Vegetation Resources Inventory

Procedures and Standards for Data Analysis Attribute Adjustment and Implementation of Adjustment in a Corporate Database

Prepared by
Ministry of Sustainable Resource Management
Resource Information Branch

for the
Resources Information Standards Committee

Version 2.0

March 2004

Published by the
Resources Information Standards Committee

National Library of Canada Cataloguing in Publication Data

Main entry under title:
Vegetation resources inventory. Procedures and standards
for data analysis, attribute adjustment and implementation
of adjustment in a corporate database. –- Version 2.0

Available on the Internet.
Also issued in printed format on demand.
Previously issued: Vegetation resources inventory.
Attribute adjustment procedures. Resources Inventory
committee, 2001.
Includes bibliographical references: p.
ISBN 0-7726-5198-1

1. Vegetation surveys - British Columbia – Methodology –
Handbooks, manuals, etc. 2. Forest surveys – British
Columbia – Methodology – Handbooks, manuals, etc.
3. Environmental sampling – Handbooks, manuals, etc.
I. British Columbia. Resources Information Standards
Committee. II. British Columbia. Ministry of Sustainable
Resource management. Resource Information Branch.
III. Title: Vegetation resources inventory. Attribute
adjustment procedures.

QK203.B7V44 2004 333.95'311'09711 C2004-960095-8

Additional Copies of this publication can be purchased from:

Government Publication Services
Phone: (250) 387-6409 or
Toll free: 1-800-663-6105
Fax: (250) 387-1120
www.publications.gov.bc.ca

Digital Copies are available on the Internet at:

http://srmwww.gov.bc.ca/risc

Acknowledgments

The Government of British Columbia provides funding of the Resources Information Standards Committee work, including the preparation of this document. The Resources Information Standards Committee supports the effective, timely and integrated use of land and resource information for planning and decision making by developing and delivering focused, cost-effective, common provincial standards and procedures for information collection, management and analysis. Representatives to the Committee and its Task Forces are drawn from the ministries and agencies of the Canadian and the British Columbia governments, including academic, industry and First Nations involvement.

The Resources Information Standards Committee evolved from the Resources Inventory Committee which received funding from the Canada-British Columbia Partnership Agreement of Forest Resource Development (FRDA II), the Corporate Resource Inventory Initiative (CRII) and by Forest Renewal BC (FRBC), and addressed concerns of the 1991 Forest Resources Commission.

For further information about the Resources Information Standards Committee, please access the RISC website at: http://srmwww.gov.bc.ca/risc/.

For questions concerning the content of this publication please contact the:

Biometrician
c/o Ministry of Sustainable Resource Management
Resource Information Branch
Registries and Resource Information Division
British Columbia Ministry of Sustainable Resource Management
P.O. Box 9358, Stn.Prov.Govt.
395 Waterfront Cresent
Victoria, B.C. Canada
V8W 9M2
Telephone: (250) 387-3592 Fax: (250) 387-5999

Table of Contents

1.0 Introduction

Purpose of this document

The following interim Vegetation Resources Inventory (VRI) Procedures and Standards for Data Analysis, Attribute Adjustment and Implementation of Adjustment to a Corporate Database are the current *minimum* standard that must be followed for the completion of a VRI ground sample inventory. They have been developed largely to provide support for the Ministry of Forests (MoF) timber supply analysis process, while also meeting Ministry of Sustainable Resource Management (MSRM) corporate data business needs.

The primary users of this manual are industrial forest sector staff who are completing VRI inventories, or other government staff who wish to gain an understanding of the VRI process.

Proponents undertaking a Forest Investment Account (FIA) funded, VRI ground sample inventory, must complete an analysis to these minimum standards and provide the required documentation, digital summaries and adjusted inventory files specified in this manual to the MSRM. If during the analysis process, proponents have concerns that these standards may be inappropriate for their business application, they *must* contact the MSRM biometrician at Resource Information Branch (RIB) for authorization for a variance to these standards.

This document references analysis and adjustment procedures for the *Probability Proportional to Size with Replacement* sample (PPSWR) selection methods. Other methods for adjusting inventory files using the previous sample selection method, *Ordered Systematic* (OS) are available from the biometrician at RIB. Users of this manual should ensure the method of analysis matches the method of sample selection.

The following adjustment procedures reference the use of the MSRM Variable Density Yield Prediction model (VDYP version 6.6d). It is not the intent of these procedures to *mandate* the use of VDYP in timber supply analysis, but to present the principles of adjustment to a standing inventory linked to a growth and yield model within a timber supply context. Proponents wishing to use an alternative model, other than VDYP, should contact the RIB biometrician for a variance to these standards.

Proponents undertaking adjustments to inventory files are encouraged to contact RIB regarding either the general process, or the specific details outlined in this manual.

Organization of this Document

This document has seven sections:

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

Section 2 provides an overview of the sample selection process.

Section 3 provides an overview of attribute adjustment procedures.

Section 4 provides standards for the analysis to derive adjustment factors for stand age, height, and volume.

Section 5 provides standards for the adjustment of VRI inventory attribute files.

Section 6 provides standards for the types of analysis and inventory file adjustment verification reports that must be generated and submitted to MSRM.

Section 7 provides standards for the minimum content required for the inventory analysis and adjustment 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 consists of two key components: estimation of vegetation characteristics from aerial photographs (VRI photo-interpreted inventory), and statistically-based ground sampling (VRI ground sample inventory). The ground sample inventory is used to adjust the initial photo estimates.

The VRI ground sample inventory 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 1), ground sampling (commonly called phase 2), attribute adjustment, and other supporting activities including database management and inventory update. The key phases are:

Initial Estimation - of 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 – provides 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[1] and,

ii) within-polygon variation (WPV) sampling.

Data Analysis – is the process of screening and preparing data for analysis and involves creating data summaries from the compiled data, developing adjustment factors and extrapolating sample data to make population inferences.

Attribute Adjustment – is the process of analysing the relationship between estimation and ground sampling data 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.

1.2 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 individual polygon values.

1.3 Future Development of Attribute Adjustment

Development of the final attribute adjustment process, database systems and attributes is under development.[2]

---

[1] http://srmwww.gov.bc.ca/risc/pubs/teveg/index.htm

[2] The final attribute adjustment standards will be linked to a new version of the VDYP growth model (version 7.0), the silvicultural reporting system (RESULTS) and associated Oracle databases.

MSRM recognizes that other VRI photo-interpreted timber attributes such as basal area and density are key inputs into the future VDYP 7 growth model and, as such, are eligible for adjustment. If proponents wish to adjust these additional attributes, they must contact the biometrician at RIB to ensure the adjustments will meet the business needs of the MoF and fit within the MSRM corporate data systems.

1.4 Roles and Responsibilities

As of fiscal 2002, the analysis of VRI ground sample inventory data and the adjustment of VRI inventory database files are the responsibility of the industrial forest sector. These activities are eligible for funding under the Forest Investment Account. The MSRM is responsible for providing the minimum standards for ground sample analysis and inventory file adjustment and auditing for adherence to these standards. MSRM also provides a mentoring and advice role to both industry and other government agencies around all phases and products of the VRI. The business components of these plans are reviewed and approved by the MoF.

1.5 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 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 1). 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.

2.0 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[3].

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

Prior to the use of PPSWR sample selection methodology, the MSRM primarily used *ordered systematic* sampling (OS) methods. Proponents should contact the biometrician at RIB regarding analysis/adjustment methods that used OS sampling.

PPWSR sampling requires that polygons be pre-stratified based on criteria identified in the VRI ground sample project plan (VPIP)[4]. 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 is commonly grouped into three or more volume classes. 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 MSRM 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 and are useful to determine if the original project objectives (level of precision or sampling error at the 95% level of probability) were achieved.

The prestratification 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

---

[3] For more information on the PPSWR sample selection procedures the reader is advised to see http://srmwww.gov.bc.ca/tib/vri/vri/reports&pub/vri_techpub.htm or contact the biometrician at RIB, MSRM.

[4] Examples of VPIPS: http://srmwww.gov.bc.ca/tib/vri/vri/reports&pub/vri_vripub.htm

immediate and any future analysis. Pre-stratification has the drawbacks of potentially increasing sample size, (and costs) for a given level of precision.

3.0 Overview of Attribute Adjustment Procedures

Attribute adjustment is the process of correcting the data from the estimation phase using ground-sampling observations. The purpose of the adjustments is to obtain unbiased overall averages and totals for the inventory unit while recognizing trends in the data by major species and age strata. The process generally involves: developing factors to adjust components of the inventory (e.g., age, height and volume by leading species, and if appropriate, age strata); applying the adjustment factors to the target population; and checking to ensure that the adjustment was applied appropriately.

3.1 Statistical Estimation of Overall Values

There are two kinds of variables collected in the VRI: **continuous** and **categorical.** Each of these requires different analysis procedures.

3.2 Continuous Variables

Continuous variables may be measured to an arbitrary degree of accuracy. For example, the height of tree may be measured in meters, tenths or one hundredths of a meter, where the measurement is not necessarily a whole number.

For the purposes of these interim adjustment procedures, only the following continuous attributes are recommended for adjustment: leading species age, leading species height (for the rank one layer) and aggregate stand volume (net decay waste and breakage) for all layers. In layered stands, it should be noted that the ground volume represents all layers combined. For these interim procedures, the phase 1, layer 1 volume should be used as the inventory estimate.

If proponents have completed a new VRI standard phase 1 inventory and wish to apply adjustments to other timber attributes such as density and basal area, they must contact the RIB biometrician for further direction.

Attribute adjustment of continuous variables is based on the relationship between the estimates and the ground observations. For the current interim procedures, the ratio of means approach is the most commonly used method to adjust the target attributes. In the interim, the use of estimators other than the ratio will be considered on a case by case basis.

3.3 Categorical Variables

Categorical variables, such as species composition, include qualitative and quantitative discrete vegetation and land description attributes which are obtained during ground data collection. Such attributes are not part of the interim age/height/volume adjustment process described in this document and will not be discussed any further in this report.

3.4 Species Composition Adjustment

The adjustment of photo-interpreted species names and percentage composition is outside the scope of this document. The photo-interpreted species composition is directly used along with other inventory variables with adjusted age and adjusted height to determine inventory volume using VDYP (see Section 4: Procedures for Adjusting Stand Age, Height and Volume).

4.0 Procedures for Adjusting Stand Age, Height, and Volume

4.1 Overview

The following procedures describe the ***current interim data analysis and attribute adjustment methods using VRI photo-interpreted and ground sample information. This process was originally developed for incorporating inventory audit data for the Fraser TSA into the timber supply analysis. To maintain consistency with past terminology, this approach will be referred to as the "Fraser Process"***. As mentioned in previous sections of this document, a more comprehensive and long-term process is under development and will be reflected in subsequent revisions of these standards.

The "Fraser Process" currently provides adjustment procedures for only three continuous variables: leading species stand age, leading species height for the rank one layer, and the aggregate volume for all layers. (See Figure 1).

The adjustment is sequential. Adjustment factors are developed for age and height from the ground sample database. These factors are applied against the inventory sample age and height attributes. The adjusted sample height and age are used to derive "attribute adjusted" inventory sample volumes using the Variable Density Yield Prediction (VDYP) model[5]. The "attribute adjusted" sample volume is then compared against the original ground volume. The ratio of the average ground sample volume to the average attribute adjusted sample volume (called the "volume adjustment factor") reflects the error associated with the yield model. The age, height and volume adjustment factors derived from this process are then applied against the entire inventory attribute database. The goal of the adjustment is to remove any overall bias in the inventory when compared against the VRI ground data for the attributes that are adjusted.

---

[5] Currently VDYP version 6.6.

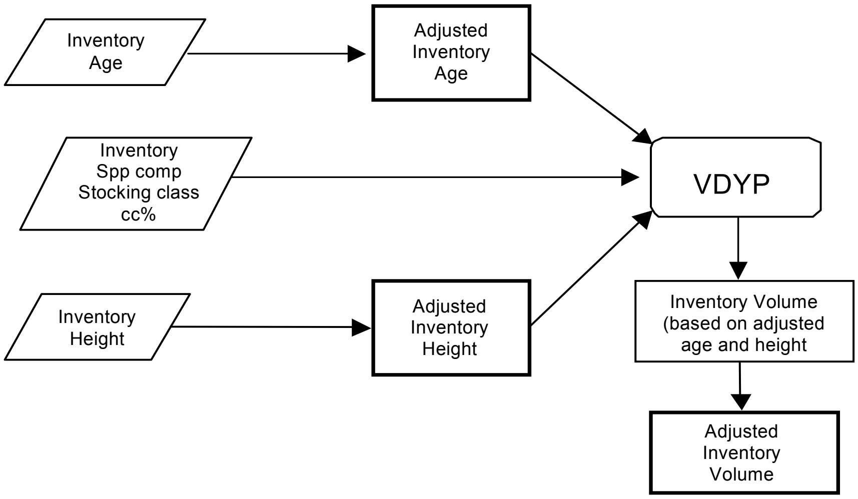


Figure 1. Data and process flowchart for Fraser Protocol.

The basic steps in the analysis and adjustment process are outlined below:

Analysis Stage

1. Merge ground sample data with the photo-interpreted data (or existing inventory), and screen the combined data to eliminate data entry errors. The ground sample volume per hectare (at an appropriate utilization level) is determined at this stage. The inventory sample information (age, height and volume) must be projected to the year of ground sampling. If a new VRI phase 1 inventory standard inventory has been completed, and a projected height for the second species is required for the height/age matching process, height should be projected using site tools.[6]
2. Review the ground and inventory sample information and determine the appropriate leading species site height and age.
3. For the sampled polygons, derive relationships (most commonly a ratio of means) between the ground measured and inventory estimated height and age attributes for the appropriate strata.
4. Using the factors derived in step 3, adjust the inventory sample height and age.
5. Using the adjusted inventory sample height and age in step 3, derive a new VDYP "attribute adjusted" inventory volume. At this point, we assume that the interim volume derived from the adjusted age and height is corrected for the majority of the attribute bias.
6. Derive the ratio of the ground average volume (identified in step 1) to the "attribute" adjusted volume in step 5. This ratio is called the volume adjustment factor. It is different from the impact ratio which is based on the average ground sample volume and the average volume derived from unadjusted attributes. In the VRI database, the volume adjustment ratio is applied against the interim VDYP yield model volume as

---

[6] See MoF Research Branch for a copy of the site tools software.

a correction for any overall inventory *model* bias. See section 5.3.1 for details on the application of the volume adjustment factor on the MS MDB files.

7. Fully document the analysis and provide appropriate tables and statistical summaries.

   The factors derived in this process would generally be summarized into a table reflecting the strata involved in the analysis, e.g., leading species or, if appropriate, age related strata.

Data base Adjustment stage

8. The VRI MDB attribute files (VIF) are generally projected to the last year of ground sampling.
9. Based on the appropriate strata, the existing height and age fields on the MDB files are adjusted by the factors derived in step 3.
10. Based on the same strata used in step 9, the volume adjustment factors derived in step 6 are assigned to the volume adjustment field on the VIF files.
11. The VIF files are processed through the MSRM VEGCAP software (which contains the VDYP yield model) and new polygon volumes and site index values are assigned to the VIF files.
12. The inventory files are thoroughly reviewed to ensure the factors were correctly applied. Summary statistics are generated comparing the impacts of the adjustments against the inventory files versus the ground sample summaries. These summaries are appended to the analysis report derived in step 7.
13. Following the successful review of this process, the adjusted VIF files are loaded to MSRM corporate data servers. Non adjusted files are stored for future reference.

Details of this process are described in the sections that follow.

The first step is to determine adjustment factors for stand age, height, and volume. Inventory attributes for each ground sample will be "paired" with the corresponding photo-interpreted (or existing inventory) attributes by polygon.

In the PPSWR sample selection process, samples observations are grouped according to the strata that were created during sample selection. Post-stratification is feasible; however, a minimum of 10 sample observations for each stratum is recommended to ensure the development of stable adjustment factors.

For the interim "Fraser Protocol" approach, adjustment factors for leading species age, height, and volume are determined based on the ratios of means for the corresponding two variables (e.g., "ground" mean age / "inventory" mean age). Other adjustment approaches such as Geometric Mean Regression, General Least Square Linear Regression, or Non-Linear Regression, are currently under evaluation. However, regardless of the models used to fit the data, the adjusted attributes must be biologically defensible and recognize the dynamic nature of the inventory. For instance, models which predict negative attribute values should not be acceptable because it is biologically impossible to have negative timber attributes. **The process must recognize age related trends in volume bias.**

There are several general analysis steps involved in the adjustment process (see Figure 2). Data is first assembled (Phase 1 inventory estimation data and Phase II VRI ground sample data) and matched for each ground-sampled polygon. If required, pre-stratification weights are calculated, and strata to be used in the analysis are identified. Adjustment factors are then determined for age and height and residual plots are examined. Finally, the volume adjustment factor is determined, and the individual polygon values for the entire population are adjusted.

The process must be documented. The product of this procedure is individual polygon adjusted values that form a VRI attribute adjusted database. The steps in the process are

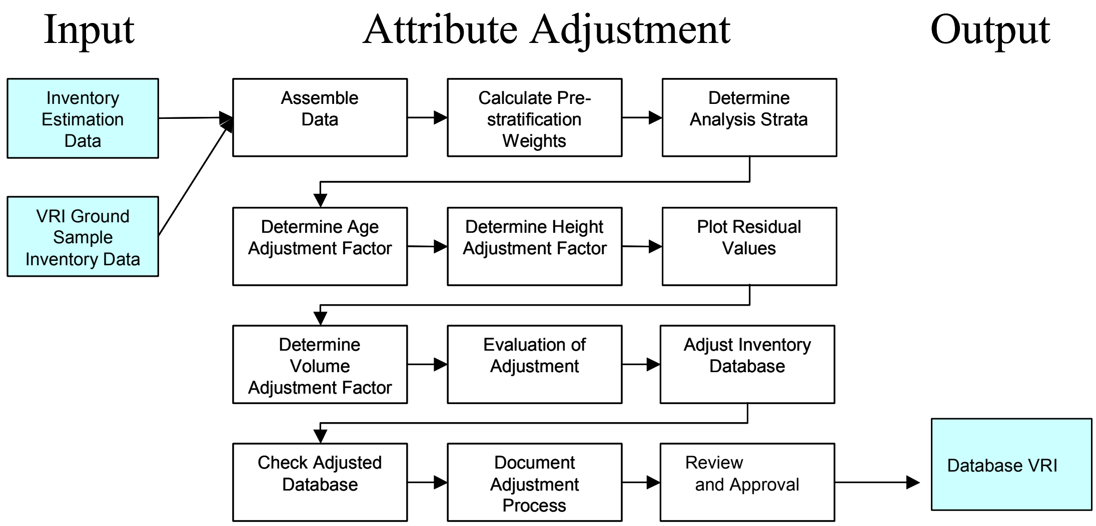


Figure 2. The process steps for the VRI Attribute Adjustment

4.2 Data /Editing and Screening

4.2.1 General Description

Generally, a data set consisting of key inventory and corresponding ground sample timber attribute data is assembled in a common Excel spreadsheet application for initial screening. Spreadsheets are a useful medium to present the information as they allow for easy viewing and manipulation of the data, and they are simple to understand and verify.

The inventory estimates may come from several sources including: new VRI photo-interpretation; traditional Forest Inventory Planning (FIP) files that have been converted to the new VRI format; silvicultural or disturbance history, etc. Normally, the same version of the population 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. The inventory data must be projected to the same date as the ground sampling. If significant changes (e.g., new photo-estimation/updates) have occurred since the time of the initial sample selection, the impacts on the analysis/adjustments should be investigated, and any issues that arise should be documented.

4.2.2 Data Assembly/Edit Procedures

The following general procedures are provided for consideration:

1. **Obtain compiled VRI ground sample data.**

   The MSRM is responsible for setting the standards for ground sample volume compilation. Proponents can obtain the necessary compilation and data entry/validation software to compile the data themselves or contact a qualified consultant. If NVAF sampling has been completed, ensure that the appropriate factors have been applied to the net factored volumes. VRI ground sample volumes must be compiled using the MSRM VRI compilation software and report values using the compiler file and output file naming conventions.

2. **Obtain inventory sample attribute data**

   Obtain inventory attribute sample information from the Vegetation Inventory Files (VIF) and project to the year of ground sampling.

3. **Stratification and weighting**

   In the PPSWR sample selection method, sample polygons are generally pre-stratified by leading species with the number of plots per strata assigned, based on the proportion of the strata to the entire population. The computation of individual strata and combined strata statistics should not present any problems if the original strata are maintained. If the sample data are re-grouped using different stratification criteria from the original, a qualified biometrician should be consulted in order to ensure the appropriate weights are applied.

   If the sample was selected using OS sample selection methods, the reader should also consult RIB for more information.

4. **Determine post-stratification criteria**

   The original population may be post-stratified based on attributes/factors that can improve adjustment. The most common post-stratification criteria are age, species, biogeoclimatic zone and disturbance history. Care must be taken to avoid creating an excessive number of strata with low sample sizes. The results based on very small sample sizes (less than 10 observations) have a higher risk of providing misleading inferences about the population. The intent of this section is to acknowledge that post stratification of a PPSWR sample, although not strongly encouraged, may be applied if the trends in data support such analysis.

5. **Define utilization standards for volume**

   The net volume type and utilization level, which will be used as the basis for determining the inventory sampling error for volume, should be defined prior to the adjustment. Generally, the selected net volume type and utilization level should correspond with what is used in the timber supply analysis for a given inventory unit. Note that the ground utilization selected for computation of the volume adjustment factor will be reflected in the adjusted inventory population.

The MSRM recognizes that management unit specific differences in minimum diameter limits thresholds and their application in the development of volume adjustment factors may create minor inconsistencies in volume reporting at the provincial level. However, these differences are considered minimal and will be addressed in the release of subsequent adjustment standards and databases.

6. Create an Excel spreadsheet table and import the inventory data and sample data

Spreadsheet tables are a useful medium for the initial screening and editing of the ground and inventory data. Care should be taken during the data preparation stage to ensure that all inventory (Phase I) and ground sample data have appropriate one-to-one correspondence for each sample polygon. (Mismatching rows or columns will have severe consequences in the subsequent data analysis). MSRM standards require that a summary file is provided. See section 6.1 for more details.

7. Check, verify, and clean any problems and errors in the data

The merged sample data may have problems that need to be identified and corrected ("cleaned") prior to the adjustment. There are many details to be determined for each attribute including missing data items, errors, invalid codes, etc. Age, height, and volume scatter plots and various summary statistics are useful to identify data entry errors, logged samples, coding mistakes, or any other unreasonable and unrealistic values. Potential outlying values should be identified and investigated. The reasons for rejecting "outliers" should be defensible and carefully documented. Ground samples with zero volume that fell in natural openings are not excluded from the analysis.

8. Treatment of recently logged samples

The issue of assessing the impacts of logged samples on the vegetated treed populations has been addressed by MSRM in the following manner:

If the population of interest is Vegetated Treed (VT) and the ground sample integrated plot centre points have fallen in a "recent" clear-cut opening, then by definition, the plot did not sample the vegetated treed population and, therefore, should be excluded from the analysis. This logic assumes that the files used for sample selection were not up-to-date with respect to harvest openings and that the ground sample does not represent the VT population. Proponents should carefully evaluate these issues on a case-by-case basis to ensure that samples are not excluded that fall in older harvest openings that have missed the depletion update process or have fallen in natural stand openings.

9. Treatment of multi-layered stands

The VRI photo-estimation phase allows for the assignment of attributes on a layer basis (up to 9 layers). On the ground, however, the identification of vegetation attributes by layer is difficult, if not impossible, to implement on a consistent basis. As such, attributes by layer are not assigned during the ground sample phase. This creates difficulties in the analysis stage, as it is impossible to match photo-estimated attributes to the ground sample attributes by layer.[7] The following section describes how MSRM deals with adjustments to multi-layered stands using the interim "Fraser process". It is recognized that this method is less than ideal and will be modified in upcoming releases of VDYP 7 and alternative adjustment methods.

In multi-layered stands, the photo-interpretation process provides a full description of timber attributes which include height of the 1st and 2nd species, age of the 1st and 2nd species; species composition; number of trees and basal area per hectare. Layers are assigned a "rank" through the Vegcap process. This process allows for the identification of the primary layer (rank 1) and the secondary layer (rank 2). The primary layer forms the largest proportion of the stand and is the main crop of interest in stand projection.

The height and age attributes for the rank 1 layer should be used for developing adjustment factors. The rationale for this approach is that the ground measurements for site index are more likely to correspond with the layer information that constitutes the largest proportion of the stand. This issue is particularly important when a stand contains a small proportion of residual trees which may have captured the site index information. Rigorous screening should be conducted to ensure that the residual trees are excluded from consideration in such situations.

Note: A log of all data issues identified in this phase of the analysis and their subsequent treatment must be maintained and documented in the final VRI analysis report.

4.3 Calculate Pre-Stratification Weights

4.3.1 Sample weights

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

1. sampling unit weight, 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 probability proportional to size with replacement 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.
2. 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.
3. 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).

---

[7] On a provincial basis, multi-layered (> 1 layer) stands comprise approximately 5% of the provincial forested land base, while individual management units may contain up to 20%.

The sampling unit weight and the strata proportion weights are most commonly used (see equations in Sections 4.6.2, 4.7.2, 4.8.2 & 4.9.1). If samples are selected with different sampling intensities within a stratum, then an additional weight must be applied. A biometrician should be consulted to ensure the correct weights are applied in such circumstances.

4.4 Determine Analysis Strata: Post-Stratification

4.4.1 Description

Post-stratification can improve the accuracy of adjustment. It may not be desirable to adjust inventory attributes using a single factor for the entire unit as this may distort the dynamic nature of the inventory, especially volume and age relationships. Bias trends observed in the residual analysis may require revisiting and revising the initial post-stratification.

4.4.2 Procedure

The relationships between ground and inventory attributes can be reviewed using the following suggested criteria for defining the analysis strata:

- biological similarity (e.g., Phase I leading species grouping, BEC, age);
- similarity of inventory and ground relationships for the attributes (e.g., similarity of ratios for different groupings of samples);
- correspondence to analysis units; and
- number of samples available.

Note: Post-stratification will not change the pre-stratification weights. See section 4.3

4.5 Procedure for Selecting VRI Height and Age Sample Data in Preparation for Computing Adjustment Factors

4.5.1 Description

Height and age are critical attributes in VRI in that they are key to determining site index. In turn, site index is an essential element in gauging site productivity, a key element in timber supply analysis. As such, considerable care is required in screening sample data used to derive adjustment factors for these attributes. The following procedures for adjusting height and age are presented in order to ensure consistency in VRI adjustments. Proponents are encouraged to use these methods in analysis and provide any comments regarding their application to RIB.

The principles underlying the screening process are based on the following assumptions and factors:

1. The sample data is typically stratified by Phase I leading species. For this reason, as much effort as possible should be made to ensure that most of the heights used in the derivation of the adjustment factors are associated with the Phase I (photo-estimated) and Phase II (ground) leading species. A reasonable level of mismatch should be expected, and the levels of agreement and disagreement should be reported as meta-data in the analysis data log section of the final VRI analysis documentation.
2. The stratification by Phase I leading species is an attempt at assembling stands of similar growth characteristics in one stratum. The screening process should take this into account in cases where ground and photo data are mismatched.
3. 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.
4. 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.
5. It is assumed that both Phase 1 (photo-estimation) and Phase 2 (ground data) define leading species based on the same reference utilization level (>=4.0 cm dbh).
6. It is assumed that the rank 1 layer height and age for phase 1 are matched with equivalent ground sample (phase 2) data, even though the ground data are not portioned by layer.

There are currently two formats for attributes in the photo-estimation (Phase 1) inventories in general use in BC: the current VRI format (VIF) and the old Forest Inventory Planning file (FIP) format. The VRI format (VIF) provides timber attributes for both the first and second leading species, while the old FIP format provides attributes for only the first, or leading species. For data obtained from the old FIP files, slight modifications will be required to facilitate height and age adjustment that conform to VRI standards. For example, since FIP data does not provide second species height and age, Cases 2, 4 and 6 that are described in Section 4.5.2: Matching Phase I and II Data, would not apply.

4.5.2 Matching Phase I and II Data

In the VRI, height and age are closely linked. For that reason, the screening process described below is applicable to both attributes. In the following description, all references to the data should be interpreted as including both height and age data.

The adjustment ratio for height (and age) is computed based on a selection of sample tree data (top height, leading and/or second species site tree height) depending on the ability to match Phase I and Phase II height data by leading or second species. The seven cases listed below should cover all combinations of species data. Based on analysis of prototype data for the Sunshine Coast, it is expected that 80% of the polygons will be matched using Cases 1, 2, 3 or 4. If there is a match, the paired observations are used to compute the adjustment ratio as described below in Section 4.5.3: Analysis of Data.

Case 1. If the Phase II leading species (determined at the > 4.0 cm DBH utilization level) is the same as the Phase I leading species at the *species code level*, 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 at the *species code level*, match the Phase I second species data with the Phase II leading species data.

Case 3. If the Phase II leading species is the same as the Phase I leading species at the *Sp0*[8] *code level*, match the Phase I leading species data with the Phase II leading species data. Note that the 16 possible Sp0 codes approximate genus (with some exceptions such as PY and PL, etc.). Please refer to Appendix B for a table that translates species codes to the appropriate Sp0 code.

Case 4. If the Phase II leading species is the same as the Phase I second species at the *Sp0 code level*, match the Phase I second species data with the Phase II leading species data.

Case 5. If a match cannot be found using Cases 1, 2, 3 or 4, 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 6. If a match cannot be made based on Case 5, 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 7. If no other case applied, the polygon shall be dropped from the age/height adjustment analysis. Case 7 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 or in error. The age/height attributes for that polygon only shall be dropped from the age and height adjustment analysis. The polygon will still be useful for analysis of other timber attributes such as volume and basal area. The height and age for a non-matching polygon will still be adjusted using the derived ratio, and it will still contribute to the derivation of the volume adjustment.

---

[8] Sp0 codes are generated based on genus and species similarities. A translation table for converting species to Sp0 is provided in Appendix B.

Phase I and II Height and Age Data Matching Decision Tree

Case 1
Phase I leading species matches Phase II leading species at species code level?
yes: matched
no: check for Case 2

Case 2
Phase I second species matches Phase II leading species at species code level?
yes: matched
no: check for Case 3

Case 3
Phase I leading species matches Phase II leading species at Sp0 code level?
yes: matched
no: check for Case 4

Case 4
Phase I second species matches Phase II leading species at Sp0 code level?
yes: matched
no: check for Case 5

Case 5
Phase I leading species and Phase II leading species both conifer (or hardwood)?
yes: match regardless of species
no: check for Case 6

Case 6
Phase I second species and Phase II leading species both conifer (or hardwood)
yes: match regardless of species
no: check for Case 7

Case 7
Phase I leading and second species conifer (hardwood); Phase II leading species hardwood (conifer)?
Do not include this polygon in the analysis

Additional criteria for handling incomplete data:

1. If both age and height data are missing for the matching species (either Phase I or II), do not include the polygon in the analysis.
2. If 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, for a given polygon, if age is missing but height is available, this polygon would contribute to the height adjustment ratio but it would not contribute to the age adjustment ratio.
3. Where leading and second species composition (either Phase I or II) is equal and one or both of age and height are missing, the second species may be taken to provide complete data or a match.

4.5.3 Analysis of data

1. For a single cluster ("sample"), depending on the Case identified for the Phase I and Phase II data match, select the appropriate Phase I height (and age) estimate. This is only applicable if there are Phase I heights for both the first and second species. Note that if the Phase I second species 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 must be completed manually, as required, using the MoF Research Branch product "site tools."

2. For the Phase II leading species (defined as the leading species by basal area/ha determined at the > 4.0cm DBH utilization level), compute the average height from the Phase II top height, leading, and/or second species site tree height (TLS) tree data. If the Phase I and Phase II data match is based on the species code level (i.e., Case 1 or 2), then the Phase II average height for the leading species in a cluster is obtained from the "Mean HT for TLS" field and the appropriate species code line in the cluster-by-species summary file (e.g., Project name_smy_ncs.xls, Age_tls, Ht_tls).

   If the Phase I and Phase II data match is based on the Sp0 code level (i.e., Case 3, 4, 5 or 6), then the Phase II average height for the leading species in a cluster is obtained from the "Mean HT for TLS" field and the appropriate Sp0 code line in the cluster-by-species summary file. Note that data matches at the Sp0 code level may require computing a weighted average "Mean HT for TLS" for two or more Phase II species that contribute to the same Sp0 code.
3. Repeat step one for all clusters ("samples") in a stratum (or population). This gives a collection of relevant data (averages). Summaries of the matching process must be documented in the Sample Database/Analysis Application Check file in section 6.1.
4. Compute the averages of Phase I and II data for all clusters on a sample by strata basis, using the PPSWR analysis methodology. Compute the adjustment ratio using these means. Variations from the ratio of means approach will be considered on a case-by-case basis, depending on the nature of the data in a given project. Note that the sample size for the height and age means may differ since one of age or height may be missing for a given sample.
5. Apply the adjustment factor (step 3) to both first and second species for each Phase I polygon in the stratum (or population). The adjustment factor is applied to all samples in the stratum regardless of whether or not that sample contributed to the computation of the adjustment factor (i.e., where Phase II age and/or height was missing for a given sample or where there was "no match" for the sample, the Phase I age and height is still adjusted for that sample).

4.5.4 Adjustment of VRI Inventory Second Leading Species

For VRI standard inventories, the second leading species is adjusted using the same ratio developed for the first leading species.

4.6 Determine Height Adjustment Factor

4.6.1 Description

The height adjustment factor is determined based on the relationship of the ground to inventory height. If the data has been pre-, or post-stratified, a height adjustment factor will be determined for each stratum.

4.6.2 Procedure

1. Plot the data (ground vs. inventory) to visually analyze the relationship.
2. Calculate the Ht adjustment factor or ratio for the probability proportional to size with replacement (PPSWR) sample selection method as follows:

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

where:

R_{shj} = height ratio for sub-stratum j in stratum h
$\hat{Y}_{hji}$ =. y_{hji} . a_{hji} = ground measured height times area for polygon i in sub-stratum j of stratum h.
X_{hji} = x_{hji} . a_{hji} = photo interpreted or unadjusted inventory height times area for polygon i in sub-stratum j of stratum h.
P_{hji} = a_i/A_{hj} = selection probability of polygon i in sub-stratum j of stratum h
n_{hj} = number of sample observations in substratum j of stratum h
a_{hji} = area of polygon i in sub-stratum j of stratum h.
A_{hj} = area of sub-stratum j of stratum h

3. Compute the residuals ($\hat{Y}_{hji}$- $\hat{R}_{hj} X_{hji}$) where $\hat{Y}_{hji}$ and ($\hat{R}_{hj} X_{hji}$) are the ground and ratio-adjusted heights. A plot of the computed residuals against the *adjusted* values provides information on how well the ratio defines the relationship between ground and photo interpreted or inventory heights (with the polygon area applied as defined above - see the example in Section 4.8.3). For an ideal fit, the residuals should appear as a “band” of points distributed, uniformly on both the positive and negative sides of the zero line. Occasionally, the points will be arranged to form a trend cutting across the zero line. Such occurrences are symptoms of an inappropriate ratio relationship. A biometrician should be consulted to determine the course of action to take in such circumstances.
4. If residual plots are not satisfactory, explore alternative post stratification for the adjustment or alternative adjustment models.

4.7 Determine Age Adjustment Factor

4.7.1 Description

The age adjustment factor is determined based on the relationship of the ground to inventory age attributes. If the data has been pre-, or post-stratified, an age adjustment factor will be determined for each stratum.

4.7.2 Procedure

The procedure for age adjustment is similar to the procedure for height adjustment.

1. Perform a preliminary visual analysis of the relationship between ground and photo interpreted Age by plotting ground age against photo-based age.
2. Calculate the Age adjustment factor or ratio for the Probability proportional to size with replacement (PPSWR) sample selection method as follows:

$$\hat{R}_{shj} = \frac{\sum_{i=1}^{n_{hj}} \hat{Y}_{hji} / p_{hji}}{\sum_{i-1}^{n_{hj}} X_{hji} / p_{hji}} = \frac{\sum_{i=1}^{n_{hj}} \hat{Y}_{hji} / a_{hji}}{\sum_{i=1}^{n_{hj}} X_{hji} / a_{hji}}, \quad p_{hji} = \frac{a_{hji}}{A_{hj}} \qquad [2]$$

where:

R_{shj} = age ratio for sub-stratum *j* in stratum h
$\hat{Y}_{hji}$ =. $y_{hji} \,.\, a_{hji}$ = ground measured age times area for polygon *i* in sub-stratum *j* of stratum *h*.
X_{hji} = $x_{hji} \,.\, a_{hji}$ = photo interpreted or unadjusted inventory age times area for polygon *i* in sub-stratum *j* of stratum *h*.
$P_{hji} = a_i / A_{hj}$ = selection probability of polygon *i* in sub-stratum *j* of stratum *h*
n_{hj} = number of sample observations in substratum *j* of stratum *h*
a_{hji} = area of polygon *i* in sub-stratum *j* of stratum *h*.
A_{hj} = area of sub-stratum *j* of stratum *h*

3. Compute the residuals ($\hat{Y}_{hji}$- $\hat{R}_{hj} X_{hji}$) where $\hat{Y}_{hji}$ and *(*$\hat{R}_{hj} X_{hji}$*)* are the ground and ratio-adjusted age values. A plot of the computed residuals against the *adjusted* values provides information on how well the ratio defines the relationship between ground and photo interpreted or inventory ages (with the polygon area applied as defined above - see the example in Section 4.8.3). For an ideal fit, the residuals should appear as a “band” of points distributed, uniformly on both the positive and negative sides of the zero line. Occasionally, the points will be arranged to form a trend cutting across the zero line. Such occurrences are symptoms of an inappropriate ratio relationship. A biometrician should be consulted to determine the course of action to take in such circumstances.
4. If residual plots are not satisfactory, explore alternative post stratification for the adjustment or alternative adjustment models.

4.8 Determine Volume Adjustment Factor

4.8.1 Description

The adjusted values of height and age are used to compute interim (Phase I) inventory sample volumes in the VDYP 6 model. The interim inventory volumes, along with compiled ground sample volumes, are used to determine the volume adjustment factors. It is important to note that the interim inventory volumes are based on adjusted heights and ages. The ground volume should ideally be adjusted with the NVAF; however, in the case where an NVAF has not been completed, net factored "only" volumes (i.e. net factored volumes without NVAF) should be used. It should be noted that net factored "only" volumes will, in most cases, overestimate volume and that stakeholders will need to address this issue. The utilization levels and standards that correspond to the volume adjustment factors must be carefully documented.

As with height and age, the inventory and ground volume must be computed to conform to the PPSWR sample selection methodology.

4.8.2 Procedure

1. After heights and ages are adjusted, use the adjusted values to compute inventory volumes in VDYP. Plot the data (ground volume vs. interim inventory volume) to visually analyze the relationship.
2. Calculate the volume adjustment factor or ratio[9] for the Probability proportional to size with replacement sample selection method as follows:

$$\hat{R}_{shj} = \frac{\sum_{i=1}^{n_{hj}} \hat{Y}_{hji} / p_{hji}}{\sum_{i-1}^{n_{hj}} X_{hji} / p_{hji}} = \frac{\sum_{i=1}^{n_{hj}} \hat{Y}_{hji} / a_{hji}}{\sum_{i=1}^{n_{hj}} X_{hji} / a_{hji}}, \quad p_{hji} = \frac{a_{hji}}{A_{hj}} \qquad [3]$$

where:

R_{shj} = volume ratio for sub-stratum j in stratum h
$\hat{Y}_{hji}$ =. y_{hji} . a_{hji} = ground measured volume times area for polygon i in sub-stratum j of stratum h.
X_{hji} = x_{hji} . a_{hji} = photo interpreted or unadjusted inventory volume times area for polygon i in sub-stratum j of stratum h.
P_{hji} = a_i/A_{hj} = selection probability of polygon i in sub-stratum j of stratum h
n_{hj} = number of sample observations in substratum j of stratum h
a_{hji} = area of polygon i in sub-stratum j of stratum h.
A_{hj} = area of sub-stratum j of stratum h

3. Compute the residuals ($\hat{Y}_{hji}$- $\hat{R}_{hj} X_{hji}$) where $\hat{Y}_{hji}$ and ($\hat{R}_{hj} X_{hji}$) are the ground and ratio-adjusted volume values. A plot of the computed residuals against

---

[9] More details on the PPSWR sampling design are available at the following site: http://www.for.gov.bc.ca/resinv/Veginv/techdocs/techdocuments.htm

the adjusted values provides information on how well the ratio defines the relationship between ground and photo interpreted or inventory volumes (with the polygon area applied as defined above - see the example in Section 4.8.3). For an ideal fit, the residuals should appear as a "band" of points distributed, uniformly on both the positive and the negative sides of the zero line. Occasionally, the points will be arranged to form a trend cutting across the zero line. Such occurrences are symptoms of an inappropriate ratio relationship. A biometrician should be consulted to determine the course of action to take in such circumstances.

4. If residual plots are not satisfactory, explore alternative post stratification for the adjustment or alternative adjustment models. If examination of the residuals indicates that alternative or additional post-stratification is required, adjustment factors for age and height will also need to be recalculated to correspond to the new post-stratification. (See section 4.6)

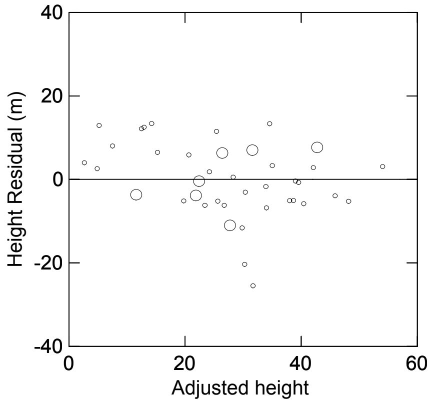
4.8.3 Examine Residuals Plots

The performance of the model can be assessed using residual analysis (see the example of residual plots below for height adjustment of the Hemlock/Balsam stratum in the test data set).

For every sample point, the residual can be calculated by:

"Residual" = ground (Phase II) value – adjusted inventory value

Example: Residuals of height for Cw/Hw/B strata for the test data set.



The horizontal line at the zero residual value indicates where the adjusted inventory value would be equal to the ground value. Points above the line represent underestimation in the adjusted values, whereas points below the line represent overestimation in the adjusted values. In some software packages, the size and shape of the ploted points can be altered. In the illustration, sizes of data points indicate the relative weights of the samples.

4.9 Computation of the Variance of the Ratio

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.

4.9.2 The Separate & Combined Ratio estimators for PPSWR[10]

In this discussion, the sampling unit is the polygon.

4.9.2.1 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*

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

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

---

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

$v(\hat{R}_{ch})$ = 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)

4.9.2.1 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:

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

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

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[11] to:

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

---

[11] This simplification also applies to Equation 5.

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:

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

[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:

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

$$v(\hat{Y}_{ch}) = X_h{}^2 v(\hat{R}_{ch}) \qquad [11]$$

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.

4.9.2.4 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)[12]

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.

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

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

$$[12]$$

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

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

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

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

---

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

4.9.2.5 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.

4.9.3 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 MSRM first. Each situation will be examined on a case-by-case basis.

5.0 Adjust the Inventory Database

5.1 Description

Using the adjustment factors determined from Sections 4.6, 4.7 and 4.8, the height, age and volume attributes in the inventory unit database are adjusted. This is the so-called "Fraser Process". The implementation of the adjustment process occurs within MS Access file processing systems.

5.2 General Procedure

1. Identify the population to be adjusted

   Ideally, the inventory population to be adjusted should be identical to the population used to select the samples. However, because there is a time delay between population assembly for sample selection and adjustment, the original population may have changed due to natural or man-made disturbances. If the depletion update information is available, it should be applied to the database prior to the development of adjustment factors. This updated information should then be the foundation for the statistical adjustment. To achieve the depletion update and the statistical adjustment successfully to meet the timber supply review needs, requires close collaboration and coordination with the Timber Supply Branch. Ultimately, the TSB is the key user of this information, and they need to be closely involved with the process.

   It is important to note any restrictions to the adjustment population of interest that may have arisen during the adjustment factor analysis. If the adjustment factors were developed from a subset of the original population used to select the samples (for example, if the analysis was restricted to only those samples greater than 30 years of age) then the adjustments must be applied to a similar inventory population.

2. Determine post-stratification for adjustment

   The target population data should be screened to ensure that all polygons can be identified by the strata which were specified during the development of adjustment factors. Each polygon in the target population should belong to only one stratum. If there is ambiguity, and a polygon can potentially be identified with more than one stratum, then the population data contains errors, and these should be corrected before the adjustment process.

   The appropriate adjustment factors should be applied to the relevant strata for which they were developed.

3. Inventory file adjustment

   The general process used to adjust the inventory files are as follows:

- Adjust the inventory age, by polygon and by stratum

Adjusted Age = Age Adjustment Factor x Inventory Age[13]

- Adjust the inventory height, by polygon and by stratum

Adjusted Height = Height Adjustment Factor x Inventory Height[14]

The adjusted age and height are used to generate interim (attribute-adjusted) volume for each polygon using VDYP 6 or any other yield model that uses height and age as the key drivers in yield prediction. The interim inventory file volumes are required for the volume adjustment checks in section 6.4.

- Populate the volume adjustment factor fields in the MDB files with the appropriate volume adjustment factor. Then generate adjusted volumes which will reflect the impact of adjusted height and age.

Final Adjusted Inventory Volume = Volume Adjustment Factor x Interim Volume

4. Inventory file adjustment check

   The process is verified and documented.

5.3 Detailed Adjustment Procedures

The following detailed procedures have been developed to assist proponents in adjusting VIF format MS Access MDB files. In particular, they specify what procedures should be carried out to ensure that the adjustments were applied correctly to the MSRM corporate database and defines the types of summary reports that must be supplied to MSRM.

5.3.1 Adjusted Database Fields

The following section identifies the MS Access Tables and Field Names that are to be adjusted using the VDYP 6 yield model. This documentation assumes users are not adjusting stands < 30 years of age[15].

The MDB Tables and Fields[16] that need to be adjusted are:

1) Layer Table:
Field Name:

VOLUME_ADJUSTMENT_FACTOR
REFERENCE_YEAR

---

[13] Ensure age is projected to the year of ground sampling.

[14] Ensure height is projected to the year of ground sampling.

[15] Recently, MSRM has been advising proponents to not sample stands < 30 years of age due to the complexity and lack of business need associated with the adjustments.

[16] MDB data dictionary: http://srmwww.gov.bc.ca/tib/drm/odd/overview.html

2) **Leading_species table:**
Field Name:

AGE
HEIGHT

In order for the VDYP 6.0 yield model to predict and project volumes correctly, the original *field names* that are adjusted (AGE/HEIGHT) must not change. This is counter intuitive as the adjusted field names exist on the current MDBs; however, as the VDYP 6 model has not been programmed to recognize the adjusted field names, the original field names must be maintained. These issues will be addressed with the release of VDYP7 and a revised VIF file structure.

5.3.2 Steps to apply adjustments to the MDB file using the VIFUPDATE versions of VDYP (version 6.6d)

The following section identifies the steps required to adjust the VIF MDB files.

1. Save a backup of the original file before proceeding with the following adjustments.
2. Project stand attributes (height and age) to year of ground sampling used to derive adjustments.
3. Apply adjustment factors (e.g. ratios or regression coefficients) to AGE_PROJECTED and HEIGHT_PROJECTED fields in the Leading_species table.
4. Copy AGE_PROJECTED values to AGE, HEIGHT_PROJECTED to HEIGHT fields.
5. Change contents of REFERENCE_YEAR field to year used in step 2.
6. Load VOLUME_ADJUSTMENT_FACTOR values.
7. Note: for stands that were > 30 years of age prior to adjusted but <31 years of age after adjustment, VDYP expects to have the photo-interpreted (Layer) ESTIMATED_SITE_INDEX, (Layer) ESTIMATED_SITE_INDEX_SPECIES, and (Layer) SUPPLIED_SITE_INDEX_SOURCE fields populated.

8 Run VEGCAP software that will project and apply the adjusted volumes to the VIF files. For users not working with corporate government database systems, please contact Tim Salkeld at RIB for more information. (Tim.Salkeld@gems4.gov.bc.ca)

6.0 Analysis and Adjustment Documentation

The following section identifies the types of confirmation reports that are required in order to verify that the analysis and adjustments were applied using appropriate standards. The summary reports specified in this section are a mandatory requirement and must be supplied to MSRM.

6.1 Sample Data Base/Analysis Application Check File

In order for the MSRM to verify that the VRI ground sample analysis conforms to the minimum standards, the following ground sample database summary file must be provided to the MSRM. The file must be provided in MS Excel format.

The following tables specify the appropriate VRI ground sample compiler *file and field* names that must be used to populate the indicated table.[17]

---

[17] VRI compiler file and field names are referenced against "Data Dictionary for Vegetation Resources Inventory and National Forest Inventory Timber Data" http://srmwww.gov.bc.ca/tib/vri/vri/vri_data.htm

Table 1: VRI general attributes and identifiers

Common Name	**VRI sample number**	**VRI project number**	**Mapsheet**	**Polygon**	**Sample plan strata**	**Final analysis strata**	**Original VPIP sample plan weight**	**Final analysis sample weight**	**BCLCS**	**Polygon area**	**Special cruise number or PSYU**	**FIZ code**	**BGC zone & sub zone**	**Owner-ship code**	**Forest cover Rank code**	**Layer ID**	**Type ID at reference**	**Projected date**	**Year of ground sampling**
Database or text file attribute name	samp_no	Proj_id	Map_no	Poygon	The strata defined in the original VPIP; referenced against the original sample selection files.	The strata defined in the final analysis if differenct from the sample plan strata; based on inventory population; may incorporate age-related trends.			BCLCS Level 1	polyarea	Special cruise number or PSYU	FIZ_CD	BEC	OWN_CD	rank_cd	Layer ID	typid_ref	proj_dt	Meas_dt
Compiler output summary file name	smy_c or smy_cs	smy_c or smy_cs					na	na											Smy_c
MS MDB table name	na	na	layer	layer	na	na	na	na	resultant	resultant	Non tree attributes	resultant	resultant	resultant	tree cover	layer	tree cover	version	na
example	101	DMH1	092P088	409	VT PINE	VT PINE			VTUTMOP	20.70	141	G	IDFmw2	62-c	1	1	1	010101	2001

Table 2: VRI inventory attributes

Common Name	**Inventory species composition from 1st to 6th**	**Inventory type group**	**Reference year**	**Input first spp height**	**Input first spp age**	**Input second spp height**	**Input second spp age**	**Projected 1st spp height (to year of ground sampling)**	**Projected 1st spp age (to year of ground sampling)**	**Projected 2nd spp height (to year of ground sampling)**	**Projected 2nd spp age (to year of ground sampling)**	**Crown closure**	**Projected stocking class**	**Site index**	**Inventory Estimated Volume (with specified utilization)**
Database or text file attribute name	sspcs1, pct1, sspcs2, pct2, …, sspcs6, pct6	itg	ref_yr	ht_in	age_in	s_ht_in	s_age_in	ht_prj	age_prj	2nd spp height not automatically projected; must do manually if required for match	2nd spp age not automatically projected; must do manually if required for match	crwn_cls	stkcl_pr	site_idx	This volume must be generated independently using VDYP batch with the appropriate utilization
MS MDB table name	tree species	tree cover	layer	species	species	species	species	layer	layer	na	na	tree cover	tree cover	tree cover	na
example	PL 90 S 10	28	72	14.5	70	6.0	35	17.9	99			50	3		167.6

Table 3: VRI ground compiled attributes

Common Name	**Ground species % by BA at 4cm+ for live trees only**	**Ph II leading spp at 4cm+**	**Mean total age (TLS trees) for Ph II lead spp**	**Mean top height (TLS trees) for Ph II lead spp**	**Inventory lead spp.**	**Inventory 2nd spp**	**Case # for match**	**Inventory age for match**	**Inventory height for match**	**Ground age for match**	**Ground height for match**	**Net factored "only" vol/ha, close U, net dwb, live trees only, at specified utilization**	**NVAF vol/ha, close U, net dwb, live trees only, at specified utilization**
Database or text file attribute name	SPB_CPCT	Derived from the SPB_CPCT	aget_tls	ht_tls	Derived from the inventory data	Derived from the inventory data	See VRI ground sample analysis standards Sec.4.5. These are the inventory and ground age & height values used to compute the age & height adjustment factors.					vht_nwb	nvl_nwb
Compiler output summary file name	smy_nc	Derived from smy_nc	smy_ncs	smy_ncs								smy_nc	smy_nc
example	PL 86 S 14	PL	77	27.2	PL		1	99	17.9	77	27.2	191.341	188.9483

6.2 Stand Level Adjustment Application Check File

In order for the MSRM to verify that the inventory adjustment factors were appropriately applied to the VIF files, the following population summary file must be provided to the MSRM. The summary file must be generated for at least two mapsheets in the project area and should ideally cover the range of the population. The summary must be provided in MS Excel format.

Table 4: Data to be included in summary file for stand level adjustment application check.

	Original data	***Adjusted Data***
	Layer.MAP_ID	Layer.MAP_ID
	Layer.POLYGON_ID	Layer.POLYGON_ID
	Layer.LAYER_ID	Layer.LAYER_ID
	Layer.FOR_COVER_RANK_CD	Layer.FOR_COVER_RANK_CD
	Layer.ESTIMATED_SITE_INDEX_SPECIES	Layer.ESTIMATED_SITE_INDEX_SPECIES
	Layer.ESTIMATED_SITE_INDEX	Layer.ESTIMATED_SITE_INDEX
	Layer.SUPPLIED_SITE_INDEX_SOURCE	Layer.SUPPLIED_SITE_INDEX_SOURCE
	Layer.VOLUME_ADJUSTMENT_FACTOR	Layer.VOLUME_ADJUSTMENT_FACTOR
	Layer.SITE_INDEX	Layer.SITE_INDEX
	Leading_species.AGE	Leading_species.AGE
	Leading_species.HEIGHT	Leading_species.HEIGHT
	Leading_species.PROJECTED_AGE	Leading_species.PROJECTED_AGE
	Leading_species.PROJECTED_HEIGHT	Leading_species.PROJECTED_HEIGHT
	Leading_species.UPDATE_AGE_DATE	Layer.reference_Year.
	Layer.CROWN_CLOSURE	
	Rank 1 layer species	Adjustment Stratum
	Rank 1 layer species percent	Age * Ratio – a calculated value (unadjusted age * age adjustment ratio) to compare against the assigned adjusted value.
		Height Ratio - a calculated value (unadjusted ht * ht adjustment ratio) to compare against the assigned adjusted value.

The example in Table 5 compares information from the *Layer* table and the *Leading Species* tables, before and after adjustment. Basic calculations are performed on this data to ensure that the adjustments were applied to the polygons correctly. The example in Table 5 shows inventory data that has been projected to 2001 (the year of ground sampling), adjusted and then projected to the current year (2003).

The *Layer* table comparison ensures that the Reference_Year is set to the year of ground sampling, that the Volume_Adjustment_Factor field is populated, and where required, that stands adjusted from >30 years old to <31 years old have their Estimated_Site_Index_Species, Estimated_Site_Index, and Estimated_Site_Index_Source fields populated.

The *Leading Species* table comparison ensures that the Age and Height fields are properly adjusted. This is confirmed by comparing actual assigned adjusted values (i.e., polygon values after the adjustment has been applied) with calculated values to check the adjustment process (see the "Calculated ratios and differences" section of Table 5). For example, the Heights in the 2001 data must be replaced with Projected_Heights. These new Heights will then be adjusted, so by multiplying the Projected_Heights by the Ht_Adj value, the calculated heights should match the Heights in the 2003 *Leading Species* table. The same applies to Age and Projected_Age information[18]. The Differences fields in the "Calculated Ratios and Differences" section of Table 5 are simply derived by subtracting the Age and Height information in the 2003 data from the values just calculated. The *Species* table information is useful in confirming that the polygons have been assigned to the correct adjustment strata.

---

[18] The comparison is illustrated and referred to in Table 5 by letter-referencing relevant columns.

Table 5: Data for checking the adjusted inventory file

Unadjusted, Projected to 2001
Layer Table

Map_id	Polygon_id	Layer_id	For_cover_rank_cd	Estimated_site_index_species	Estimated_Site_index	Supplied_site_index_source	Volume_adjustment_factor	Site_index	REFERENCE_YEAR	Crown_closure
092I033	1	1	1					9.1	1987	20
092I033	2	1	1					11.1	1987	60
092I033	3	1	1					15.1	1987	70
092I033	4	1	1					11.1	1987	60
092I033	5	1	1					13.1	1987	50
092I033	6	1	1					12.8	1987	70
092I033	7	1	1					13.9	1987	7
092I033	8	1	1					14.6	1987	40
092I033	9	1	1					12.2	1987	60
092I033	10	1	1					13.5	1989	50
092I033	11	1	1					10.7	1987	60
092I033	12	1	1					6.9	1987	50
092I033	13	1	1					9.2	1987	60
092I033	14	1	1					12	1987	50
092I033	15	1	1					11.5	1987	60
092I033	16	1	1					12.5	1987	40
092I033	17	1	1					11.5	1987	60

Adj and Prj to 2003
Layer Table

Estimated_site_index_species	Estimated_Site_index	Supplied_site_index_source	Volume_adjustment_factor	Site_index	REFERENCE_YEAR
			1.418	10.1	2001
			1.329	11.5	2001
			1.191	14.8	2001
			1.329	11.5	2001
			1.329	13.2	2001
			1.329	13.1	2001
			1.000	13.9	2001
			1.329	14.5	2001
			1.575	15.9	2001
			1.575	16.9	2001
			1.575	13.4	2001
			1.137	6.8	2001
			1.575	12	2001
			1.191	11.7	2001
			1.329	11.7	2001
			1.575	14.5	2001
			1.329	11.7	2001

Unadjusted, Projected to 2001
Leading Species Table

Layer_id	Species_id	Age	Height	Projected_age [A]	Projected_height [B]
1	1	200	18	214	18.5
1	1	100	16	114	17.1
1	1	120	24	134	25.4
1	1	100	16	114	17.1
1	1	120	20	134	20.9
1	1	100	18	114	19.2
1	1	120	22	134	23.3
1	1	160	24	174	24.6
1	1	120	22	134	23.9
1	1	140	26.4	152	27.7
1	1	200	28	214	28.9
1	1	140	13	154	13.7
1	1	200	26	214	27
1	1	110	18	124	19.3
1	1	120	18	134	18.9
1	1	220	30	234	30.9
1	1	120	18	134	18.9

Adj and Prj to 2003
Leading Species Table

Age [C]	Height [D]	Projected_age (i.e. 2003)	Projected_height (i.e. 2003)
166	18.7	168	18.8
101	16.5	103	16.7
155	26.6	157	26.8
101	16.5	103	16.7
119	20.1	121	20.2
101	18.5	103	18.7
134	23.3	136	23.5
155	23.7	157	23.8
94	22.6	96	23
107	26.1	109	26.4
150	27.3	152	27.5
165	14.0	167	14.1
150	25.5	152	25.7
143	20.2	145	20.3
119	18.2	121	18.3
164	29.2	166	29.4
119	18.2	121	18.3

Actual Applied Adjustment Ratios

Poly	Stratum	HT ADJ Ratio [E]	AGE ADJ Ratio [F]	VOL ADJ
1	FNV31	1.009	0.775	1.418
2	PTV101	0.963	0.889	1.329
3	WFTV139	1.046	1.154	1.191
4	PTV101	0.963	0.889	1.329
5	PTV101	0.963	0.889	1.329
6	PTV101	0.963	0.889	1.329
7		1.000	1.000	1.000
8	PTV101	0.963	0.889	1.329
9	SBDTV31	0.944	0.701	1.575
10	SBDTV31	0.944	0.701	1.575
11	SBDTV31	0.944	0.701	1.575
12	PNV31	1.022	1.069	1.137
13	SBDTV31	0.944	0.701	1.575
14	WFTV139	1.046	1.154	1.191
15	PTV101	0.963	0.889	1.329
16	SBDTV31	0.944	0.701	1.575
17	PTV101	0.963	0.889	1.329

Calculated Ratios and Differences

Age * Ratio i.e. A*F [G]	Height * Ratio i.e. B * E [H]	Age Difference [i.e. C - G]	Height Difference [i.e. B - H]
165.85	18.7	0.15	0.00
101.35	16.5	-0.35	0.00
154.64	26.6	0.36	0.00
101.35	16.5	-0.35	0.00
119.13	20.1	-0.13	0.00
101.35	18.5	-0.35	0.00
134.00	23.3	0.00	0.00
154.69	23.7	0.31	0.00
93.93	22.6	0.07	0.00
106.55	26.1	0.45	0.00
150.01	27.3	-0.01	0.00
164.63	14.0	0.37	0.00
150.01	25.5	-0.01	0.00
143.10	20.2	-0.10	0.00
119.13	18.2	-0.13	0.00
164.03	29.2	-0.03	0.00
119.13	18.2	-0.13	0.00

Species Table

Polygon_id	Layer_id	Species_id	Species_cd	Species_percent
1	1	1	FD	70
1	1	2	PY	30
2	1	1	PL	90
2	1	2	FD	10
3	1	1	FD	70
3	1	2	PL	30
4	1	1	PL	90
4	1	2	FD	10
5	1	1	PL	60
5	1	2	FD	40
6	1	1	PL	70
6	1	2	FD	30
7	1	1	FD	70
7	1	2	PL	30
8	1	1	PL	100
9	1	1	SE	60
9	1	2	B	20

6.3 Overall Population Adjustment Check File

In order for the MSRM to assess the overall impacts of the adjustment process against the inventory population files, the following summaries must be generated. The data should be supplied in text format in the final analysis summary report.

Area summaries.

Sum Total Area
Sum VT Area
Sum Area by Operability
Sum Area by THLB
Sum Area by any other layer used to define the populations (ie BEC)
Sum Area by Strata
Sum Area by Strata by Map
Sum Area of Adjusted Polygons

Original data:

Sum Volume[19] and Sum Area for VT Stands - Total
Sum Volume and Sum Area for VT Stands > 60 Years (approximate Inventory audit comparison)[20]
Sum Volume and Sum Area by Age Class, VT
Sum Volume and Sum Area by Strata

Adjusted data:

Sum Volume and Sum Area for VT Stands - Total
Sum Volume and Sum Area for VT Stands > 60 years (approximate audit comparison)
Sum Volume and Sum Area by Age Class, VT
Sum Volume and Sum Area by Strata

6.4 Volume Adjustment factor check:

In order for the MSRM to assess the reliability of the adjustments against the inventory files, the following summary table must be provided.

---

[19] “Volume” for a polygon is the volume/ha multiplied by the polygon area. The “sum volume” divided by the “sum area” will generate the area weighted volume/ha. The volume utilization must be identified.

[20] http://srmwww.gov.bc.ca/tib/vri/vri/vri_audits.htm

	Sample							**Population**				
strata	**Inventory volume (m3/ha) (PH1)**	**Attribute Adjusted volume**	**Ground volume (PH2)**	**Ph2/Ph1 ratio (R)**	**Sampling Error @95% level of probability**	**Final Adjusted volume (m3ha)**		**Inventory volume [A]**	**Attribute adjusted volume [B]**	**Final Adjusted volume [C]**	**Ratio of Adjusted volume to Attribute Adjusted volume [C/B]**	**Impact Ratio [C/A]**
All Strata												
A												
B												
C												
D												
E												
F												
G												

The following table summarizes the table attributes:

Sample:

Strata: The final strata definitions used to adjust the inventory

Inventory “Sample” Volume: The average sample volume, compiled at a utilization level consistent with the ground sample for the total sample and the individual strata populations. The utilization must be specified.

Attribute Adjusted ‘Sample Volume: The average attribute adjusted inventory sample volume, compiled at a utilization level consistent with the ground volumes.

Ground Sample Volume: The average ground sample volume, compiled at a utilization level consistent with the inventory volume, for the total and individual strata populations. The utilization must be specified.

PH2/PH1 ratio: The weighted ratio of the average ground sample to inventory volume for the total and strata populations.

Sampling Error of the PH2/Ph1 Ratio: The sampling error of the ratio of the PH2/PH1 at the 95% level of probability.

Final Adjusted Volume: The final average adjusted sample (adjusted by age and height and the volume adjustment factor) volume at a utilization level consistent with the inventory and ground samples.

Population:

Inventory Population Volume: The average inventory population volume, compiled at a utilization level consistent with ground sample for the total and individual strata populations. The volume utilization must be specified. In addition, the population definition for the adjustment must also be specified (e.g. VT stands >30 years of age, etc.)

Inventory Population Attribute Adjusted Volume: The average population, attribute adjusted volume compiled at a utilization level consistent with ground sample volumes for the total and individual strata populations.

Final Adjusted Population Volume: The final average adjusted population (adjusted by age and height and the volume adjustment factor) volume at a utilization level consistent with the inventory and ground samples.

Ratio of Final Population Adjusted Volume to Population Attribute Adjusted Volume: The ratio of the average final population average adjusted volume to the arithmetic average attribute adjusted inventory volume for each strata and the population.

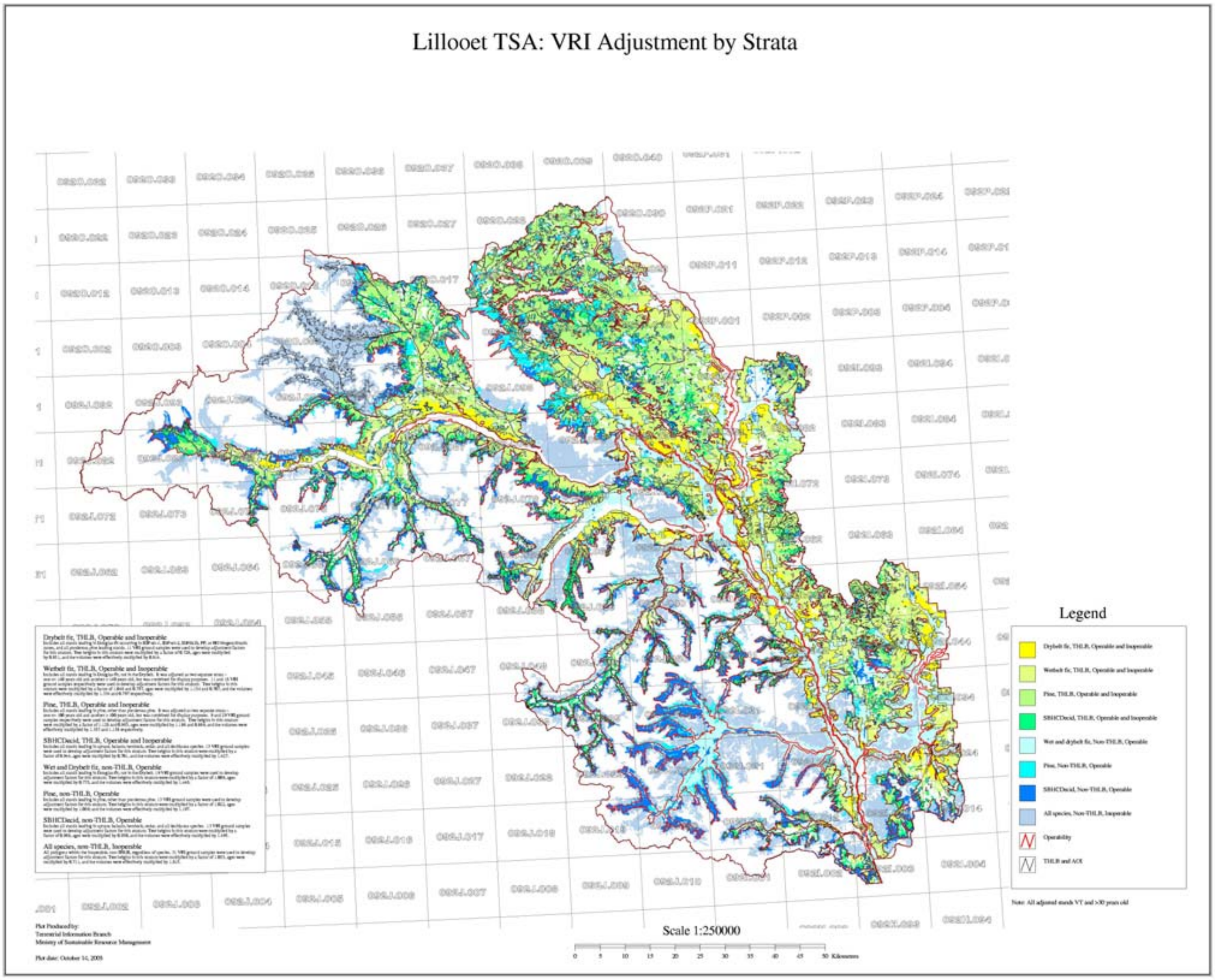
Impact Ratio: The ratio of the average inventory volume to the final adjusted arithmetic average population volume by strata and the total.

6.5 GIS Coverage Check

In order for the MSRM to spatially confirm that the appropriate adjustments were applied to the inventory files, a GIS colour themed map must be produced.

There is no specific format that is required, but a 1:250,000 scale map showing colour themed strata is most commonly developed.

An example of the Lillooet TSA is attached:


Lillooet TSA: VRI Adjustment by Strata
Drybelt fir, THLB, Operable and Inoperable
Wetbelt fir, THLB, Operable and Inoperable
Pine, THLB, Operable and Inoperable
SBHCDecid, THLB, Operable and Inoperable
Wet and Drybelt fir, non-THLB, Operable
Pine, non-THLB, Operable
SBHCDecid, non-THLB, Operable
All species, non-THLB, Inoperable
Legend
Drybelt fir, THLB, Operable and Inoperable
Wetbelt fir, THLB, Operable and Inoperable
Pine, THLB, Operable and Inoperable
SBHCDecid, THLB, Operable and Inoperable
Wet and drybelt fir, Non-THLB, Operable
Pine, Non-THLB, Operable
SBHCDecid, Non-THLB, Operable
All species, Non-THLB, Inoperable
Operability
THLB and AOI
Note: All adjusted stands VT and >30 years old
Plot Produced by:
Terrestrial Information Branch
Ministry of Sustainable Resource Management
Plot date: October 14, 2003
Scale 1:250000
0 5 10 15 20 25 30 35 40 45 50 Kilometres

7.0 Documentation

The attribute adjustment process must be carefully documented and submitted to MSRM for audit. The documentation must contain the following minimum information.

7.1 Data Log

All steps of the adjustment process should be documented in a data log, along with a record of all changes to the data and the reasons for those changes.

7.2 Original Data

The data used in sample selection should be archived to form a point of reference for subsequent analysis. By the time the ground sampling is completed, changes may have occurred affecting the original data set used in sample selection. If depletion or other attribute updates occur prior to the development of adjustment coefficients, a new separate data set containing the updates should be archived to form the basis for the final statistical adjustment. In some cases, depletion or non-depletion updates may result in unexplained polygon area changes. An effort should be made to explain why the changes in polygon area occurred.

All data should conform to the VRI data formats and standards.

7.3 Summary Report

All stratum-based adjustment factors should be summarized in a standard report along with the analysis results, such as data scatter plots, data tables, residual plots, and sampling errors. The following report outline is proposed for data analysis reports.

Outline

- Executive summary
- Description on the inventory unit
- Description of Phase I and Phase II issues
  Inaccessible / dropped plot summary
- Description of data screening process
- Presentation of results
- Discussion of planned and attained Sampling error
- Summary and conclusion
- Original sampling plan

7.4 Risk and Uncertaintity

The final report must identify, as much as possible, issues associated with the analysis that could represent weakness or increased uncertaintity associated with the adjustments. One of the key critteria for assessing risk and uncertaintity is the individual and overall sampling error for a VRI project. All analysis reports should contain summary tables showing these sampling errors..

Appendix A: VRI Adjustment Reports

Examples of completed VRI adjustment reports can be found at the following website:

http://srmwww.gov.bc.ca/tib/vri/vri/reports&pub/vri_vripub.htm

The final analysis report must contain the following:

A.1 Executive summary

The executive summary should outline the study objectives and indicate whether they were met satisfactorily. It should be a brief description.

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.

A.2 Description on the inventory unit

A description of the inventory unit should indicate where it is located within the province of British Columbia. It should describe the Biogeoclimatic zones that exist in the inventory unit and any other physiographic factors that might influence growth in the unit.

A.3 Description of Phase I and Phase II issues

A summary of the status of photo-interpreted inventory 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 as they may impact the statistical adjustment.

A.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.

A.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. It could be the ratio or least squares analysis. If other methodologies are used, a rationale should be provided.

This section should contain a number of tables showing various results, which may include summary statistics on various attributes of interest, adjustment factors and the variances associated with them by strata, and overall statistics. The results section should contain a discussion on the implication of the results. It should also provide possible explanations for unexpected results.

A.6 Discussion of planned and attained Sampling error

A section on sampling error should be provided. This 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.

A.7 Summary and conclusion

The summary and conclusion section puts 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.

A.8 Sampling plan

The original VRI ground sample inventory project implementation plan (VPIP) must be appended to the final analysis report. The final analysis will be compared against original plan objectives.

Appendix B: Conversion table from species to Sp0 codes

Sp0 code	*Species code*
AC	A AC AH
AT	AT
B	B BA BB BC BG BL BM BP
C	C CI CW IG IS J JR
D	D DG DM DR VB VP VV W XH XH
E	E EA EP EW
F	F FD XC ZC
H	H HM HW

Sp0 code	*Species code*
MB	GP KC M MB MR MV QG RA
L	L LA LE LT LW
PA	PA PF
PL	PJ PL PM PR
PW	PS PW
PY	PY
S	S SA SB SE SS SW SX
Y	CP CY TW Y YC

Appendix C: Impact of Expert Matching Process

The following example illustrates how the matching process and analysis works for each of the cases described above. For the analysis, the first five cases are the Phase II sample. The full ten cases represent the population. This figure includes shading and is best printed in colour.

Case	Phase I				Phase II leading spp		Adjusted (Expert matching	
	1st sp	Ht1	2nd Sp	Ht2	sp0	Sp0 Hts	Adj Ht1	Adj Ht2
1	Hm	25	Fd	22	H	30	25.4	22.3
2	Cw	30	S	28	Sw	28	30.5	28.4
3	A	18	S	30	Ac	15	18.3	30.5
4	S	35	A	22	D	20	35.5	22.3
5	Ba	18	S	30	A	25	18.3	30.5
6	Fd	35	S	30	Fd	35	35.5	30.5
7	S	28	Fd	32	Fd	33	28.4	32.5
8	Hw	25	S	25	Hm	30	25.4	25.4
9	Fd	28	M	19	Fd	30	28.4	19.3
10	S	34	Fd	26	Fd	38	34.5	26.4
Population Averages		27.60	26.55556	26.40		28.40	28.02	26.80
	Ratio							
Ratio: Phase II / Phase I leading	1.028985507	28.4	27.3	27.2				
Ratio: Phase II / Expert Matched	1.069456067	29.5	28.4	28.2				
Ratio: Phase II / Phase II leading	1.075757576	29.7	28.6	28.4				
Sample Averages		25.2	23.25	26.4		23.6		
	Ratio							
Ratio: Phase II / Phase I leading	0.936507937	25.8	25.8	25.8				
Ratio: Phase II / Expert Matched	1.015053763	28.0	28.0	28.0				
Ratio: Phase II / Phase II leading	0.893939394	24.7	24.7	24.7				

If we use the "Expert Matching" criteria (described in Section 4.5.2: Matching Phase I and II Data) for determining the sample x-mean for use in calculating the ratio, then we cannot determine the population mean of the x using that criteria because the rule requires knowledge of the y. Applying that rule to the leading species (x) will therefore introduce some level of bias in estimates since population means for the leading species and the expert matching are not the same mean (and we cannot compute the latter). If the expert rules lead to better height adjustments for most polygons, then we may expect that the adjusted polygon values will provide a better estimate of the population mean of the x (expert adjusted heights).

In the example, using the sample data (first five cases), three ratios (Phase II mean divided by Phase I leading, expert matched, and Phase I second) are computed (.936, 1.01, .89). These are then applied to the population mean of x using the three sets of data (Phase I leading, expert matched, and Phase I second). In this example, the expert-matching x definition seems to provide the best estimate of population mean for Phase II leading species height (28.0). It may not always turn out this "nice." The right-most columns of the table show adjusted Phase I leading and second species heights using the expert-matching ratio. Neither of the population means of these adjusted heights (28.02 and 26.80) are the same as population mean of Phase II leading species height (28.4). This example also shows that there is no reason to expect the estimates using the expert matched data are unbiased.

Appendix D: PPSWR Estimators: A review by Dr. Bill Warren

Although, on the surface, there would appear to be no problem with the estimators listed by CJS, namely the inflation, ratio and regression estimators, their application to the data provided raises some questions.

The estimates of the total volume (Y) and their estimated standard errors are as follows (Computational details are appended).

Stratum I Infl. $\hat{Y} = 149669 \cdot se(\hat{Y}) = 16750$

Ratio $\hat{Y} = 1523221 \cdot se(\hat{Y}) = 170907$

Regr. $\hat{Y} = 1485002 \cdot se(\hat{Y}) = 167190$

Stratum II Infl. $\hat{Y} = 2003805 \cdot se(\hat{Y}) = 208437$

Ratio $\hat{Y} = 1759665 \cdot se(\hat{Y}) = 214187$

Regr. $\hat{Y} = 1357285 \cdot se(\hat{Y}) = 231125$

The fact that the estimated standard error of the ratio estimate is greater than that of the inflation estimate can be readily explained (see below), but that the standard error of the regression estimate should be greater than that of the inflation estimate, in both strata, and greater than that of the ratio estimate in stratum II, is surprising. One would think that the result of fitting two parameters would have to be at least as good as that obtained when a single parameter is estimated.

1. **Notation.**

Before going further, I would like to clarify some notation. CJS uses $\hat{Y}_i$ to denote the ground-sample estimate of the total volume on polygon I; the ^ is used because it is an estimate, since the ground sample covers only a fraction of the area of the polygon. However, this could be confused with, e.g., $\hat{Y}_i = \hat{a} + \hat{b}X_i$, the regression estimate based on the auxiliary variable (the photo-interpreted polygon volume). Accordingly, I am going to use Y_i for the ground-estimated polygon total volume and reserve $\hat{Y}_i$ for the regression or other such estimate of that volume.

Unless Otherwise indicated, all summations will be over the sample (of size n = 10). Strata will be considered individually, so that no subscript indicating strata will be used.

2. Estimators.

The inflation estimator is given as:

$$Y = (1/n)\sum(Y_i / p_i)$$

with estimated variance:

$$Var(\hat{Y}) = \frac{1}{n}\frac{\sum(Y_i / p_i - \hat{Y})}{n-1}$$

where: $p_i = a_i/ A$ (a_i = the area of polygon i, $A=\sum^N a_i$)

Let me write: $y_i = Y_i / a_i$, i.e., the volume per hectare on polygon i, which is what is presented in the data listing. Then:

$$\hat{Y} = A\sum y_i / n = A\bar{y}$$

and

$$V(\hat{Y}) = \frac{1}{n}\cdot\frac{\sum\Big((AY_i / a_i) - (A\sum\big(\cancel{Y}_i / a_i\big)/ n)\Big)^2}{n-1}$$

$$= \frac{A^2}{n(n-1)}\cdot\sum\left(y_i - \bar{y}\right)^2$$

The ratio estimator is :

$$\hat{Y}_{rt} = \hat{R}X, with:$$

$$\hat{R} = \frac{\sum\left(Y_i / p_i\right)}{\sum\left(X_i / p_i\right)} = \frac{\sum\left(Y_i / a_i\right)}{\sum\left(X_i / a_i\right)} = \frac{\sum y_i}{\sum x_i}$$

The estimated variance of $\hat{Y}_{rt}$ is given as:

$$V(\hat{Y}) = \frac{\sum (e_i / p_i)^2}{n(n-1)}, where\_e_i = Y_i - \hat{R}X_i$$

Thus, $$\frac{e_i}{p_i} = \left(\frac{Y_i}{p_i} - \hat{R}\frac{X_i}{p_i}\right) = A\left(\frac{Y_i}{a_i} - \hat{R}\frac{X_i}{a_i}\right) = A(y_i - \hat{R}x_i) = A(y_i - \hat{y}_i)$$, and

$$V(\hat{Y}_{rt}) = \frac{A^2}{n(n-1)} \sum (y_i - \hat{y}_i)^2$$

Regression Estimator:

The regression estimator is:

$\hat{Y}_{rg} = N\hat{a} + \hat{b}X$, where[21]

$$\hat{b} = \left[(\frac{1}{n} \cdot \sum \frac{X_i Y_i}{p_i}) - \left(\frac{\hat{X}\hat{Y}}{(1/n)\sum(1/p_i)} \right) \right] / \left[\frac{1}{n} \cdot \sum \frac{X_i}{p_i}^2 - \frac{X^2}{(1/n)\sum(1/p_i)} \right]$$

$$\hat{a} = \frac{\left(Y - \hat{b}X\right)}{(1/n)\sum(1/p_i)}$$

$$\hat{Y} = \frac{1}{n} \cdot \sum \frac{Y_i}{p_i}, ...and\_.\hat{X} = \frac{1}{n} \cdot \sum \frac{X_i}{p_i}$$

---

[21] â is a regression intercept while a_i are polygon areas.

After some algebraic manipulation, we have:

$$\hat{b} = \frac{\left[\sum(1/n)\sum(a_i x_i y_i)\right] - \sum x_i \sum y_i}{\left[\sum(1/n)\sum\left(a_i {x_i}^2\right)\right] - (\sum x_i)^2}$$

$$\hat{a} = \frac{\left[\sum y_i - \sum x_i\right]}{\sum(1/a_i)}$$

The estimated variance of the regression estimator is given as:

$$V(\hat{Y}_{rg}) = \frac{\sum(e_i/p_i)^2}{n(n-1)}, where:-..e_i = Y_i - \hat{a} - \hat{b}X_i$$

Thus, $\dfrac{e_i}{p_i} = \left(\dfrac{Y_i}{p_i} - \dfrac{\hat{a}}{p_i} - \dfrac{\hat{b}X_i}{p_i}\right) = A\left(y_i - (\hat{a}/a_i) - \hat{b}x_i\right)$

Then $V(\hat{Y}_{rg}) = \dfrac{A^2}{n(n-1)} \cdot \sum(y_i - \hat{y}_i)^2, ...with...\hat{y}_i = \hat{a}/a_i + \hat{b}x_i$.

Thus, in all three cases, the estimated variance of Y can be expressed as:

$$V(\hat{Y}) = \frac{A^2}{n(n-1)} \cdot \sum(y_i - \hat{y}_i)^2$$

where $\hat{y} = \bar{y},\_or..\hat{R}x_i\ \_or..(\hat{a}/a_i) + \hat{b}x_i$; in the case of the inflation, ratio and regression estimators, respectively.

3. How can the inflation estimate have a smaller estimated variance than the ratio estimate?

In section 2 above, we see that the estimated variance can be expressed in terms of the sum of squares of the residuals ($y_i - \hat{y}_i$). In figures 1 and 2, the y_i are plotted against the x_i for strata I and II, respectively. The horizontal line represents the mean of the y_i values (y^{bar}) and the diagonal line $R^{hat}x_i$

One can see that in Stratum I there is little, if any, relationship between the y_i and x_i. The residual for point 3 and 5 for the ratio estimates far exceed, indeed are roughly double, those of the inflation estimates. For point 10, the residuals are approximately equal, albeit of opposite sign. The net result is that the sum of squares of the residuals of the ratio estimates far exceed that of the inflation estimates.

For stratum II, there is a somewhat better relationship between the y_i and x_i. The larger residual for the ratio estimate of point 10 is balanced by the smaller residuals of point 7 and 9. Visually, there seems no clear advantage, and the net result is a slightly higher sum of squares for the ratio-estimate residuals.

The failure of the ratio estimate to produce a smaller standard error than the inflation estimates is due, therefore, to the weak, and in the case of Stratum I, seemingly non-existent relationship between the y_i and x_i, i.e., between the ground and photo-interpreted polygon volumes per hectare.

4. Design-based, model-based and model-assisted inference.

CJS writes "In the ratio estimator, the estimated ratio between the Phase I and II total within the sample is used to adjust the Phase I population total. This is known as model assisted sampling – the presumed relationship between the Phase I and Phase II totals is used to improve the estimation process (Sarndal et al., 1992)".

> *Conquest writes "[In a model-assisted approach], ancillary data is made use of by specifying what is called a super-population model" between the ancillary variables and the response(s) of interest. The model is then used to either suggest a design or an estimator. In the estimation case, although the model suggests the estimators, the estimators are evaluated with design-based properties (Opsomer, et al, 2001). For example, regression estimation uses a covariate within the context of a linear model to obtain a more precise estimate of a total or a mean of a population.*
>
> *The regression estimator, although slightly biased, results in a much higher (design-based) precision than an estimator that does not take advantage of the ancillary data. In the design case, the model is used to suggest a probability sampling design that takes advantage of the relationship between the ancillary variables and the response. If the model holds, these strategies can help the sampler estimate parameters of construct optimal sampling designs."*

Thus, ratio and regression estimation can be viewed as model-assisted estimation. An important point is that, in the model-assisted approach, the properties of the estimators are from a design-based perspective. This is in contrast to a total model-based approach. To illustrate, consider the simple model $y_i = bx_i + e_i$. In terms of a model, the e_i could be considered to be independent random variables with zero expectation and variance $k_i\sigma^2$. Consider an estimator of b as

$$\sum w_i y_i / \sum w_i x_i = \sum w_i x_i b / \sum w_i x_i + (\sum w_i e_i / \sum w_i x_i) = b + \sum w_i e_i / \sum w_i$$

Now, under a purely model-based approach, expectations are taken with respect to the assumed underlying stochastic model. Accordingly, the estimator is unbiased, whatever the choice of the w_i. By using the assumed values of k_i, the w_i may be chosen to minimize the variance of the estimate.

Design-based expectations are formed as the average over all N^n (N_n^N if sampling without replacement) possible outcomes, appropriately weighted by the inclusion probabilities if something other than simple random sampling is employed. Only under special circumstances will an estimator be design unbiased, although the bias will commonly be small and even negligible.

Thus an estimator can be model unbiased, but design biased. Further, if the model is reasonably consistent with the data, the appropriately weighted estimate will have relatively small, if not minimum variance.

5. Estimator properties

The inflation estimator is a (design) unbiased estimator of $Y=\sum^{N} Y_i$. As a simple example, consider $N=3$, $n=2$. Since sampling is with replacement, there are 6 possible distinct outcomes, namely: Y_1Y_1, Y_1Y_2, Y_1Y_3, Y_2Y_2, Y_2Y_3, and Y_3Y_3, with probabilities p^2_1, $2p_1p_2$, $2p_2p_3$, p^2_2, $2p_2p_3$ and p^2_3, respectively. The possible estimates are, therefore:

$$Y_1/p_1,\ (Y_1/p_1+y_2/p_2)/2,\ (Y_1/p_1+Y_3/p_3)/2,\ Y_2/p_2,\ (Y_2/p_2+Y_3/p_3)/2 \text{ and } Y_3.p_3.$$

The expectation is then:

$$Y_1p_1+(Y_1p_2+Y_2p_1)+(Y_1p_3+Y_3p_1)+Y_2p_2+(Y_2p_3+Y_3p_2)+y_3p_3$$

$$=Y_1(p_1+p_2+p_3) + Y_2(p_1+p_2+p_3) + Y_3(p_1+p_2+p_3) \text{ (sum of all } p_i = 1)$$

$$=Y_1+Y_2+Y_3 = Y.$$

By considering $E\left(\hat{Y}^2\right) - E^2\left(\hat{Y}\right)$, it can be confirmed that the variance of $\hat{Y}$ is:

$$V(\hat{Y}) = \frac{1}{n}\sum_{i=1}^{N} p_i\left(Y_i / p_i - Y\right)^2$$

and that $V(\hat{Y}) = \frac{1}{n}\sum \frac{\left(Y_i / p_i - \hat{Y}\right)^2}{n-1}$

is a (design) unbiased estimator of $V(\hat{Y})$.

The ratio and regression estimators are not, however, design unbiased estimators of Y. Accordingly, it would be too much to expect that the $\sum(e_i/p_i)^2/n(n-1)$ would be unbiased estimators of $V(\hat{Y})$. Indeed, as noted by CJS, several estimators of $V(\hat{Y})$ have been suggested, of which the above is, perhaps, the most simple. CJS initially writes:

$$\sum((e_i/p_i - (1/n) \sum e_i/p_i))^2 / n(n-1)$$

but, in the case of the three estimators of Y so far considered, $\sum e_i/p_i=0$.

6. The general Linear Model

Consider the linear model:

$$\underline{Y} = \underline{a} + b\underline{X} + e, \quad or \quad \underline{Y} = A\underline{p} + \underline{e}$$

Assume E($\underline{e}$) = $\underline{0}$, V($\underline{e}$) – Vσ^2, where in particular V=D(v_i), where D denotes a diagonal matrix, i.e., the ei are independent random variables with expectation zero but non-homogeneous variance.

The matrix of independent variable(s) (A) is specified as:

$$A = \begin{vmatrix} 1 & 1 & 1 & \ldots & 1 \\ X_1 & X_2 & X_3 & \ldots & X_n \end{vmatrix}$$

The normal equations for estimating $\underline{b}$ are:

$$AV^{-1}A'e = AV^{-1}\underline{Y}$$

With, here, $V^{-1} = D(1/v_i)$. Thus:

$$\begin{vmatrix} \sum 1/v_i & \sum X_i/v_i \\ \sum X_i/v_i & \sum X_i^2/v_i \end{vmatrix} \begin{vmatrix} \hat{a} \\ b \end{vmatrix} = \begin{vmatrix} \sum Y_i/v_i \\ \sum X_i Y_i/v_i \end{vmatrix}$$

whence:

$$\hat{b} = \left[\sum \frac{1}{v_i} \sum \frac{X_i Y_i}{v_i} - \sum \frac{X_i}{v_i} \sum \frac{Y_i}{v_i}\right] \Big/ \left[\sum \frac{1}{v_i} \sum \frac{X_i}{v_i} - \left(\sum \frac{X_i}{v_i}\right)^2\right]$$

which, with a bit of algebraic manipulation, can be shown to be

$$\hat{b} = \left[\frac{1}{n} \sum \frac{X_i Y_i}{v_i} - \frac{\hat{X}\hat{Y}}{(1/n)\sum(1/v_i)}\right] \Big/ \left[\frac{1}{n} \sum \frac{X_i{}^2}{v_i} - \left(\frac{\hat{X}^2}{(1/n)\sum(1/v_i)}\right)\right]$$

where, $\hat{X} = \frac{1}{n}\sum \frac{X_i}{v_i}, ...and...\hat{Y} = \frac{1}{n}\sum \frac{Y_i}{v_i}$.

$$\hat{a} = \left(\sum \frac{Y_i}{v_i} - \hat{b}\sum \frac{X_i}{v_i}\right) / \sum (1/v_i) = \frac{\left(\hat{Y} - \hat{b}\hat{X}\right)}{(1/n)\sum (1/v_i)}$$

The regression estimator is equivalent to the GLM estimator with $V(e_i)$ proportional to the polygon areas, a_i.

Note that we are, in effect, regressing $Y_i/\sqrt{a_i}$ on on $1/\sqrt{a_i}$ and $X_i/\sqrt{a_i}$ (without intercept) and thus minimizing:

$$\sum \left(\frac{Y_i}{\sqrt{a_i}} - \frac{\hat{a}}{\sqrt{a_i}} - \frac{\hat{b}X_i}{\sqrt{a_i}} \right)^2$$

where as $V(\hat{Y})$ involves:

$$\sum \left(\frac{Y_i}{a_i} - \frac{\hat{a}}{a_i} - \frac{\hat{b}X_i}{a_i} \right)^2$$

Let us now restrict the general linear model to pass through the origin, i.e.,:

$\underline{Y} = b\underline{X} + \underline{e}$ with again, $V(\underline{e}) = D(v_i)\sigma^2$

The GLM estimator of b is then:

$$\hat{b} = \sum \left(\frac{X_i Y_i}{v_i} \right) / \sum \left(\frac{X_i^{\,2}}{v_i} \right)$$

Compare this with the ratio estimator:

$$\hat{R} = \sum\left(\frac{Y_i}{p_i}\right) / \sum\left(\frac{X_i}{p_i}\right)$$

The ratio estimator is thus equivalent to the GLM estimator with $v_i \propto a_iX_i$, i.e., the product of the polygon area and total photo interpreted volume.

Here, we are in effect, regressing Yi/√(aiXi) on √(Xi/a) (no intercept), and minimizing:

$$\sum\left(\frac{Y_i}{\sqrt{a_iX_i}} - \hat{b}\sqrt{X_ia_i}\right)^2$$

whereas V(Y) involves:

$$\sum\left(\frac{Y_i}{a_i} - \frac{\hat{b}X_i}{a_i}\right)^2$$

6.2. The selected estimator for the variance of Ŷ appears to be dictated by the sampling design, i.e., the pi and not the underlying assisting model, which assumes a different variance structure according to whether ratio or regression estimation is involved. It is known that ratio estimation generally performs well if the variability of the response variable increases with that of the auxiliary variable, whereas the regression estimator performs well when the variation is relatively constant over the range of data.

7. Alternatives

It would seem logical that, in a comparison of model-assisted estimators, the same model should underlie both. Accordingly, it would seem to be of interest to examine regression estimation under $v_i \propto a_iX_i$, and ratio estimation under vi prop a_i. We begin by considering the first mentioned. The normal equations are then:

$$\begin{vmatrix} \sum(1/a_iX_i) & \sum(1/a_i) \\ \sum(1/a_i) & \sum X_i/a_i \end{vmatrix} \begin{vmatrix} \hat{a} \\ b \end{vmatrix} =' \begin{vmatrix} \sum Y_i/a_iX_i \\ \sum Y_i/a_i \end{vmatrix}$$

As before, we have

$$Y_i = \hat{a} + \hat{b}X_i, ...y_i = \frac{Y_i}{a_i} = \frac{\hat{a}}{a_i} + \hat{b}x_i$$

and it turns out that $\sum(y_i - \hat{y}_i) = 0$.

$$\hat{Y} = N\hat{a} + \hat{b}X$$

and, as for V(Ŷ), we take:

$$V(\hat{Y}) = \frac{A^2}{n(n-1)} \cdot \sum(y_i - \hat{y}_i)^2 \, .$$

This leads to the following:

Strutum I Ŷ = 1364380, se(Ŷ) ≈ 163278 Sampling error = 23.9%

Strutum II Ŷ = 1605655, se(Ŷ) ≈ 191403 sampling error = 23.8%

Recall that the ratio estimates, under *D($a_i X_i$)* are:

Stratum I Ŷ = 1523221 se(Ŷ) ≈ 170907 Sampling error = 22.4%

Stratum II Ŷ = 1759665 se(Ŷ) ≈ 214187 sampling error = 24.3%.

In both strata, the estimated standard error of the regression estimate is less than that of the ratio estimate.

Next consider ratio estimation under $v_i \propto a_i$. Then,

$$\hat{b} = \frac{\sum(X_i Y_i / a_i)}{\sum(X_i^2 / a_i)}, ...and... \hat{Y} = \hat{b}X \, .$$

However, ∑(yi – Ŷi) ≠ 0. Accordingly, we take:

$$V(\hat{Y}) = \frac{A^2}{n(n-1)} \cdot \sum(y_i - \hat{y}_i)^2 - \left[\frac{1}{n} \cdot \left(\sum(y_i - \hat{y}_i)\right)^2\right]$$

to obtain 168782 and 196172 for stratum I and II, respectively.

While under D(ai), the regression estimate has a smaller estimated standard error than the ratio estimator in stratum I, the reverse holds in stratum II.

However, as noted above, we are in effect regressing Yi/√ai on 1/√ai and Xi/√ai. The residual sums of squares for the regression and ratio fits for stratum II are, respectively:

$$\sum\left(\frac{Y_i}{\sqrt{a_i}}-\frac{\hat{a}}{\sqrt{a_i}}-\frac{\hat{b}X_i}{\sqrt{a_i}}\right)^2 = 4482739 \text{ and}$$

$$\sum\left(\frac{Y_i}{\sqrt{a_i}}-\frac{\hat{b}X_i}{a_i}\right)^2 = 5470640$$

Thus the residual mean square for the regression is, as it should be, less than that for the ratio. It seems paradoxical, therefore, that V($\hat{Y}$) for the regression should be greater than V($\hat{Y}$) for the ratio estimate.

8. The GLM with b=0

Above we considered the GLM with a=0. There seems no reason why one could not also consider the case when b=0, i.e.,

$$Y_i = a + e_i$$

The normal equation is then:

$$\sum\left(\frac{1}{v_i}\right)\hat{a} = \sum\frac{Y_i}{v_i}, \ldots i.e.,$$

$$\hat{a} = \frac{\sum(Y_i/v_i)}{\sum(1/v_i)}, \ldots and$$

$$Y = \frac{N\sum(Y_i/v_i)}{\sum(1/v_i)}$$

The inflation estimator is:

$$Y = \frac{1}{n}\sum\left(Y_i / p_i\right) = \frac{A}{n}\left(Y_i / a_i\right) = A\frac{\sum y_i}{n}.$$

Seemingly, the only way that the inflation and GLM estimators can be made equivalent is if $v_i = p_i = 1/N = a'/A$, *if* $a_i = a'$, for all i.

Thus the inflation estimator does not appear to fit into the GLM framework. There is, of course, no auxiliary variable; however, there is a model assumption concerning $V(e_i)$

For completeness, I have examined the GLM with $v_i \propto a_i$ and $v_i \propto a_iX_i$; the results, however, do not appear to be credible and are not presented here.

9. Discussion

The regression and ratio estimators have been presented as model assisted in that, although suggested by some potential underlying model, the properties, e.g., expectations, are determined as if design-based. Accordingly, the estimators are biased, albeit slightly in most cases, but exact expression for the variance of such estimators is complicated. Various suggestions have been made for the variance estimators. The one presented in CJS takes the form:

$$\frac{A^2}{n(n-1)} \cdot \sum\left(y_i - \hat{y}_i\right)^2$$

possibly by analogy with the given estimator of the variance of the inflation estimate – which can also be expressed in the same form.

Application of these estimators to the provided data set results in some surprises, most notably that the estimated variance of the regression estimate of the population total is not always less than that of the ratio estimate, in spite of the fact that the regression estimate uses an additional parameter and, thus, its precision would be expected to be at least as good as that of the ratio estimate.

A closer examination reveals that the model underlying the regression estimate differs from that underlying the ratio estimate in more than just the additional parameter. Put in a general linear model context, we find that the regression estimate essentially assumes that residuals, the e_i in $Y_i = a + bX_i + e_i$ have variance proportional to the polygon areas, a_i, but those in the case of the ratio estimate would have variance proportional to the product of polygon area and the auxiliary variable, a_iX_i.

It would seem logical that a comparison of the ratio and regression would require the underlying model to be the same for each; however, analysis made under a common assumption, i.e., ratio and regression estimates under $v_i \propto a_i$ and again under $v_i \propto a_iX_i$ does not always eliminate the paradox.

In all three cases (inflation, ratio and regression) the estimated variance of the population total takes the form given above with $y_i = Y_i/a_i$ and $a_i/A = p_i$. The y_i is the estimate of Y_i/a_i by whatever estimation method is chosen, yet the expression for the variance estimate involves solely the p_i and contains nothing that would reflect the undoubtedly different precisions of the chosen y_i.

More specifically, we can write;

$$(y_i - \hat{y}_i) = \left(\frac{Y_i}{a_i} - \frac{\hat{Y}_i}{a_i}\right)$$

So that, in the variance estimator, we are looking at : $\sum\left(\frac{(Y_i - \hat{Y}_i)}{a_i}\right)^2$. However, if $vi \propto a_i$, we are minimizing:

$$\sum\left(\frac{(Y_i - \hat{Y}_i)}{\sqrt{a_i}}\right)^2$$

and if $v_i \propto a_i X_i$, we would be minimizing:

$$\sum\left(\frac{(Y_i - \hat{Y}_i}{\sqrt{a_i X_i}}\right)^2$$

Now, the more precise an estimate of yi the smaller the difference (y_i - $\hat{y}_i$), and one would think, the smaller the estimated variance. But it is clear that this does not always happen and gives rise to speculation about the viability of the given variance estimator.

For stratum I, the two ratio and two regression estimates range from c. 1,357,000 to 1,523,000 with estimated standard errors ranging from c. 163,000 to 171000. The inflation estimate is c. 1,492,000 with estimated standard error c. 117,000.

These estimates of Y are similar, probably because of the lack of a relationship between the ground and photo-interpreted volumes, at least as demonstrated by the n points of the sample. The reason for the smaller standard error of the inflation estimate has been illustrated in Figure 1. The picture could change, however, with a different sample.

For Stratum II there is less consistency, the ratio and regression estimates ranging from c. 1,357,000 to 1760,000 with standard errors ranging from c. 191,000 to 231,000. The inflation estimate is c. 2,004,000 with a standard error of c. 208,000. The auxiliary variable has apparently reduced the estimate of the population total, but with, at best, a marginal reduction in standard error.

If these results were typical[22], one would have to question the merit of including the auxiliary variable.

---

[22] Apparently, the data is a mock-up for illustration purposes. It was not impossible to simulate a true PPSWR sample with ground measurements from a small population of 200 polygons as a practical example.

Margaret Penner has suggested three analyses:

(i) Create one regression for all strata combined, then use this regression to estimate a population total volume.

In comparing separate and combined regressions Cochran writes:

> *"Hard and fast rules cannot be given to decide whether the separate or combined estimate is better in any specific situation: some exercise of judgement is required in making a choice. The defects of the separate estimate are that it is more liable to bias when samples are small within the individual strata and that its variance has a larger contribution from sampling errors in the regression coefficients.*
>
> *The defect of the combined estimate is that its variance is inflated if the population regression coefficients differ from stratum to stratum. If we are confident that the regressions are linear and if β_h appears to be the same in all strata, so far as can be judged, the combined estimate is to be preferred.*
>
> *If the regression appears linear (so that the danger of bias seems small) but β_h seems to vary from stratum to stratum, the separate estimate is advisable."*

[This was written in relation to sampling without replacement, but the same considerations would no doubt apply in sampling with replacement.]

Similar argument applies to the separate and combined ratio estimates.

With the data in hand, a common regression could be applicable if for no other reason than that the relationships are so weak.

(ii) In a separate step, estimate strata totals, using a ratio estimator.

Given that, as model-assisted estimators, the ratio and regression estimators, even within a stratum, are based on different models, their comparison is questionable. Accordingly, I would be most unwilling to compare a total based on separate ratio estimates with a combined regression estimate.

(iii) Constrain the sum of the strata totals to be equal to the regression based overall total from (i).

I have difficulty in seeing the point to this. If there is a need to estimate strata totals, then the use of a separate ratio estimator would seem appropriate. One could consider the combined estimator if the relationships were the same. There would be a slight complication in estimating the variance of the individual total in that the $\sum(e_i / p_i)$ for an individual stratum would not then, in general, $= 0$.

The bottom line is, I think, that the relationships in each data set need to be examined, and an estimator appropriate to the situation, including the objectives, be selected.

10 Further Observations

CJS includes a section on geometric mean regression estimators, on the grounds that both X and Y are subject to variation about the true value. With X there is no sampling error, *per se*; observations are made on all sample units (polygons) in the population(?). While there may also be some random error in photo interpretation, the concern is, I believe, more towards bias in photo interpretation, and the ground samples are used to adjust for that bias. Unfortunately, the ground samples are subject to sampling error since, in general, only one cluster of sample plots is used to represent the whole polygon (so that there is good reason why the variance of the e_i should increase with polygon area as well as with volume). If this error were random, i.e., evenly scattered about zero, and sufficient polygons were sampled, regression of *Y* on *X* should be effective in adjusting for photo interpretation bias. Since the objective is the estimation of *Y* for a given *X*, rather than estimating the relationship between two random variables, least-squares regression of *Y* on *X* would seem preferable to geometric mean regression.

The tighter the relationship between the *X* and *Y*, the more effective regression (or ratio) estimation will be. This would presumably occur if Y could be made a more accurate representation of what the polygon contained. Thus, it may be advantageous to increase the number of sample clusters within a polygon, but at the same time, for cost considerations, reduce the number of polygons sampled. Also, since the *X* are known for all polygons, their total is known without sampling error. The important thing is then to estimate as accurately as possible the relationship between *X* and *Y*. In this respect, adherence to a probability based sampling design may not be critical; there may be a role for purposive sampling. Note: purposive sampling is not the same as subjective sampling.

APPENDICES

APPENDIX 1

Application

1. The inflation Estimator

$$\hat{Y} = \frac{1}{n}\cdot\sum\frac{Y_i}{p_i} = \frac{A}{n}\cdot\sum\frac{Y_i}{a_i} = \frac{A}{n}\sum y_i$$

$$V(\hat{Y}) = \frac{1}{n}\sum\left(\frac{Y_i / p_i - \hat{Y}}{n-1}\right) = \frac{A^2}{n(n-1)}\sum\left(\frac{Y_i}{a_i} - \left(\frac{1}{n}\sum\frac{Y_i}{a_i}\right)\right)^2$$

$$= \frac{A^2}{n(n-1)}\sum\left(y_i - \bar{y}\right)^2 = \frac{A^2}{n(n-1)}\left[\sum y_i - \left(\frac{(\sum y_i)^2}{n}\right)\right]$$

Stratum I:

$$\hat{Y} = \frac{3669}{10}\cdot 4065.6 = 1{,}491{,}668.6$$

$$V(\hat{Y}) = \frac{3669^2}{90}\cdot 91130.71 = (116750.5)^2$$

Stratum II:

$$\hat{Y} = \frac{4182}{10}\cdot 4791.5 = 2003805.3$$

$$V(\hat{Y}) = \frac{4182^2}{90}\cdot 223576.19 = (208437.4)^2$$

2. Ratio Estimator:

$$\hat{Y} = \frac{\sum(Y_i / p_i)}{\sum(X_i / p_i)} \cdot X = \frac{\sum(Y_i / a_i)}{\sum(X_i / a_i)} \cdot X = \frac{\sum y_i}{\sum x_i} \cdot X, \ldots \hat{R} = \frac{\sum y_i}{\sum x_i}$$

$$V(\hat{Y}) = \frac{A^2}{n(n-1)} \sum \left(y_i - \hat{R} x_i\right)^2$$

Stratum I

$$\hat{Y} = \frac{4065.6}{3618.6} \cdot 1355747.3 = 1{,}523{,}220.6$$

$$V(\hat{Y}) = \frac{3669^2}{90} \cdot 195285.56 = (170907.7)^2$$

Stratum II:

$$\hat{Y} = \frac{4791.5}{5299.9} \cdot 1946374.1 = 1759665.6$$

$$V(\hat{Y}) = \frac{4182^2}{90} \cdot 236080.79 = (214{,}187.0)^2$$

3. Regression Estimator

$$\hat{Y} = N\hat{a} + \hat{b}X$$

$$\left|\begin{array}{cc} \sum(1/a_i) & \sum X_i/a_i \\ \sum X_i/a_i & \sum X_i2/a_i \end{array}\right| \left|\begin{array}{c} \hat{a} \\ b \end{array}\right| \left|\begin{array}{c} \sum Y_i/a_i \\ \sum X_iY_i/a_i \end{array}\right|$$

$$\text{or} \left|\begin{array}{cc} \sum 1/a_i & \sum x_i \\ \sum x_i & \sum a_ix_i \end{array}\right| \left|\begin{array}{c} \hat{a} \\ b \end{array}\right| \left|\begin{array}{c} \sum y_i \\ \sum a_ix_iy_i \end{array}\right|$$

$$V(\hat{Y}) = \frac{A^2}{n(n-1)} \cdot \sum \left(y_i - \hat{a} / a_i - \hat{b} x_i\right)^2$$

Stratum I:

$$\begin{vmatrix} 0.657082 & 3618.6 \\ 3618.6 & 33534377 \end{vmatrix} \begin{vmatrix} \hat{a} \\ b \end{vmatrix} = \begin{vmatrix} 4065.6 \\ 36988497 \end{vmatrix}$$

b = 1.0729391, â = 278.59608

$\hat{Y}$ = 109 x â + b x 1946374.1 = 1357284.8

$$V(\hat{Y}) = \frac{3669^2}{90} \cdot 186883.09 = (167190.4)^2$$

Stratum II:

$$\begin{vmatrix} 1.3922721 & 5299.9 \\ 5299.9 & 66973571 \end{vmatrix} \begin{vmatrix} \hat{a} \\ b \end{vmatrix} \begin{vmatrix} 4791.5 \\ 48160536 \end{vmatrix}$$

b = 0.6393545, â = 1007.6945

$\hat{Y}$ = 112 · â + b · 1946374.1 = 1357284.8

$$V(\hat{Y}) = \frac{4182^2}{90} \cdot 274894.97 = (251124.6)^2$$

4. **Regression Estimator, where: $v_i \propto X_i$**

$$\hat{Y} = N\hat{a} + \hat{b}X$$

$$\begin{vmatrix} \sum(1/a_iX_i) & \sum X_i/(a_iX_i) \\ \sum X_i/(a_iX_i) & \sum X_i^2/(a_iX_i) \end{vmatrix} \begin{vmatrix} \hat{a} \\ b \end{vmatrix} =' \begin{vmatrix} \sum Y_i/(a_iX_i) \\ \sum X_iY_i/(a_iX_i) \end{vmatrix}$$

$$\text{or} \begin{vmatrix} \sum(1/a_i^{2}{}_{xi}) & \sum(1/a_i) \\ \sum(1/a_i) & \sum x_i \end{vmatrix} \begin{vmatrix} \hat{a} \\ b \end{vmatrix} =' \begin{vmatrix} \sum y_i/a_ix_i \\ \sum y_i \end{vmatrix}$$

Stratum I

$$\begin{vmatrix} 1.6707294\text{x}10^{-4} & 0.6570802 \\ 0.6570802 & 3618.6 \end{vmatrix} \begin{vmatrix} \hat{a} \\ b \end{vmatrix} = \begin{vmatrix} 0.793546 \\ 4065.6 \end{vmatrix}$$

b = 0.9132753, â = 1157.8828

$$\hat{Y} = 109\hat{a} + \hat{b}1355747.3 = 1364379.7$$

$$V(\hat{Y}) = \frac{3669^2}{90} \cdot 198239.76 = (163278.4)^2$$

Stratum II

$$\begin{vmatrix} 1.6570762\text{x}10^{-3} & 1.3922721 \\ 1.3922721 & 5299.9 \end{vmatrix} \begin{vmatrix} \hat{a} \\ b \end{vmatrix} =' \begin{vmatrix} 0.793546 \\ 4065.6 \end{vmatrix}$$

b = 0.7872159, â = 444.8371

$$\hat{Y} = 112\hat{a} + \hat{b} \cdot 1946374.1 = 1605654.9$$

$$V(\hat{Y}) = \frac{4182^2}{90} \cdot 188525.99 = 1914028$$

5. Ratio Estimator $v_i \propto a_i$

$$\hat{R} = \frac{\left(\sum X_i Y_i / a_i\right)}{\left(\sum X_i^{\,2} / a_i\right)}$$

Stratum I:

$$\hat{R} = 36988497/33534377 = 1.1030023$$

$$\hat{Y} = \hat{R}X = 1355747.3 \cdot \hat{R} = 1495392.4$$

$$\sum \left(y_i - \hat{y}_i\right) \neq 0$$

$$V(\hat{Y}) = \frac{A^2}{n(n-1)} \cdot \left[\sum \left(y_i - \hat{y}_i\right)^2 - \frac{1}{n}\left(\sum (y_i - \hat{y}_i)\right)^2\right]$$
$$= \frac{3669^2}{90} \cdot \left[191006.79 - \frac{1}{10} \cdot 74276^2\right] = 168{,}780.7^2$$

Stratum II

$$\hat{R} = 48160536/66973571 = 0.7190976$$

$$\hat{Y} = 1946374.1 \cdot \hat{R} = 1{,}399{,}633$$

$$V(\hat{Y}) = \frac{4182^2}{90} \cdot \left[294146.67 - \frac{1}{n} \cdot 980.35^2\right] = 196{,}172.0^2$$

APPENDIX 2

Construction of example with N=3, n=2.

$$V(\hat{Y}) = E(\hat{Y}^2) - E^2(\hat{Y})$$

$$E(\hat{Y}^2) = p_1{}^2(Y_1 p_1)^2 + p_2{}^2(Y_2 p_2)^2 + p_3{}^2(Y_3 / p_3)^2 + \frac{2p_1p_2}{4}(Y_1/p_1 + Y_2/p_2) +$$

$$\frac{2p_1p_2}{2}(Y_1/p_1 + Y_3/p_3) + 2p_2p_3(Y_2/p_2 + Y_3/p_3)$$

$$= Y_1{}^2 + Y_2{}^2 + Y_3{}^2 + \frac{1}{2}\frac{Y_1{}^2}{p_1}(p_2 + p_3) + \frac{1}{2}\frac{Y_2{}^2}{p_2}(p_1p_3) + \frac{1}{2}\frac{Y_3{}^2}{p_3}(p_1p_2) + Y_1Y_2 + Y_1Y_3 + Y_2Y_3$$

$$= Y_1{}^2 + Y_2{}^2 + Y_3{}^2 + \frac{1}{2}\frac{Y_1{}^2}{p_1}(1 - p_1) + \frac{1}{2}\frac{Y_2{}^2}{p_2}(1 - p_2) + \frac{1}{2}\frac{Y_3{}^2}{p_3}(1 - p_3) + Y_1Y_2 + Y_1Y_3 + Y_2Y$$

$$= \frac{1}{2}\left(Y_1{}^2 + Y_2{}^2 + Y_3{}^2\right) + \frac{1}{2}\left(\frac{Y_1{}^2}{p_1} + \frac{Y_2{}^2}{p_2} + \frac{Y_3{}^2}{p_3}\right) + Y_1Y_2 + Y_1Y_3 + Y_2Y$$

$$= \frac{1}{2}\left(\frac{Y_1{}^2}{p_1} + \frac{Y_2{}^2}{p_2} + \frac{Y_3{}^2}{p_3}\right) + \frac{1}{2}\left(Y_1 + Y_2 + Y_3\right)^2$$

$$V(\hat{Y}) = E(Y^2) - \left(Y_1 + Y_2 + Y_3\right)^2 = \frac{1}{2}\left(\frac{Y_1{}^2}{p_1} + \frac{Y_2{}^2}{p_2} + \frac{Y_3{}^2}{p_3}\right) - \frac{1}{2}\left(Y_1 + Y_2Y_3\right)^2$$

Now:

$$\frac{1}{2}\left[p_1\left(\frac{Y_1}{p_1}-Y\right)^2+p_2\left(\frac{Y_2}{p_2}-Y\right)^2+p_3\left(\frac{Y_3}{p_3}-Y\right)^2\right]$$

$$=\frac{1}{2}\left[p_1\left(\frac{Y_1^2}{p_1^2}-2\frac{Y_1Y}{p_1}+Y^2\right)+p_2\left(\frac{Y_2^2}{p_2^2}-2\frac{Y_2Y}{p_2}+Y^2\right)+p_3\left(\frac{Y_3^2}{p_3}-2\frac{Y_3Y}{p_3}+Y^2\right)\right]$$

$$=\frac{1}{2}\left[\frac{Y_1^2}{p_1}+\frac{Y_2^2}{p_2}+\frac{Y_3^2}{p_3}-2Y(Y_1+Y_2+Y_3)+(p_1+p_2+p_3)Y^2\right]$$

$$=\frac{1}{2}\left[\frac{Y_1^2}{p_1}+\frac{Y_2^2}{p_2}+\frac{Y_3^2}{p_3}-Y^2\right],\_i.e.\_...$$

$$V(\hat{Y})=\sum_{i=1}^{3}p_i\left(\frac{Y_i}{p_i}-Y\right)^2$$

APPENDIX 3

$$E\left[\frac{1}{n}\sum\left(\frac{Y_i}{p_i}-\hat{Y}\right)^2\right],\_again\_N=3,...n=2$$

In the case of the sample being Y1,Y1 (or Y2,Y2; or Y3Y3).

$$\hat{Y}=\frac{1}{2}\left(\frac{Y_1}{p_1}+\frac{Y_1}{p_1}\right)=\frac{Y_1}{p_1},\quad so\quad that\quad \frac{Y_i}{p_i}-\hat{Y}=0$$

Thus

$$E\left[\frac{1}{n}\sum\left(\frac{Y_i}{p_i}-\hat{Y}\right)^2\right]$$

$$=2p_2p_2\left[\frac{1}{2}\left(\frac{Y_1}{p_1}-\frac{1}{2}\left(\frac{Y_1}{p_1}+\frac{Y_2}{p_2}\right)\right)^2+\frac{1}{2}\left(\frac{Y_2}{p_2}-\frac{1}{2}\left(\frac{Y_1}{p_1}+\frac{Y_2}{p_2}\right)\right)^2\right]+2p_1p_2[\ldots]+2p_2p_3[\ldots]$$

$$=\left[\left(\frac{1}{2}\frac{Y_1^{\,2}}{p_1^{\,2}}+\frac{1}{2}\frac{Y_2^{\,2}}{p_2^{\,2}}\right)-\frac{Y_1Y_2}{p_1p_2}\right]p_1p_2+[\ldots]p_1p_3+[\ldots]p_2p_3$$

$$=\frac{1}{2}\frac{Y_1^{\,2}}{p_1}(p_2p_3)+\frac{1}{2}\frac{Y_2^{\,2}}{p_2}(p_1p_3)+\frac{1}{2}\frac{Y_3^{\,2}}{p_3}(p_2p_3)-Y_1Y_2-Y_1Y_3-Y_2Y_3$$

$$=\frac{1}{2}\frac{Y_1^{\,2}}{p_1}(1-p_1)+\frac{1}{2}\frac{Y_2^{\,2}}{p_2}(1-p_2)+\frac{1}{2}\frac{Y_3^{\,2}}{p_3}(1-p_3)-Y_1Y_2-Y_1Y_3-Y_2Y_3$$

$$=\frac{1}{2}\left[\frac{Y_1^{\,2}}{p_1}+\frac{Y_2^{\,2}}{p_2}+\frac{Y_3^{\,2}}{p_3}-Y_1^{\,2}-Y_2^{\,2}-Y_3^{\,2}-2Y_1Y_2-2Y_1Y_3-2Y_2Y_3\right]$$

$$=\frac{1}{2}\left[\frac{Y_1^{\,2}}{p_1}+\frac{Y_2^{\,2}}{p_2}+\frac{Y_3^{\,2}}{p_3}-Y^2\right]$$

Thus: $E\left[\frac{1}{2}\sum\left(\frac{Y_i}{p_i}-\hat{Y}\right)^2\right]=\frac{1}{2}\sum p_i\left(\frac{Y_i}{p_i}-Y\right)^2=V(\hat{Y})$