Vegetation Resources Inventory

Interim Procedures and Standards for Statistical Adjustment of Baseline VRI Timber Attributes

Prepared by

Ministry of Forests and Range

Forest Analysis and Inventory Branch

Version 1.1

January 2008

Table of Contents

1 INTRODUCTION 1

1.1 Background 1

1.2 Overview of Vegetation Resources Inventory 2

1.3 The Role of Attribute Adjustment... 3

1.4 Future Development of Attribute Adjustment.................................... 4

1.5 Roles and Responsibilities 4

1.6 Principles of Adjustment... 4

2 SAMPLE SELECTION 6

2.1 Background 6

3 DESCRIPTION OF BASELINE ATTRIBUTES TO BE ADJUSTED 8

3.1 Background 8

3.2 Baseline adjustable timber attributes 8

3.3 Overview of interim baseline attribute adjustment process 10

3.4 Data Base Configuration... 10

4 DATA PREPARATION AND ASSEMBLY 12

4.1 Merging the Phase I and Phase II data... 12

4.2 Projecting Phase I data to year of ground sampling 13

4.3 Matching height and age data... 14

4.4 Potential Error Sources and Data Screening 16

5 DATA STRATIFICATION...19

5.1 Purpose of Stratification... 19

5.2 Pre-Stratification... 19

5.3 Post-Stratification 20

5.4 Rules on stratification 20

5.5 Sample weights 20

6 ATTRIBUTE ADJUSTMENT PROCESS 22

6.1 Attributes to adjust 22

6.2 Adjustment process 22

6.3 Species composition conversion in FIP inventories 24

7 DERIVATION OF ADJUSTMENT FACTORS 25

7.1 General background 25

7.2 Procedure 25

7.3 Summary statistics from analyses 26

8 COMPUTATION OF POPULATION STATISTICS 29

8.1 Estimation of population total and variance of the population total 29

8.2 Estimation of sampling error and confidence intervals for population total 33

9 ANALYSIS REPORT CONTENTS 35

9.1 Executive summary 35

9.2 Description of the inventory unit 35

9.3 Description of Phase I and Phase II issues 35

9.4 Description of data screening process 36

9.5 Presentation of results 36

9.6 Discussion of planned and attained Sampling error 36

9.7 Summary and conclusion 36

9.8 Sampling plan 37

9.9 Check-list for Attribute adjustment deliverables 37

10 REFERENCES 38

11 APPENDIX A 39

1 Introduction

Purpose of this Document

The purpose of this document is to define the standard process that is recommended for use in adjusting VRI timber attributes. The larger part of the document describes the adjustment methodologies for Vegetated Treed (VT) polygons.

> *Variations from these procedures will be given due consideration, and should be approved after peer review according to the standard VRI change management process.*

The anticipated audience for this document includes biometricians, technicians and consultants involved in VRI data analysis. A basic understanding of biometrics is required to use the document. It is, however, recommended that technicians who may wish to perform the analysis described herein, consult a qualified professional biometrician to approve their work.

Organization of this Document

This document has nine sections:

Section 1 provides a broad overview of the VRI and the role of attribute adjustment.

Section 2 provides a brief overview of the sample selection process.

Section 3 provides a description of the baseline attributes to be adjusted

Section 4 describes the steps in the adjustment data preparation and assembly

Section 5 describes data stratification and weighting

Section 6 provides an overview of attribute adjustment procedures.

Section 7 provides standards for the analysis to derive adjustment factors

Section 8 provides standards for the computation of adjustment analysis statistics

Section 9 provides standards for the minimum content required for the inventory analysis and adjustment documentation.

1.1 Background

The principles and premises on which the Vegetation Resources Inventory (VRI) is based are outlined in the final report of the Vegetation Inventory Working Group (1994). Also, Dr. Kim Iles wrote a number of documents on the various aspects of the Vegetation Resources Inventory (Iles, 1995).

Briefly, the VRI consists of three major processes that include:

1) Phase I - Photo estimation of vegetation attributes on all polygons in an inventory unit
2) Phase II - Ground data collection on a small sample of the photo interpreted polygons, and
3) Statistical adjustment of continuous attributes in the target population, using the relationships between the photo interpreted estimates and the data collected on the ground.

This document represents a progression from the basic adjustment processes, commonly called the Fraser protocol. In that procedure, only three attributes -- height, age and volume (net decay, waste and breakage) -- were adjusted. Ratios of means were recommended for adjusting these attributes.

The Fraser protocol was adopted to cope with short-term timber supply analysis (TSR) information needs. The three attributes considered in the protocol were critical to inventory attribute projection during TSR. As the implementation of the VRI progressed, it was recognized that the adjustment of timber attributes had to maintain the inherent relationships that exist among the attributes. The Fraser protocol maintained the relationships in a way that was easily manageable in TSR.

The protocol in this document has also been designed primarily to accommodate timber supply-related uses. It represents a generalized approach that has not been tailored to meet specific needs of other users. However, newer innovative methodologies may be developed in the future, and after appropriate testing and peer reviews, the standards will evolve to accommodate alternative approaches.

The current statistical adjustment procedures are interim and focus on only six VRI attributes. It is expected that additional attributes will be added to the adjustment process in the future. Appendix A provides draft methodology for a modified adjustment process that is under current investigation.

1.2 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, as well as individual polygon attributes for timber and non-timber resources. This information is used to support:

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

- National Forest Inventory (NFI) goals, including reporting of the Criteria and Indicators of sustainable forest management as defined by the Canadian Council of Forest Ministers.

The VRI process involves initial estimation (commonly called Phase I), ground sampling (commonly called Phase II), attribute adjustment, and other supporting activities including database management and inventory update. The key phases are:

Initial Estimation - forest or 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 when combined with ground data, may improve the precision of estimation of the overall totals and averages for an inventory unit. Initial estimates could come either from existing forest inventories (pre-VRI) or from new delineation and estimation of vegetation characteristics completed to VRI standards.

Ground Sampling – accurate measurement of selected vegetation characteristics. These measurements may be 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 and more accurate than the photo-based estimates. Other ground sampling supporting activities include:

i) net volume adjustment factor (NVAF) sampling and,

ii) within-polygon variation (WPV) sampling.

Data Analysis and Attribute Adjustment – screening and preparing data for analysis and involves creating data summaries from the compiled data, analyzing the relationship between estimation and ground sampling data, developing adjustment factors and extrapolating sample data to make population inferences, and making appropriate adjustments to the initial estimates. In the short term, adjustment procedures identified in this document are the minimum standard for adjustments to vegetation inventories. As enhanced standards are developed, they will replace these interim standards.

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

1.3 The Role of Attribute Adjustment

Attribute adjustment is the process of adjusting the initial estimates, either photo-interpreted or existing inventory, using ground-sampling observations. The objectives of adjustment are twofold:

1. to obtain overall averages and totals for an inventory unit that are statistically unbiased and respect the biological and dynamic nature of vegetation types, and
2. to adjust the existing or new photo-interpreted estimates to obtain adjusted individual polygon values.

1.4 Future Development of Attribute Adjustment

The development of a protocol for adjusting attributes in accordance with the VDYP7 growth and yield model requirements is now complete. A fixed number of attributes will be adjusted (see section 2.2).

The data needs of inventory information users are dynamic, and as such, the protocol described in this document will undergo continual updates and improvements to keep up with the changing needs. The current protocol adjusts five attributes, in accordance with VDYP7 requirements. However, a total of 17 baseline adjustable attributes have been identified (see Section 3.2).

1.5 Roles and Responsibilities

The responsibility for completing VRI data analysis is now diversified to include a role for licensees, the MOFR and other related agencies. Data analysis activities are eligible for funding under the Forest Investment Account. The MOFR is responsible for providing the minimum standards for ground sample analysis and inventory file adjustment and auditing for adherence to these standards. MOFR 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 MOFR with concurrence from district, regional and headquarters inventory program staff.

Note that a general review of the VRI design, including the adjustment process and future needs, was completed by Dr. Stephen Titus in 2007.

1.6 Principles of Adjustment

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 such that averages or totals determined after adjustments have been made, are the same as were found during the statistical estimation or ground sampling phase.

In addressing the two questions above, the VRI adjustment 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 can be improved by ground based adjustments.
3. Adjustments must be developed by logical strata to reflect observed or expected differences in the relationship between the ground sample data (phase 2) and the

inventory estimates (Phase I). For instance, if old growth stands are expected to have different height, age or volume ratios than younger second growth stands, then strata should be created through pre or post stratification to ensure the appropriate adjustment factors are applied in such as manner as not to distort the dynamic nature of the inventory data. (e.g., if necessary, the adjustments should recognize species and age relationships). The adjustments must be applied in a manner that recognizes both that the inventory will be projected over time in timber supply analysis, and that the relationship between the photo-based estimates and ground samples will likely differ across different stand types and conditions. A single overall management unit average that is accurate today may not accurately reflect timber supply considerations in the future, if for example, the adjustments for young forests differ greatly from that for old (currently harvestable) stands.

4. Relationships between the ground data and the photo-interpreted information should be used to adjust the photo-interpreted information, regardless of whether there are statistically significant differences between the ground and photo means for any attribute.
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

2.1 Background

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.

Prior to the use of PPSWR sample selection methodology, the MOFR primarily used *ordered systematic* sampling (OS) methods. Proponents should contact the biometrician at Forest Analysis and Inventory Branch (FAIB) regarding analysis/adjustment methods that used OS sampling.

The following section focuses on describing the analysis and adjustment procedures for a VRI ground sample inventory selected using PPSWR methodology.

PPWSR 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 in the sample analysis.

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 MOFR 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.

3 Description of Baseline Attributes to be Adjusted

3.1 Background

To facilitate the implementation of the statistical adjustment of the VRI database, an internal Ministry of Forests team was formed to determine a workable process. This team produced a list of attributes, which they anticipated would be required by a wide variety of inventory information users. Generally, the most important information needs were identified as being related to timber supply analysis and inventory update.

The timber supply process requires information to predict future timber yields and thus provides information for decision-makers. Inventory update on the other hand is used to update the inventory information to make it current.

The need to satisfy the information requirements for these two primary end-uses tends to dominate the statistical adjustment process. However, there is recognition that there are other inventory information end-uses that must be addressed and a report by the VRI Attribute Adjustment Working Group in 1999 listed most of the major inventory data end-users, and defined what information they required.

The list of attributes classified as “baseline” or “Class 1” has changed since the completion of the report by the VRI Attribute Adjustment Working Group. What were previously called Class 1 attributes have now become the “baseline adjustable timber attributes”. They are the base set of continuous timber attributes that were intented to be adjusted and stored in the corporate database.

Note that the interim procedures laid out in this report restrict the adjustment to 5 of the total 17 baseline adjustable timber attributes. As the adjustment procedure is further refined and tested, it is anticipated that additional baseline attributes will be incorporated into the adjustment.

3.2 Baseline adjustable timber attributes

In the initial development of the VRI adjustment process, the list of baseline attributes included stand height, age, basal area, number of stems per hectare and various net volumes. This list has been refined over time, and now includes the following attributes:

1) Height of 1st species – rank 1 layer
2) Height of 2nd species – rank 1 layer
3) Age of 1st species – rank 1
4) Age of 2nd species – rank 1

5) Site index[1] of 1st species
6) Site index of second species
7) Estimated site index
8) Species composition at 7.5cm+ dbh – all layers combined [2]
9) Number of trees/ha at 7.5cm+ dbh utilization – all layers & species combined
10) Basal Area at 7.5cm+ dbh utilization – all layers & species combined
11) Basal Area at 12.5cm+ dbh utilization – all layers & species combined
12) Lorey Ht at 7.5cm+ dbh utilization level – all layers & species combined
13) Whole Stem volume at 7.5cm+ dbh – all layers & species
14) Whole stem volume at 12.5cm+ dbh – all layers & species
15) Volume net top & stump (CU) 12.5cm+ dbh – all layers & species combined
16) Volume net top, stump (CU + decay) & decay 12.5cm+ dbh – all layers & species combined
17) Volume net top, stump (CU +decay & waste), decay & waste 12.5cm+ dbh – all layers & species combined

The old photo-based inventories were mainly concerned with the estimation of stand height, age, species composition and crown closure. These attributes were used to derive net volume at 12.5 cm or 17.5 cm utilization levels using the variable density yield prediction program VDYP version 6. With the implementation of the current VDYP version 7, there are additional attributes attributes that have become relevant to yield prediction. The list of 17 attributes that appears above is an artifact of the new VRI and VDYP7.

The baseline attributes listed above includes some attributes that will not be adjusted directly (i.e. age of second species) but rather will be derived based on the adjustment of other attributes (i.e. age of first species).

The interim adjustment procedures described in this document focus on the adjustment of only six attributes on this list. However, as previously mentioned, it is expected that additional attributes will be included in the process as adjustment procedures are further developed and refined.

[1] This site index will be derived from adjusted age and adjusted height of the first species.

[2] Species composition is potentially adjustable but the technical application of the adjustment method (sequential balancing) is complicated and controversial. Since it could result in potential negative consequences, species composition adjustment is not currently part of the official MOFR adjustment process.

3.3 Overview of interim baseline attribute adjustment process

The steps involved in the attribute adjustment process are described in detail in Section 6.2. The adjustment itself is sequential in nature, with 2 main stages. At the first stage, an initial set of adjustment factors is developed using the ground sample data and the Phase I inventory data. This first set of adjustment factors is then input into VDYP7, where the factors are applied against the inventory values and additional attributes are derived. At the second stage, another set of adjustment factors are developed for attributes that reflect the adjustments at the first stage. These second stage adjustment factors are input into VDYP7 (the VRIADJUST internal module) that then completes the inventory file adjustment process. The interim adjustment process described in this document has been developed using VDYP7 Console.

Of the 17 baseline adjustable timber attributes outlined in the previous section, only 6 require direct derivation of adjustment factors in the current interim process:

1) Height of 1st species – rank 1 layer
2) Age of 1st species – rank 1 layer
3) Basal Area at 7.5cm+ dbh utilization – all species combined
4) Number of trees/ha at 7.5cm+ dbh utilization – all species combined
5) Lorey Ht at 7.5cm+ dbh utilization level – all layers & species combined
6) Volume net top, stump (CU +decay & waste), decay & waste 12.5cm+ dbh – all species combined

The first four attributes in this list are adjusted at the first stage of the adjustment process whereas Lorey height and volume are adjusted at the second stage.

3.4 Data Base Configuration

The Vegetation Resource Inventory Management System[3] (VRIMS) is the system that manages:

- Veg Inventory data collection,
  - Updates
  - Reinventory
  - Adjustments
- Quality Assurance,
- Data Loading, and
- Data Storage.

The current VRI data model has been re-engineered into the VRIMS data model. VRIMS uses one Oracle Spatial database that is separated into 5 storage areas:

---

[3] This description of VRIMS was based on material presented by Marc Rousseau, MOFR Vegetation Update Specialist, in spring 2007.

1) Staging Area (Database): this area stores information coming in from RESULTS, information coming in from the update group as natural disturbances, information coming as backlog free-to-grow assessments, and information coming in from re-inventories.

2) Estimated Area (Database): this area stores the data for the staging area after it has been QA'd and integrated. The estimated data is also copied to the Adjusted Area (database).

3) Adjusted Area (Database): this area contains a copy of the estimated area data after it has been completed (projected to the reference year) and is used as the base data to which adjustment factors are applied.

4) Reprojected Area (Database): this area stores the data derived from the reprojection service. The reprojection service takes the data from the adjusted area and reprojects it through VDYP7 (grows the trees).

5) Veg Comp Layers: once the annual reprojection is complete, the replication service extracts data from the VRIMS projected area into the Veg Comp Layer, history table, label table and VDYP& Yield input table

The main working format for VRI data will be the Personal Geodatabase (PGDB). VRIMS includes a custom tool to extract data from the Oracle spatial database into a PGDB. This tool can extract data from either the estimated, adjusted, or projected area.

The Phase I data required for the statistical adjustment analysis will be extracted as a PGDB and the VRI statistical adjustment factors will come into VRIMS via the Adjusted Area. However, full implementation of the adjustment procedure in VRIMS is still under development.

4 Data Preparation and Assembly

4.1 Merging the Phase I and Phase II data

Data preparation for the statistical adjustment 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 (Phase II) come from the VRI data compilation process and contain a wide variety of attributes that may not be directly used in the statistical adjustment process. The photo interpreted inventory (Phase I) 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 developing adjustment factors. However, because a time delay often exists between sample selection and adjustment (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 adjustment analysis should be investigated and any issues that arise should be documented. For example, if polygon redelineation 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 statistical adjustment process.

For the attributes that are targeted for statistical adjustment (see section 3.3), it is necessary to merge the Phase II ground attributes with corresponding photo interpreted Phase I 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 done[4].

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 mapsheet 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.

4.2 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 adjustment analysis can begin. If the Phase I inventory is a new VRI, VDYP7 will invoke an internal module called VRISTART to perform the projection. If the inventory is *not* a new VRI, an internal module of VDYP7 called FIPSTART will automatically be run to perform the projection and to 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

---

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

must be documented. Projecting the individual samples to their specific year of ground sampling is considered a refinement to the adjustment process.

4.3 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).

The compiled height and age data from these T, S & L trees is what is used in the adjustment of the Phase I age and height attributes. In order to determine the adjustment 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 adjustment factors 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 adjustment factors 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 applied, the polygon shall be dropped from the age/height adjustment 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 adjustment process. The age/height attributes for that polygon only shall be dropped from the age/ height adjustment 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.

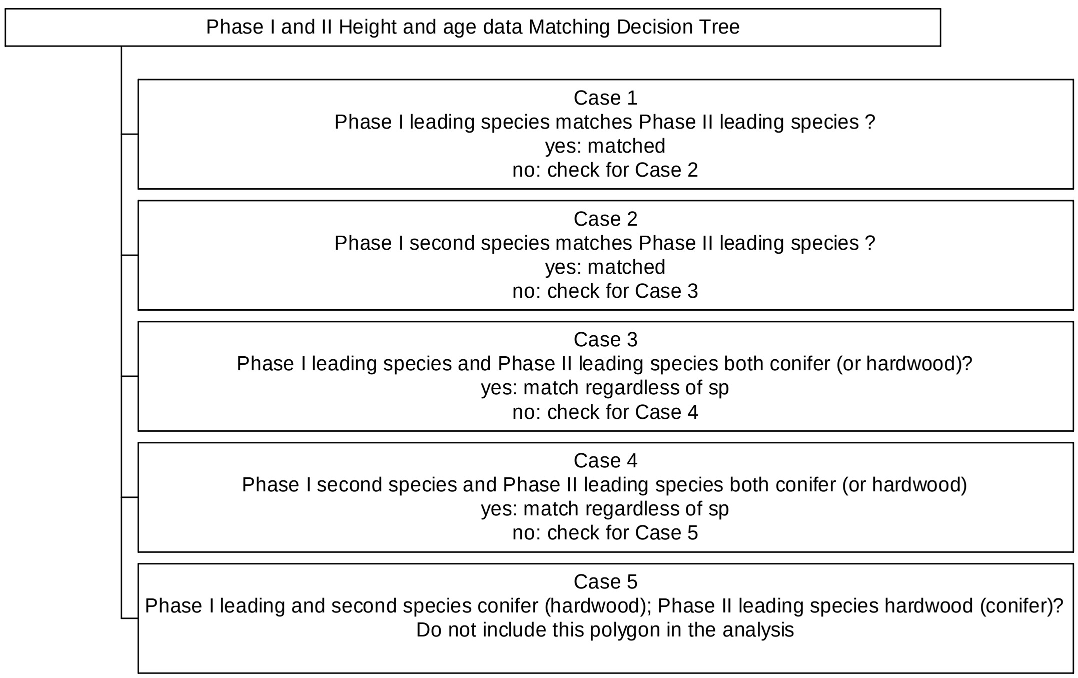

Figure 1: Phase I and II height and age data matching

There are a number of important additional criteria for handling incomplete Phase I or II data[5]:

[5] 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.

- 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 height adjustment ratio but it would not contribute to the age adjustment ratio.
- If the first and second species in a ground sample observation are the same as the corresponding first and second species in the photo interpreted observation, the second species data can be used if the first species values are deemed to be suitable. In most cases the site height and age of the second species represent the dominant and co-dominant trees in a stand, the same way the site height and age of the first species do.
- 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.

4.4 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 five categories:

1) Incorrect or missing Phase I data.
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 4.1 (Merging Phase I and Phase II Data) and 4.3 (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 adjustment factors are developed 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 to not make sense. For instance, a sample record where observed ground height is claimed to be 70 metres 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 screening[6] 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.

4.4.1 Screening height and age

The objective of the height and age matching procedure outlined in Section 4.3 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 carefully investigated.

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 development[7].

---

[6] For example, upper and lower bounds, missing values checks, etc.

[7] 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

4.4.2 Screening basal area & trees per hectare

For inventory projects where the photo interpreted information is collected according to the new VRI standards, the basal area (BA) and trees per hectare (TPH) attributes are interpreted directly from the photos. For projects based on old 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.

4.4.3 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. Adjusted height, age and BA are used as inputs to the VRISTART module in the derivation process. For this reason, the height, age and BA screening, which were described in the previous sections, contribute to the validation of the Phase I component for volume. The screening process for volume may also 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.

---

small counted age. This sample had a significant impact on the value of the adjustment factor for age. After carefully documenting these factors, it was decided to exclude this sample from the derivation of the age adjustment factor.

5 Data Stratification

5.1 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 statistical adjustment 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 individual adjustments relevant to other strata.

5.2 Pre-Stratification

Prior to 2000, there was no direct pre-stratification of target populations before sample selection. The stratification was achieved indirectly through sorting of the polygons in the sampling frame. In a sense, stratification was implicit rather than explicit. During 2000, the sample selection process was changed from selection using a sorted list to selection with probability proportional to size with replacement (PPSWR). The latter works best if it is applied in conjunction with 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 adjustment factor statistics.

The modification in sample selection methodology implies that different types of estimators (equations) will be used to suit the sample selection methodology. This will be discussed further in Section 7. Pre-stratification has several advantages, which include:

a) the strata are identified prior to ground data collection, and sample size augmentation can occur if required.

b) the number of strata is determined prior to analysis. This limits the possibility of data manipulation to find suitable strata after ground data collection

c) the needs of key inventory information users are identified prior to ground data collection.

Pre-stratification has some disadvantages as well, and these include:

a) it can lead to higher cost of the inventory if separate sampling error objectives are set by strata.

b) the inventory information users may identify additional strata which were not identified at the sample selection planning stage.

c) if there is miss-classification in the attributes used to define the strata, the errors are embedded in the sample selection process.

5.3 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 miss-classification 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 to data analysis, if it is determined that more data is require to augment an important stratum.

5.4 Rules on stratification

It is recommended that pre- or post-stratification be based on leading species and/or logical age breaks. Both 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 encouraged.

5.5 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 Section 7 and 8). 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.

6 ATTRIBUTE ADJUSTMENT PROCESS

6.1 Attributes to adjust

A list of timber attributes that are available for adjustment i.e. directly and indirectly is provided in section 3.2. This list includes attributes that will be adjusted indirectly, such as height of the second species, age of the second species, site index of the first species, and site index of the second species. Other attributes included on this list, such as species composition and estimated site index, are generally not adjustable at this time and await refinements in the adjustment process.

The height and age of the second species should be adjusted using corresponding adjustment factors derived for the height and age of the first species. This is a non-statistical adjustment. It is simply an easier method for maintaining the observed photo interpreted relationships in the adjusted data. The site indexes of the first and second species will be derived from the adjusted heights and ages of the first and second species, respectively.

The abbreviated list (from section 3.3), which includes only the attributes that require direct derivation of adjustment factors for the current interim process, 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 and waste at 12.5 cm+ dbh utilization level.

6.2 Adjustment process

Adjustment factors for the attributes listed above are developed in a sequential process, with the first four attributes adjusted at the first stage and the last attribute (i.e., volume) adjusted at the second stage. One set of adjustment factors is produced for each analysis stratum. For example, if the sample data is post-stratified into three leading species groups, one set of adjustment factors will be produced for each leading species stratum. *All analysis for adjustment factor development is based on the rank 1 layer of the photo-interpreted inventory.*

The process is outlined in the steps that follow:

<u>Step 1:</u> 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: Derive adjustment factors for height of first species, age of first species, trees per hectare, and basal area at 7.5cm+ dbh utilization level (see Section 7 for detailed calculations), on a strata basis. This is the first stage of the adjustment process.

Step 3: Input the adjustment factors from Step 2 into VDYP7 and run the model. The adjustment factor input for VDYP7 must specify strata definitions, strata assignments for each polygon, and adjustment factors for each stratum.

Step 4: VDYP7 will generate a wide range of derived attributes. From the VDYP7 output, screen out and compute adjustment factors for:

- Lorey height at 7.5cm + dbh utilization level
- Volume net top, stump, decay and waste at 12.5cm+ dbh utilization level

This is the second stage of the adjustment process. Note that in this step, there is the potential to adjust a number of additional attributes including:

- Basal area at 12.5cm utilization level
- Whole stem volume at 7.5cm utilization level
- Whole stem volume at 12.5cm utilization level
- Volume net top and stump at 12.5cm utilization level
- Volume net top stump and decay at 12.5cm utilization level

At this time, the direct derivation of individual adjustment factors for these additional attributes is not recommended due to issues surrounding harmonization requirements[8] in VDYP7. However, adjustment factors for these attributes must still be input into VYDP7 for Step 5 below.

---

[8] Harmonization ensures that volume and basal area relationships are maintained across utilization levels and volume levels. Attributes output by VDYP7 after the first stage of adjustments will automatically meet the harmonization requirements. However, the adjustment factors derived at the second stage of adjustment may potentially disrupt the inherent relationships among volume and utilization levels as the polygons are projected over time. For example, if the ratio for whole stem volume at 7.5 is smaller than the ratio for whole stem volume at 12.5, the volume at 12.5cm could eventually overtake the volume at 7.5cm and hence violate the harmonization requirement. VDYP7 will not run error-free if harmonization conditions are not met for all polygons. With the interim procedures described in this document, management of these relationships is handled by VDYP7 internally. However, in the future, further procedures for adjustment and harmonization may be incorporated into the standard procedure.

As a result, adjustment factors of 1.0 should be used for basal area at 12.5cm+ dbh. The four volume attributes (i.e., whole stem volume at 7.5cm, whole stem volume at 12.5cm, volume net top and stump at 12.5cm and volume net top, stump and decay at 12.5cm utilization level) should be adjusted using the same adjustment factor computed for volume net top, stump, decay and waste at 12.5cm utilization level. ***If this is not done, the volumes will not be harmonized and VDYP7 will generate errors.***

Step 5: Input the adjustment factors from Step 4 into VDYP7 and run the model. VDYP7 will run an internal module called VRIADJUST which will produce adjusted volumes net, top, stump, decay, waste and breakage at 12.5cm+ dbh and 17.5 cm+ dbh utilization levels.

6.3 Species composition conversion in FIP inventories

Species composition in FIP inventories is volume-based. However, in VDYP7 and new VRI inventories, species composition is basal area-based. As a result, it is possible in some cases for species composition to change when a FIP inventory is processed through VDYP7. This may have implications for stratum assignments for VRI ground samples. In general, stratum assignment should be based on the leading species at the time of sample selection and should not be altered[9].

[9] If a sample switches strata, it must carry with it the appropriate sample selection weight based on the stratum at time of sample selection i.e. the number of hectares represented by one sample within a stratum. To avoid potential issues associated with weights, it is recommended that samples be kept within their original sample selection stratum.

7 DERIVATION OF ADJUSTMENT FACTORS

7.1 General background

There are many methods of deriving adjustment factors for the 5 attributes listed in section 3.3. The principles described in this section are a "best practice". Other biometricians may propose methods that are just as effective. The approach proposed in this document is based on using a ratio of means estimator as an adjustment factor. The adjustment factors are determined based on the relationship of the ground attributes to the inventory attributes. If the data has been pre-, or post-stratified, an adjustment factor will be determined for each stratum. The process of deriving adjustment 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:

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

7.2 Procedure

The basic procedure to derive the adjustment factors involves plotting the data, computing the properly weighted ratio of means, and using analysis of the residuals to evaluate the adjustment factor, 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 adjustment factor, based on the PPSWR sample selection method, as follows:

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

where:

R_h = Ratio of means adjustment factor for stratum h
Y_h = Sum ((polygon ground attribute x polygon area) /probability of polygon selection) for the sample records.
X_h = Sum ((polygon photo interpreted attribute x polygon area) /probability of polygon selection) for the sample records.
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).
n_h = number of sample observations in stratum h

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

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

$\hat{Y}_{hi}$ = Adjusted value of an attribute using the ratio 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-adjusted photo-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 satisfactory[10], alternative post stratification or alternative adjustment models should be investigated. Exploration of post-stratification alternatives would typically be done after the examination of the residuals after the first stage of the adjustment (i.e. based on the height, age, BA and TPH residuals). However, evaluation of the adjustment based on volume residuals could also be done.

7.3 Summary statistics from analyses

After completion of the analysis and derivation of the adjustment factors for the five attributes as required in these interim 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 statistical adjustment in the corporate database. An example of such a summary table is provided in Table 1. Note that the methodology for computing sampling error is provided in Section 8.

[10] i.e. indicative of potential bias or lack of fit

Table 1: An example of summary statistics and adjustment factors that are required for the six baseline attributes that must be adjusted based on the current interim VRI adjustment procedures.

Attribute	*Stratum*	*n*	*Mean weighted Phase II value, by stratum ($\overline{Y}_h$)*	*Mean weighted Phase I value, by stratum ($\overline{X}_h$)*	*Ratio of means adjustment factors*	*Sampling error %*
Height of 1st 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 1st 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 estimated impact of the adjustment must also be computed (see Table 2). Note that the estimated volume *impact* ratio for the sample is not the same as the volume adjustment factor. For the volume adjustment factor, the denominator in the ratio is the VDYP7 volume produced after the first stage adjustment factors (age, height, BA, TPH) have been applied. For the estimated volume impact, the denominator in the ratio is the VDYP7 prior to applying any adjustment factors.

In addition, the sampling error for the estimated total adjusted volume must be computed to determine if the adjustment has met the sampling error target specified in the VRI Project Implementation Plan (VPIP). Once again, Section 8 provides the methodology for computing the estimates shown in Table 2.

Note that since TSR is based on volumes net decay, waste & breakage, the assessment of the overall volume impact and the computation of the overall sampling error for the total volume is based on the same volume (i.e. volume net decay, waste & breakage).

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

Volume Impact: merchantable volume net dwb @ 12.5cm+ dbh	*n*	*Total area (ha)*	*Overall estimated total adjusted inventory volume (m^3)*	*Overall estimated total unadj'd VDYP7 inventory volume (m^3)*	*Overall estimated adjustment impact Ratio*	*Sampling error for total adjusted volume (as % of total adjusted volume)*
Overall	*77*	*1,297,186*	*323,736,512*	*321,462,023*	*1.007*	*12.53%*

As a means of verifying the results when the adjustment is applied to the population, it is also important to provide similar tables based on the inventory population of interest after all adjustment factors have been applied. The population verification should ideally be done at the stratum level. However, at a minimum, the overall volume impact of the adjustment on the population must be compared with the overall estimated volume impact based on the sample.

The most important attribute for the population adjustment verification is volume. To perform the verification, the unadjusted total volume in the population (i.e. prior to the adjustment) is compared with the adjusted total volume in the population (i.e. after all 6 attributes have been adjusted in VDYP7 in the two stage process). This ratio of adjusted total volume to unadjusted total volume – which can be interpreted as the impact of the adjustment on the population total volume – should be compared with the volume impact estimated from the sample as a means of verifying that the adjustment has been applied to the population correctly.

8 COMPUTATION OF POPULATION 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.

8.1 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.

8.1.1 The Separate & Combined Ratio estimators for PPSWR[11]

In this discussion, the sampling unit is the polygon.

8.1.1.1 Notation for the equations

$\hat{Y}_{hji}$ = y_{hji} . a_{hji} = ground measured attribute times area for polygon i in sub-stratum j of stratum h.

X_{hji} = x_{hji} . 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

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

---

[11] Equations 4-13 were initially provided by Dr. Bill Warren. Some of the equations were modified by MOFR 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

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)*

8.1.1.2 General description of the ratio estimators

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

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

The recommended variance estimate is:

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

$$where: \quad e_{hji} = \hat{Y}_{hji} - \hat{R}_{hjs} X_{hji}, \quad X_{hj} = \sum_{i=1}^{n_{hj}} X_{hji}, \quad j = 1,2,...,n_{hj}$$

This variance computation can also be expressed as:

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

In equation [6a], the sum $\sum e_{hji}/a_{hji}$ is equal to 0.0, therefore, the equation simplifies[12] to:

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

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

$$\hat{Y}_{shj} = X_{hj} \hat{R}_{shj}$$

[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:

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

[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:

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

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

[12] This simplification also applies to Equation 5.

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

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

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:

$$\hat{Y}_{ch} = X_h \hat{R}_{ch}, \quad and$$

[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.

8.1.1.3 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)[13]

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 estimator for the two "major" strata and the combined estimator for the three "minor" strata.

---

[13] William G. Cochran. 1963. Sampling Techniques. John Wiley and Sons.

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

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

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

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

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

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

8.1.1.4 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.

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

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

8.2 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 α = 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:

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

9 Analysis Report Contents

The analysis report should provide complete documentation of the adjustment procedure and results. Information should be provided to allow the adjustment factor development 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.

9.1 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.

9.2 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 Biogeoclimatic 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 to which the adjustment factors will be applied 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.

9.3 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 statistical adjustment. 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 statistical adjustment 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 developing the adjustment factors, must also be documented.

9.4 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.

9.5 Presentation of results

The presentation of the results is the most important component of the statistical adjustment report. This section should explain which adjustment methodology was used (i.e. ratio-of-means, least squares, etc.). 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, adjustment factors and the sampling error associated with them by strata, and overall statistics.

An important part of the results section is the verification of the adjustment when it is applied to the population (sample estimated impact of the adjustment compared with actual population impact). The results section should contain a discussion on the implication of the results. It should also provide possible explanations for unexpected results.

9.6 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.

9.7 Summary and conclusion

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

9.8 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 statistical adjustment report. It provides completeness to the VRI. In this format all of the information relevant to the adjustment will be located in the same place.

9.9 Check-list for Attribute adjustment deliverables

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

1. VSIP
2. VPIP
3. Original sample list
4. Population list used to select sample
5. Combined Phase 1 & 2 data used in creating adjustment factors
6. 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 2 data collection, analysis issues and the analysis
    e) a presentation of ratios (or adjustment factors) and their sampling errors
    f) a statement of the risk and uncertainty associated with volume estimation.
    g) recommendations for the application of the adjustment factors
    h) a file with adjusted attributes provided in a format that can be fed into VRIMs
7. A project approval form signed off the proponents & MOFR approvers.

10 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.

11 Appendix A

The following description of a proposed modification to the current adjustment methodology is excerpted from a report prepared by Dr. Margaret Penner, Forest Analysis Ltd. (Penner 2009a and 2009b).

This proposed modification to the adjustment methodology provides unbiased stratum estimates for HT7, AGE7, BA7, TPH7, WSV7, HL7, BA12, WSV12, VCU12, VD12 and VDW12. Rather than using a ratio of means adjustment for trees/ha at the 7.5 cm utilization level and various volumes, an adjustment relationship has been designed to maintain logical consistencies between the various attributes for a polygon.

In the current methodology (herein referred to as the Baseline_06 method), the TPH7 ratio is applied to all polygons to estimate the adjusted TPH7 (i.e. TPH7'). The polygon DQ, based on the adjusted BA7' and TPH7' values, is estimated as:

(1) $$DQ7_{base} = \sqrt{\frac{BA7'}{TPH7'*0.00007854}}$$

In VDYP7Console, volumes will not be produced if the DQ is < 7.5cm. Occasionally, the DQ projected by VDYP7Console is < 7.5cm. This is always accompanied by a processing error indicating the stand is "too short" to be processed and the BA and TPH have not been updated. All polygons that are processed without error will have a projected DQ ≥ 7.5 cm. Although the projected DQ7 may be >7.5 cm, the adjusted $DQ7_{base}$ may be less than 7.5 cm[14]. It is this case that the modified adjustment methodology for BA, TPH and DQ seeks to address.

The modified adjustment methodology ensures that the adjusted $DQ7_{base}$ (refered to as DQ7' in the modified approach) is not less than 7.5cm. In the modified approach, the typical ratio of means is used to adjusted BA7 but rather than using a ratio of means adjustment for trees/ha at 7.5cm utilization (i.e. TPH7), the adjustment relationship is designed to maintain unbiased strata estimates of TPH7 and the logical contraint DQ7' ≥ 7.5 cm.

In the proposed modification, the stratum target TPH7 is estimated as the sum of the adjusted TPH7 (i.e. TPH7'=TPH7 * TPH7 adjustment ratio), by stratum for the population. The k coefficient in the following equation is estimated[15], by strata, such that the adjusted DQ7 (i.e. DQ7') and adjusted BA lead to new adjusted TPH7 that sum to the stratum target TPH7[16]. (

(2) $$DQ7' = 7.5 + k \cdot (DQ7 - 7.5)$$

[14] Penner 2009b looked at the frequency of occurrence of such cases. In the Prince George TSA, VDYP7Console generated a "too short" error and a missing value for DQ7 for about 1.5% of the polygons. The adjusted DQ7' was less than 7.5 cm for about 0.35% of the polygons.

[15] The k coefficient is determined using a simple root finding algorithm.

[16] Stratum totals are area-weighted.

In the current Baseline_06 method, volume adjustment factors are only computed for one volume (volume net top, stump, decay and waste at 12.5cm+ dbh utilization level i.e. VDW12). The ratio for VDW12 is then used to adjust WSV7, WSV12, VCU12, VD12, and VDW12. Basal area/ha at 12.5cm+ dbh (i.e. BA12) is not adjusted. In contrast, the proposed modified adjustment methodology provides individual adjustments for each of the five VDYP7 volumes plus one for BA12 that maintain all the following expected logical relationships (i.e. harmonization):

(3)
$$DQ7' \geq 7.5$$
$$BA7' \geq BA12'$$
$$WSV7' \geq WSV12' \geq VCU12' \geq VD12' \geq VDW12'$$

In this approach, we first define the following functions:

(4) $\log it(x) = \log\left(x \middle/ (1-x) \right)$ with x modified to be in the domain (0.001, 0.999)

(5) $\log it^{-1}(y) = \exp(y) \big/ (1+\exp(y))$ with the result modified to have range (0.001, 0.999)

The adjustments have the following form:

(6) $WSV7' = \log it^{-1}\left[a_2 + \log it(WSV7/(BA7 \cdot HL7)\right] \cdot (BA7 \cdot HL7')$

(7) $BA12' = \log it^{-1}\left[a_3 + \log it(BA12/BA7\right] \cdot BA7$

(8) $WSV12' = \log it^{-1}\left[a_4 + \log it(WSV12/WSV7\right] \cdot WSV7'$

(9) $VCU12' = \log it^{-1}\left[a_5 + \log it(VCU12/WSV12\right] \cdot WSV12'$

(10) $VD12' = \log it^{-1}\left[a_6 + \log it(VD12/VCU12\right] \cdot VCU12'$

(11) $VDW12' = \log it^{-1}\left[a_7 + \log it(VDW12/VD12\right] \cdot VD12'$

The coefficients in equations (6) to (11) are determined using a search algorithm that finds the coefficient such that the sum of the adjusted attribute over the sampled polygons equaled the sum of the ground measured values. The search algorithm is to five significant digits. For example, for equation (6), find a_2 such that

(12) $$\sum_{ground} WSV7' = \sum_{ground} samp\_wsv7$$

This modified approach is effectively the same as estimating the adjusted strata mean as a mean of ratios. As such, the variance can be estimated using the formulas shown in Section 8 of this document and confidence interval calculated. Note the same sample plots are used to estimate the first set of adjustments (AGE, HT7, BA7, TPH7) and the second set (VDW7) and these have correlated errors. If two totally separate data sets were used for the first and second sets of adjustments, the correlations would be zero and the MSRM (2004) sampling error estimators could be used for both adjustments. Here the correlation between the errors in the first and second set of adjustments is ignored.

The modified algorithm described here is straightforward to implement and the gains (unbiased strata estimates and consistency between attributes) justify adoption of this algorithm. It is the most statistically defensible approach and provides unbiased strata estimates and ensures compatibility between adjusted attributes at the polygon level. It does require additional programming and may initially appear complex. In addition the modified adjustment algorithm may require modification to the VRIMS data structure which currently allows for regression adjustment but not more complex adjustments.