Hw 40 Bl 20
0031-0012-Q01	All_Other	Mat		5	Ac 70 S 10 Bl 10 Pl 05 Ep 05	Ac 45 Pl 20 Sx 15 At 10
0031-0013-Q01	All_Other	Mat		5	Hw 90 Ba 10	Hw 80 Bl 20
0031-0015-Q01	All_Other	Old		5	Hm 67 Bl 33	Bl 70 Hm 30 Ba 10
0031-0016-Q01	All_Other	Old		5	Bl 88 Ac 06 Hw 06	Sx 80 Bl 20
0031-0018-Q01	All_Other	Old		5	Ba 73 Bl 14 Hw 09 S 04	Ba 50 Hw 30 Bl 20
0031-0019-Q01	All_Other	Old		5	Bl 54 S 23 Hm 23	Bl 60 Hw 30 Ba 10
0031-0020-Q01	All_Other	Old		5	Bl 90 Ba 10	Bl 60 Ba 30 Hw 10

Leading species was identified correctly in 7 of 16 samples, or 44% correct

APPENDIX C

11 INTERACTIVE ANALYSIS TEMPLATE

[Macro Workbook](#)

[Sample Data](#)

How to use the example data and the Macro

The Macro is custom-made for VRI/Audit data analysis involving several attributes, which are, Height, Age, Basal area at 7.5cm utilization, Trees per hectare at 7.5cm utilization level, Lorey height, and volume net decay waste, breakage, and species composition.

Below are 4 steps you need to follow to test-run the macro before you use it with your own data.

1. The provided example data and the analysis workbook should be saved in the same directory on your computer.
2. In the analysis workbook you should “enable” macros before making any attempts to run the routine.
3. After enabling the macros map the path to the example data.
4. Run the Macro