



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO CIENCIAS FORESTALES

DESARROLLO DE TÉCNICAS ESTADÍSTICAS PARA LA EVALUACIÓN Y MONITOREO DE ECOSISTEMAS FORESTALES EN MÉXICO

EFRAÍN VELASCO BAUTISTA

TESIS

**PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:**

DOCTOR EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO

2015

La presente tesis titulada “**Desarrollo de técnicas estadísticas para la evaluación y monitoreo de ecosistemas forestales en México**”, realizada por el alumno Efraín Velasco Bautista, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

DOCTOR EN CIENCIAS FORESTALES

CONSEJERO Y DIRECTOR DE
TESIS



DR. HÉCTOR MANUEL DE LOS SANTOS
POSADAS

ASESOR



DR. JOSÉ RENÉ VALDEZ LAZALDE

ASESOR



DR. GILBERTO RENDÓN SÁNCHEZ

ASESOR



DR. HUGO RAMÍREZ MALDONADO

ASESOR



DR. MIGUEL ACOSTA MIRELES

Montecillo, Texcoco, Estado de México, julio de 2015.

Dedicatoria

A todos los actores del sector forestal interesados en la búsqueda de Métodos Cuantitativos Modernos para la Evaluación y Monitoreo de los Ecosistemas Forestales de México.

Al Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), cuyo principio básico es “generar conocimientos científicos e innovaciones tecnológicas para contribuir al desarrollo productivo, competitivo, equitativo y sustentable del sector forestal, agrícola y pecuario en beneficio de la sociedad mexicana”.

A todos los investigadores forestales-estadísticos que han dedicado gran parte de su vida en generar conocimientos científicos sobre la teoría de inventarios forestales de largo plazo, los cuales fueron la base de las ideas plasmadas en esta investigación.

A mi familia, por apoyarme en todo momento.

Sinceramente

Efraín Velasco Bautista

Agradecimientos

Al Colegio de Postgraduados por haberme dado la oportunidad de continuar con mi preparación académica. De manera especial extendiendo mi gratitud a los profesores del Postgrado en Ciencias Forestales y en Estadística que contribuyeron en mi formación doctoral.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por haber financiado mis estudios doctorales.

Al Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), particularmente al Centro Nacional de Investigación Disciplinaria en Conservación y Mejoramiento de Ecosistemas Forestales (CENID COMEF), por permitir que siguiera con mi preparación científica en el ámbito de la Biometría Forestal.

A la Comisión Nacional Forestal (CONAFOR), especialmente a la Gerencia de Inventario Forestal y Geomática, por haber proporcionado las bases de datos para que el desarrollo de la presente investigación fuera posible.

Al Dr. Héctor M. de los Santos Posadas, mi profesor consejero, por avanzar conmigo en la búsqueda de métodos cuantitativos modernos relacionados a la Evaluación y Monitoreo de Ecosistemas Forestales; y por la motivación brindada en los momentos difíciles que la vida me presentó.

A los integrantes de mi Consejo Particular, por estar atentos en los avances de la investigación, por sus valiosas enseñanzas y sugerencias, pero sobre todo por su amistad. Dr. Héctor Manuel de los Santos Posadas, Dr. José René Valdez Lazalde, Dr. Gilberto Rendón Sánchez, Dr. Hugo Ramírez Maldonado y Dr. Miguel Acosta Mireles, siempre estaré agradecido con Ustedes.

*Sinceramente
Efraín Velasco Bautista*

Contenido

LISTA DE CUADROS	viii
LISTA DE FIGURAS	ix
DESARROLLO DE TÉCNICAS ESTADÍSTICAS PARA LA EVALUACIÓN Y MONITOREO DE ECOSISTEMAS FORESTALES EN MÉXICO	xi
RESUMEN	xi
DEVELOPMENT OF STATISTICAL TECHNIQUES FOR THE EVALUATION AND MONITORING OF FOREST ECOSYSTEMS IN MEXICO	xii
ABSTRACT	xii
CAPÍTULO I	1
ASPECTOS GENERALES SOBRE LA INVESTIGACIÓN	1
1.1. INTRODUCCIÓN	1
1.2. PLANTEAMIENTO DEL PROBLEMA	3
1.3. OBJETIVOS	4
1.4. REVISIÓN DE LITERATURA	5
1.5. MATERIALES Y MÉTODOS	10
1.6. CONCLUSIONES	10
1.7. LITERATURA CITADA	11
CAPÍTULO II	14
COMPARACIÓN DE ESTIMADORES PARA PARÁMETROS DE INTERÉS FORESTAL EN SELVA MEDIANA DE MÉXICO	14
2.1. RESUMEN	14
ESTIMATORS COMPARISON FOR FOREST PARAMETERS IN MEDIUM RAINFOREST OF MEXICO	15
2.2. SUMMARY	15
2.3. INTRODUCCIÓN	16
2.4. OBJETIVO	17
2.5. REVISIÓN DE LITERATURA	17
2.5.1. Diseño del conglomerado y zona de inclusión	17
2.5.2. Notación	22
2.5.3. Estimadores	23
2.6. MATERIALES Y MÉTODOS	34
2.7. RESULTADOS Y DISCUSIÓN	37
2.8. CONCLUSIONES	51
2.9. LITERATURA CITADA	53
ANEXO 2.1. ANÁLISIS DEL SESGO DEL ESTIMADOR DE RAZÓN EN MUESTREO ALEATORIO SIMPLE (adaptación de Pérez, 2000)	55
ANEXO 2.2. LISTA DE ESPECIES REPRESENTATIVAS DE SELVA MEDIANA SUBPERENNIFOLIA Y SUBCADUCIFOLIA	59
CAPÍTULO III	60
COMPARACIÓN DE ESTIMADORES PARA PARÁMETROS FORESTALES EN UNA POBLACIÓN SIMULADA	60
3.1 RESUMEN	60
ESTIMATORS COMPARISON FOR FOREST PARAMETERS IN A SIMULATED POPULATION	61
3.2. SUMMARY	61
3.3. INTRODUCCIÓN	62
3.4. OBJETIVO	63
3.5. REVISIÓN DE LITERATURA	63
3.5.1. Derivación de estimadores	63
3.5.1.1. Notación	63
3.5.1.2. Modelo general	64
3.5.1.3. Modelo Forest Inventory and Analysis	65
3.5.1.4. Modelo Van Deusen	67

3.5.1.5. Modelo media de razones.....	68
3.5.1.6. Estimador razón de medias	69
3.5.1.7. Relación entre estimadores.....	71
3.6. MATERIALES Y MÉTODOS.....	75
3.7. RESULTADOS Y DISCUSIÓN	83
3.7.1. Comparación de conglomerados de subparcelas circulares y rectangulares	83
3.7.2. Comparación de estimadores por variable y condición.....	85
3.7.2.1. Condición c0 y variable densidad (árboles ha ⁻¹)	86
3.7.2.2. Condición c1 y variable densidad (árboles ha ⁻¹)	87
3.7.2.3. Condición c2 y variable densidad (árboles ha ⁻¹)	89
3.7.2.4. Condición c0 y variable volumen (m ³ ha ⁻¹).....	90
3.7.2.5. Condición c1 y variable volumen (m ³ ha ⁻¹).....	92
3.7.2.6. Condición c2 y variable volumen (m ³ ha ⁻¹).....	93
3.8. CONCLUSIONES	98
3.9. LITERATURA CITADA.....	99
ANEXO 3.1. DERIVACIÓN DE ESTIMADORES: CASO PARTICULAR DOS OBSERVACIONES	101
ANEXO 3.2. DERIVACIÓN DE LA VARIANZA ESTIMADA DE RAZÓN A PARTIR DE LA VARIANZA ESTIMADA DE VAN DEUSEN.....	107
ANEXO 3.3. ESTIMACIONES PUNTUALES Y POR INTERVALO DE 88 MUESTRAS. VARIABLE DENSIDAD EN TRES CONDICIONES DIFERENTES.....	108
ANEXO 3.4. ESTIMACIONES PUNTUALES Y POR INTERVALO DE 88 MUESTRAS. VARIABLE VOLUMEN EN TRES CONDICIONES DIFERENTES.	110
CAPÍTULO IV	112
ESTIMACIÓN DE PARÁMETROS FORESTALES EN EL CONTEXTO DE DISEÑO DE MUESTRAS PANELIZADAS DEL INVENTARIO NACIONAL FORESTAL Y DE SUELOS DE MÉXICO (INFyS)	112
4.1. RESUMEN	112
ESTIMATION OF FOREST PARAMETERS IN THE CONTEXT OF THE PANELIZED SAMPLES DESIGN OF THE NATIONAL FOREST AND SOILS INVENTORY OF MEXICO (INFyS).....	113
4.2. SUMMARY	113
4.3. INTRODUCCIÓN	114
4.4. OBJETIVOS	115
4.5. REVISIÓN DE LITERATURA	115
4.5.1. Estimadores del muestreo con reemplazo total y remediación total	116
4.5.2. Muestreo con reemplazo parcial en dos ocasiones mediante estimadores de regresión	119
4.5.2.1. Estimador compuesto de $T_y(t_2)$	120
4.5.2.2. Estimador compuesto de $T_y(t_1)$	127
4.5.3. Muestreo con reemplazo parcial en dos ocasiones mediante Mínimos Cuadrados Generalizados.....	131
4.5.4. Muestreo con reemplazo parcial en múltiples ocasiones vía Mínimos Cuadrados Generalizados.....	136
4.5.4.1 Propiedades del estimador Mínimos Cuadrados Generalizados	144
4.5.5. Estimadores del Muestreo Replicado	145
4.5.6. Estimador Media Móvil	146
4.5.7. Estimador Compuesto Univariado	147
4.5.8. Resumen de estimadores	154
4.5.9. Estimación del cambio	155
4.6. MATERIALES Y MÉTODOS.....	159
4.7. RESULTADOS Y DISCUSIÓN	167
4.7.1. Estimación de ABHA en cada ocasión de medición.....	167
4.7.2. Estimación de ABHA en la ocasión más reciente.....	182
4.7.3. Estimación de cambio de ABHA entre puntos medios de periodo.....	185
4.7.4. Estimación de la tasa de cambio de ABHA entre puntos medios de periodo	189
4.8. CONCLUSIONES	193
4.9. LITERATURA CITADA.....	195

ANEXO 4.1. DETERMINACIÓN DE a Y d (EXPRESION (4.15)). MUESTREO CON REEMPLAZO PARCIAL EN DOS OCASIONES	198
ANEXO 4.2. PROPIEDADES DE MÍNIMOS CUADRADOS GENERALIZADOS	206
ANEXO 4.3. DESARROLLO DEL PRODUCTO DE DOS FUNCIONES DE DENSIDAD NORMAL	212
ANEXO 4.4. ESTIMACIONES DE ABHA OBTENIDAS MEDIANTE MCG, MRE Y ECU EN LA CONDICIÓN UNO	215
ANEXO 4.5. ESTIMACIONES DE ABHA OBTENIDAS MEDIANTE MCG, MRE Y ECU EN LA CONDICIÓN DOS.....	219
ANEXO 4.6. INTERVALOS DE CONFIANZA AL 95% PARA ABHA GENERADOS POR ECU PARA TODAS LAS OCASIONES DE MEDICIÓN EN LA CONDICIÓN UNO	223
ANEXO 4.7. INTERVALOS DE CONFIANZA AL 95% PARA ABHA GENERADOS POR ECU PARA TODAS LAS OCASIONES DE MEDICIÓN EN LA CONDICIÓN DOS	227
CAPÍTULO V.....	231
CONCLUSIONES Y RECOMENDACIONES GENERALES.....	231

LISTA DE CUADROS

Cuadro 2.1. Coeficientes de correlación de Pearson entre las variables dasométricas y el área del conglomerado efectivamente muestreada.....	37
Cuadro 2.2. Estadísticos de ajuste y coeficientes de regresión estimados para modelos lineales simples de variables dasométricas (y) sobre área de conglomerado (a). Modelo $y = \beta_0 + \beta_1 a$	43
Cuadro 2.3. Estimaciones puntuales y por intervalo de parámetros forestales en selva mediana subperennifolia y subcaducifolia en Quintana Roo bajo diferentes estimadores.....	44
Cuadro 2.4. Estimaciones puntuales y por intervalo de parámetros forestales en selva mediana subperennifolia y subcaducifolia en Campeche bajo diferentes estimadores.....	45
Cuadro 2.5. Estimación del número de árboles y volumen por hectárea en Orova, Slovakia (Smelko y Saborowski,1999).....	47
Cuadro 3.1. Valores de t y p de la prueba de t pareada univariada para comparar poblaciones de conglomerados de subparcelas circulares y rectangulares.....	84
Cuadro 3.2. Valores de F y p de la prueba de t pareada bivariada (densidad y volumen simultáneamente) para comparar poblaciones de conglomerados de subparcelas circulares y rectangulares.....	84
Cuadro 3.3. Sesgos relativos de la estimación y sesgos relativos de la varianza para cuatro estimadores estudiados en una población forestal artificial.....	85
Cuadro 4.1. Pesos considerados en la aplicación de media móvil.....	166
Cuadro 4.2. Estimación de área basal por hectárea (ABHA) (m^2ha^{-1}) mediante Mínimos Cuadrados Generalizados, Muestreo Replicado y Estimador Combinado Univariado bajo la condición uno (tamaño de muestras panelizadas de 1000).....	168
Cuadro 4.3. Estimación de área basal por hectárea (ABHA) (m^2ha^{-1}) mediante Mínimos Cuadrados Generalizados, Muestreo Replicado y Estimador Combinado Univariado bajo la condición dos (tamaño de muestras panelizadas de 1500).....	169
Cuadro 4.4. Estimación de área basal por hectárea (ABHA) (m^2ha^{-1}) en 2012 mediante MCG, MRE y MM bajo dos condiciones diferentes.....	183
Cuadro 4.5. Estimaciones del cambio de ABHA entre periodos para las condiciones uno y dos.....	185
Cuadro 4.6. Estimaciones de la tasa de cambio de ABHA entre periodos para las condiciones uno y dos.....	190

LISTA DE FIGURAS

Figura 1.1. Unidades de Muestreo Primarias (UMP) y Secundarias (UMS) para el Inventario Nacional Forestal y de Suelos en bosques y vegetación de zonas áridas.....	7
Figura 1.2. Unidades de Muestreo Primarias (UMP) y Secundarias (UMS) para el Inventario Nacional Forestal y de Suelos en selvas.....	7
Figura 2.1. Diseño del conglomerado que se utiliza en el Inventario Nacional Forestal y de Suelos: bosques a) y selvas b).....	18
Figura 2.2. Zonas de inclusión de árboles en bosques.....	19
Figura 2.3. Zonas de inclusión de árboles en selvas.....	20
Figura 2.4. Distribución espacial de los conglomerados utilizados en el estudio. Selva mediana subperennifolia y subcaducifolia en Quintana y Campeche, México.....	36
Figura 2.5. Diagramas de dispersión y líneas de regresión de las variables de interés sobre el área del conglomerado (ha) en selva mediana subperennifolia en Quintana Roo.....	38
Figura 2.6. Diagramas de dispersión y líneas de regresión de las variables de interés sobre el área del conglomerado (ha) en selva mediana subcaducifolia en Quintana Roo.....	39
Figura 2.7. Diagramas de dispersión y líneas de regresión de las variables de interés sobre el área del conglomerado (ha) en selva mediana subperennifolia en Campeche.....	40
Figura 2.8. Diagramas de dispersión y líneas de regresión de las variables de interés sobre el área del conglomerado (ha) en selva mediana subcaducifolia en Campeche.....	41
Figura 2.9. Distribución de las muestras bootstrap de selva mediana subperennifolia de Quintana Roo. Método media de razones.....	49
Figura 2.10. Distribución de las muestras bootstrap de selva mediana subperennifolia de Campeche. Método media de razones.....	50
Figura 3.1. Distribución aleatoria de árboles (puntos) en una población hipotética en Quintana, Roo, México, en una superficie de 10,000 hectáreas.....	78
Figura 3.2. Distribución aleatoria de árboles (puntos) en una población hipotética en Quintana, Roo, México y conglomerado (círculo grande) con subparcelas circulares (círculos chicos).....	79
Figura 3.3. Distribución aleatoria de árboles (puntos) en una población hipotética en Quintana, Roo, México y conglomerado (círculo grande) con subparcelas rectangulares.....	79
Figura 3.4. Muestreo de árboles en una población hipotética en Quintana, Roo, México, cuando el conglomerado es de subparcelas circulares.....	80
Figura 3.5. Muestreo de árboles en una población hipotética en Quintana, Roo, México, cuando el conglomerado es de subparcelas rectangulares.....	80
Figura 3.6. Sesgos relativos de la estimación. Variable densidad en condición c0.....	86
Figura 3.7. Sesgos relativos de la varianza. Variable densidad en condición c0.....	87
Figura 3.8. Sesgos relativos de la estimación. Variable densidad en condición c1.....	88
Figura 3.9. Sesgos relativos de la varianza. Variable densidad en condición c1.....	88
Figura 3.10. Sesgos relativos de la estimación. Variable densidad en condición c2.....	89
Figura 3.11. Sesgos relativos de la varianza. Variable densidad en condición c2.....	90
Figura 3.12. Sesgos relativos de la estimación. Variable volumen en condición c0.....	91
Figura 3.13. Sesgos relativos de la varianza. Variable volumen en condición c0.....	91
Figura 3.14. Sesgos relativos de la estimación. Variable volumen en condición c1.....	92
Figura 3.15. Sesgos relativos de la varianza. Variable volumen en condición c1.....	93
Figura 3.16. Sesgos relativos de la estimación. Variable volumen en condición c2.....	94
Figura 3.17. Sesgos relativos de la varianza. Variable volumen en condición c2.....	94
Figura 4.1. Muestreo con reemplazo parcial en dos ocasiones (t_1, t_2).....	119
Figura 4.2. Muestreo con reemplazo parcial en dos ocasiones (t_1, t_4).....	136
Figura 4.3. Una alternativa del muestreo con reemplazo parcial en tres ocasiones.....	137
Figura 4.4. Muestreo multitemporal panelizado con ciclo de remediación de tres años.....	139
Figura 4.5. Muestreo multitemporal panelizado con ciclo de remediación de cinco años.....	141
Figura 4.6. Representación del producto de dos funciones de densidad de probabilidad normal.....	153
Figura 4.7. Muestras panelizadas que constituyeron la pseudo-población forestal artificial.....	163

Figura 4.8. Matriz diseño utilizada en el proceso de estimación de Mínimos Cuadrados Generalizados...	164
Figura 4.9. Estimaciones de ABHA obtenidas mediante MCG, MRE y ECU en la condición uno.....	170
Figura 4.10. Estimaciones de ABHA obtenidas mediante MCG, MRE y ECU en la condición dos.....	170
Figura 4.11a. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2004 a 2005 en la condición uno.....	172
Figura 4.11b. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2006 a 2007 en la condición uno.....	173
Figura 4.11c. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2009 a 2010 en la condición uno.....	174
Figura 4.11d. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2011 a 2012 en la condición uno.....	175
Figura 4.12a. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2004 2005 en la condición dos.....	176
Figura 4.12b. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2006 2007 en la condición dos.....	177
Figura 4.12c. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2009 2010 en la condición dos.....	178
Figura 4.12d. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2011 2012 en la condición dos.....	179
Figura 4.13. Intervalos de confianza al 95% para ABHA generados por ECU en la condición uno.....	181
Figura 4.14. Intervalos de confianza al 95% para ABHA generados por ECU en la condición dos.....	181
Figura 4.15. Estimaciones de ABHA obtenidas a través de MCG, MRE y MM en la condición uno.....	184
Figura 4.16. Estimaciones de ABHA obtenidas a través de MCG, MRE y MM en la condición dos.....	184
Figura 4.17. Estimación del cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición uno.....	186
Figura 4.18. Estimación del cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición dos.....	186
Figura 4.19. Funciones de densidad normal modeladas del cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición uno.....	187
Figura 4.20. Funciones de densidad normal modeladas del cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición dos.....	188
Figura 4.21. Intervalos de confianza al 95% para cambio de ABHA generados por ECU en la condición uno.....	188
Figura 4.22. Intervalos de confianza al 95% para cambio de ABHA generados por ECU en la condición dos.....	189
Figura 4.23. Funciones de densidad normal modeladas de la tasa cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición uno.....	191
Figura 4.24. Funciones de densidad normal modeladas de la tasa cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición dos.....	191
Figura 4.25. Intervalos de confianza al 95% para la tasa de cambio de ABHA generados por ECU en la condición uno.....	192
Figura 4.26. Intervalos de confianza al 95% para la tasa de cambio de ABHA generados por ECU en la condición dos.....	192

DESARROLLO DE TÉCNICAS ESTADÍSTICAS PARA LA EVALUACIÓN Y MONITOREO DE ECOSISTEMAS FORESTALES EN MÉXICO

Efraín Velasco Bautista, Dr.
Colegio de Postgraduados, 2015.

RESUMEN

La Comisión Nacional Forestal implementó en 2004 el proyecto Inventario Nacional Forestal y de Suelos (INFyS), cuyo primer ciclo de medición se llevó a cabo de 2004 a 2007, y el segundo de 2009 a 2013. A pesar del esfuerzo realizado para la toma de datos en campo en el contexto del INFyS, actualmente no existe un marco conceptual estadístico que guíe la estimación de los parámetros de interés forestal considerando los datos de la remediación. Por lo anterior, el objetivo del presente proyecto de investigación fue generar herramientas estadísticas para la estimación de parámetros forestales en el contexto del INFyS que tomen en cuenta los datos provenientes de la remediación. El reporte contiene cinco capítulos. El Capítulo I presenta los aspectos generales de la investigación incluyendo las principales conclusiones. Enfatiza que aun cuando el conglomerado de muestreo definido en el INFyS incluye teóricamente cuatro subparcelas, por varias circunstancias existe la posibilidad que en campo la información se tome en conglomerados incompletos, lo cual conduce a conglomerados de diferentes tamaños. En el Capítulo II se presenta un análisis comparativo de los estimadores media de razones (MR), razón de medias (RM) y Forest Inventory and Analysis (FIA) para estimar parámetros de interés forestal en la selva mediana de Campeche y Quintana Roo con datos del segundo ciclo de medición del INFyS. En las diferentes condiciones estudiadas, los resultados indican que para área basal, densidad y volumen por hectárea los tres estimadores son equivalentes. En el Capítulo III se analizan las propiedades estadísticas de MR, RM, FIA y el estimador Van Deusen (VAN) utilizando una población de árboles simulados. Para el muestreo se consideraron tres condiciones de completitud de los conglomerados. Para cada una de estas condiciones, las estimaciones de volumen y número de árboles por hectárea obtenidas mediante MR, RM, FIA y VAN resultaron similares y prácticamente insesgadas. No obstante, a medida que el grado de completitud fue menor, tanto para volumen como para densidad, la varianza estimada de FIA se volvió fuertemente sesgada. En el Capítulo IV se analizan las propiedades estadísticas de estimadores para parámetros forestales cuando los datos de campo se obtienen mediante muestreo de panel rotante. Se hace referencia al Estimador de Muestreo Replicado (MRE), de Mínimos Cuadrados Generalizados (MCG) y al Estimador Compuesto Univariado (ECU). Para este estudio se consideró una pseudo-población forestal artificial resultante de la unión de siete bases de datos del INFyS, donde las remediciones fueron más consistentes. El muestreo se realizó en dos condiciones de tamaño de muestra panelizada (1000 y 1500 conglomerados). En ambas condiciones, los resultados de área basal por hectárea indican que MCG es ligeramente sesgado, pero con varianza pequeña y que MRE es insesgado, pero de varianza grande. El ECU produce estimaciones cuasi insesgadas y de mínima varianza.

Palabras clave: Inventario Nacional Forestal y de Suelos, muestras panelizadas, muestreo de panel rotante, Mínimos Cuadrados Generalizados, Estimador Combinado Univariado

DEVELOPMENT OF STATISTICAL TECHNIQUES FOR THE EVALUATION AND MONITORING OF FOREST ECOSYSTEMS IN MEXICO

Efraín Velasco Bautista, Dr.
Colegio de Postgraduados, 2015.

ABSTRACT

The National Forestry Commission implemented in 2004 the National Forest and Soil Inventory (INFyS) project, whose first measurement cycle was carried out from 2004 to 2007, and the second of 2009 to 2013. Despite the efforts made for field data collection in the context of INFyS, there is not a conceptual statistical framework to guide the estimation of the parameters of forest interest based on re-measured data. Therefore, the main objective of this research was to generate statistical tools for estimating forest parameters in the context of the INFyS. The report contains five chapters. Chapter I presents the general aspects of the research including the principal conclusions. It is emphasized that although the INFyS sampling cluster theoretically includes four subplots, for several reasons there is the possibility field data are taken into smaller clusters, leading to a sampling that practically implies different cluster size. It also emphasizes the lack a conceptual framework that considers statistical analysis from re-measurement data. In Chapter II the results of the statistical comparison of the estimators of Mean of Ratios (MR), Ratio of Means (RM) and Forest Inventory and Analysis (FIA) are presented for data of the medium rainforest of Campeche and Quintana Roo corresponding to the second measurement cycle. The results indicate that for basal area, density and volume per hectare all three estimators are equivalent. Chapter III the statistical properties the MR, RM, FIA and Van Deusen (VAN) estimators are analyzed using a simulated population of trees. Three conditions of completeness of the clusters were considered. In each of these conditions, estimates of volume and number of trees per hectare obtained by MR, RM, FIA and VAN were similar and practically unbiased. However, as the level of completeness decreases (the number of incomplete clusters increases) the estimated variance of the FIA estimator became strongly biased for both volume and density. Chapter IV the statistical properties of estimators for forest parameters are analyzed when field data are obtained through a sampling of rotating panel. Reference is made to the Replicated Sampling Estimator (RES), Generalized Least Squares (GLS) and the Univariate Composite Estimator (UCE). An artificial forest pseudo-population resulting from joining seven databases of the INFyS where re-measurements were more consistent was considered for this study. Sampling was carried out in two conditions of panelized sample size (1000 and 1500 clusters). In both conditions, the results of basal area per hectare indicate that GLS is slightly biased, but with smaller variance and that RES is unbiased, but with larger variance. The UCE produces estimates quasi-unbiased and of minimum variance.

Key words: National Forest and Soil Inventory, panelized samples, rotating panel sampling, Generalized Least Squares and the Univariate Composite Estimator.

CAPÍTULO I

ASPECTOS GENERALES SOBRE LA INVESTIGACIÓN

1.1. INTRODUCCIÓN

La cuantificación del carbono, la biomasa o el volumen maderable contenido en los ecosistemas forestales requiere de la implementación de un sistema de inventario y monitoreo, el cual a su vez debe contemplar un marco conceptual estadístico que de sustento técnico y científico a los resultados que se generan.

Desde la década de los sesentas se han realizado en México tres inventarios forestales a nivel nacional. El primero se inició en 1961 y concluyó en 1985; el segundo, llamado Inventario Forestal de Gran Visión, se realizó en 1991; el tercero, denominado Inventario Nacional Forestal Periódico, se efectuó en 1994 (SARH, 1992; SARH, 1994). Los resultados de estos inventarios no son comparables entre sí por utilizar metodologías y criterios diferentes.

La evaluación de los recursos forestales del país está contemplada en la Ley General de Desarrollo Forestal Sustentable (LGDFS), la cual establece que el Inventario Nacional Forestal y de Suelos (INFyS) debe comprender la superficie y localización de terrenos forestales y preferentemente forestales; los tipos de vegetación forestal y de suelos; la dinámica de cambio de la vegetación forestal; la cuantificación de los recursos forestales, que incluya la valoración de los bienes y servicios ambientales que generan los ecosistemas forestales; y los criterios e indicadores de sustentabilidad y degradación de ese tipo de ecosistemas forestales. Los datos comprendidos en el INFyS serán la base para el cálculo del volumen de madera y/o biomasa forestal en pie (SEMARNAT, 2015).

Con base en lo anterior se planteó establecer una línea base de la cual se pueda partir para obtener los indicadores que la LGDFS y su Reglamento establecen, así como para favorecer la planeación a largo plazo en el sector forestal y dar soporte a una serie de estudios que en el marco del cambio climático México ha comprometido con el

mundo. Fue así que en el 2004 la Comisión Nacional Forestal implementó el proyecto denominado Inventario Nacional Forestal y de Suelos.

La operación del INFyS, asociada a las políticas emanadas de la estrategia de mitigación del cambio climático, ha atraído la inversión de recursos internacionales al propiciar la implementación de proyectos como “Fortalecimiento del proceso de preparación para REDD+ en México y el Fomento a la Cooperación Sur-Sur” y “Bosque y Cambio Climático” (CONAFOR, 2012a).

El programa REDD+ —cuyo propósito es reducir las emisiones de CO₂ derivadas de los procesos de deforestación y degradación de los ecosistemas forestales del planeta, además de considerar mecanismos para la conservación, el manejo forestal sustentable y el aumento de las reservas forestales de carbono— demanda de la información generada en el INFyS, ya que para su implementación efectiva en México se requiere establecer los puntos de referencia o línea base de las reservas de carbono existentes al inicio del programa. Los resultados del INFyS también constituirán una plataforma valiosa para el sistema de Medición, Reporte y la Verificación (MRV) de las nuevas reservas de carbono de REDD+ (CONAFOR, 2010; pág. 37).

No obstante, y a pesar del esfuerzo realizado en el presente siglo para la toma de datos en campo en el contexto del INFyS, en la actualidad no existe un marco conceptual estadístico que señale cómo realizar estimaciones de los principales parámetros de interés forestal que considere los datos de las remediciones sucesivas del mismo. Por tanto, en el presente proyecto de investigación se planteó como objetivo principal coadyuvar a la generación de herramientas estadísticas alternativas para la estimación de parámetros forestales agregados en el contexto del proyecto del Inventario Nacional Forestal y de Suelos de México.

1.2. PLANTEAMIENTO DEL PROBLEMA

A pesar del trabajo intenso realizado en el presente siglo para la toma de datos en campo (miles de conglomerados con información forestal relativamente reciente) correspondientes al proyecto del Inventario Nacional Forestal y de Suelos de México, coordinado por la Comisión Nacional Forestal, actualmente no existe un marco conceptual estadístico que indique cómo realizar estimaciones de los principales parámetros agregados de interés forestal que considere la estructura del sistema de muestreo (muestras panelizadas), y en consecuencia, los datos de la remediación.

Por otra parte, si bien es cierto que el diseño del conglomerado incluye teóricamente cuatro subparcelas, por varias circunstancias existe la posibilidad de que en campo la información se obtenga en conglomerados menores a cuatro subparcelas, lo cual conduce a conglomerados de tamaños distintos. Por lo que resulta importante estudiar el comportamiento de los diferentes estimadores que se pueden ser de utilidad para la estimación de parámetros forestales agregados cuando se tienen conglomerados —de subparcelas espacialmente disjuntas— de diferente tamaño.

El esfuerzo para el estudio de estimadores en el contexto del INFyS ha sido incipiente. Velasco *et al.* (2005) propusieron estimadores de razón para algunos parámetros forestales tomando en cuenta diferentes niveles de estratificación, pero considerando únicamente una evaluación. Al considerar unidades de muestreo de tamaños diferentes, y al suponer correlación lineal entre la variable de interés y el área muestreada de la parcela, estos autores propusieron estimadores de razón. Otros estimadores estadísticos como media de razones, Forest Inventory and Analysis y Van Deusen, también pueden ser aplicados en tal situación.

1.3. OBJETIVOS

Considerando que los resultados derivados del monitoreo de los ecosistemas forestales de México son la base para la formulación de la política nacional en materia forestal y de cambio climático, que se requieren para la integración del Programa de Evaluación de los Recursos Forestales del mundo — coordinado por la FAO— y para el Sistema de Medición, Reporte y Verificación (MRV) de REDD+, y que además se han determinado como Información de Interés Nacional (DOF, 28/mayo/2014), en el presente proyecto de investigación se plantean los siguientes objetivos:

General

Coadyuvar a la generación de herramientas estadísticas para la estimación de parámetros forestales agregados en el contexto del proyecto del Inventario Nacional Forestal y de Suelos de México.

Específicos

- 1) Evaluar el comportamiento de estimadores alternativos al estimador de razón cuando se tienen unidades de muestreo de diferentes tamaños utilizando bases de datos reales del INFyS.
- 2) Evaluar el comportamiento de estimadores alternativos al estimador de razón cuando se tienen unidades de muestreo de diferentes tamaños utilizando bases de datos de una población forestal artificial.
- 3) Estudiar desde el punto de vista teórico técnicas estadísticas para la estimación de los principales parámetros de interés forestal que consideren la estructura actual del sistema de muestreo panelizado. Evaluar el desempeño estadístico de estas técnicas en una pseudo-población forestal artificial.

1.4. REVISIÓN DE LITERATURA

Desde la década de los sesentas, se han realizado en México tres inventarios forestales de alcance nacional. El primero se inició en 1961 y concluyó en 1985; el segundo, llamado Inventario Forestal de Gran Visión, se realizó en 1991; el tercero, denominado Inventario Nacional Forestal Periódico, se efectuó en 1994. En el año 2000 hubo un intento por realizar otro inventario nacional forestal, sin embargo, únicamente se actualizó la cartografía forestal por parte de la Universidad Nacional Autónoma de México. Aun cuando el principal objetivo ha sido la estimación de existencias maderables y de incrementos, el diseño de la unidad de muestreo ha variado; se han usado tanto unidades de muestreo individuales, como en conglomerados, distribuidas básicamente en forma sistemática (SARH, 1992; SARH, 1994). Estos inventarios no permiten hacer comparaciones entre sí de sus resultados por utilizar metodologías diferentes y por manejar terminología y criterios distintos. En estas condiciones tampoco es posible evaluar fácilmente la tendencia de los recursos forestales a través del tiempo.

Atendiendo lo señalado en la Ley General de Desarrollo Forestal Sustentable (LGDFS) en cuanto a la evaluación de los recursos forestales del país, y con el fin de cumplir con los compromisos que México ha establecido con otras naciones del mundo respecto a las acciones directas para mitigar el cambio climático, en 2004 la Comisión Nacional Forestal implementó el proyecto Inventario Nacional Forestal y de Suelos (INFyS). Este proyecto contempla la elaboración de un programa de inventarios nacionales forestales y de suelos con una periodicidad de 5 años y metodologías homogéneas que permitan su aplicación a escalas nacional, regional y estatal; además la comparación en el tiempo para la determinación de cambios y tendencias en los recursos forestales nacionales. El primer ciclo del INFyS inició en 2004 con la toma de datos en campo y concluyó en 2009 con la elaboración del informe de resultados que fue presentado en 2011 —en total se levantaron en campo 20,791 conglomerados—. El segundo ciclo de inventario inició con el trabajo de campo en 2009, y concluyó en 2013 (CONAFOR, 2012a; CONAFOR, 2015).

Las unidades de muestreo del INFyS se ubican en todos los ecosistemas forestales del país, incluyendo bosques, selvas y vegetación de zonas áridas y semiáridas. Otros niveles inferiores de estratificación como bosque de pino, bosque de encino, selva alta perennifolia, selva baja espinosa, matorral subtropical y matorral desértico, también son considerados. La base cartográfica utilizada corresponde a la carta de Uso del Suelo y Vegetación Serie II, III y IV, a escala 1:250 000, elaborada por el INEGI. El diseño de muestreo es estratificado sistemático por conglomerados, los cuales se localizan cada 5 x 5 km en bosques y selvas altas y medianas, cada 10 x 10 km en selvas bajas y comunidades semiáridas, y cada 20 x 20 km en comunidades áridas. Los bosques y selvas de galería también se evalúan en conglomerados cada 5 km. La red de puntos a cada 5 km está diseñada para que sea compatible con el diseño de muestreo de los inventarios forestales estatales (CONAFOR, 2012b).

El conglomerado o unidad de muestreo primaria (UMP), desde el punto de vista conceptual es una parcela circular de una hectárea (56.42 m de radio), en la cual se evalúan cuatro unidades de muestreo secundarias (UMS) o sitios, dispuestos geométricamente en forma de una “Y” invertida con respecto al norte, los cuales son circulares en el caso de bosques y vegetación de zonas áridas y rectangulares en selvas (Figuras 1.1 y 1.2). La UMS número 1 constituye el centro de la UMP y las UMS 2, 3 y 4 son consideradas como periféricas (satélite). La distancia del centro de la UMS 1 al de cada una de las UMS periféricas es de 45.14 m cuando los sitios son circulares y de 36.42 m cuando son rectangulares; el azimut para localizar las UMS 2, 3 y 4 a partir del centro de la UMS 1 es de 0°, 120° y 240°, respectivamente. Las unidades de estudio (UE) son los árboles u otros elementos biológicos presentes en las UMS (Velasco *et al.*, 2005; CONAFOR, 2012b). El diseño del conglomerado, es decir, la configuración geométrica de las UMS, es muy similar al que usa el Forest Inventory and Analysis (FIA) Program del Servicio Forestal de Estados Unidos (Bechtold y Scott, 2005; McRoberts *et al.*, 2005). En cada UMS se determinan las especies arbóreas y se les mide el diámetro normal (DN) y la altura total de los árboles mayores a 7.5 cm de DN.

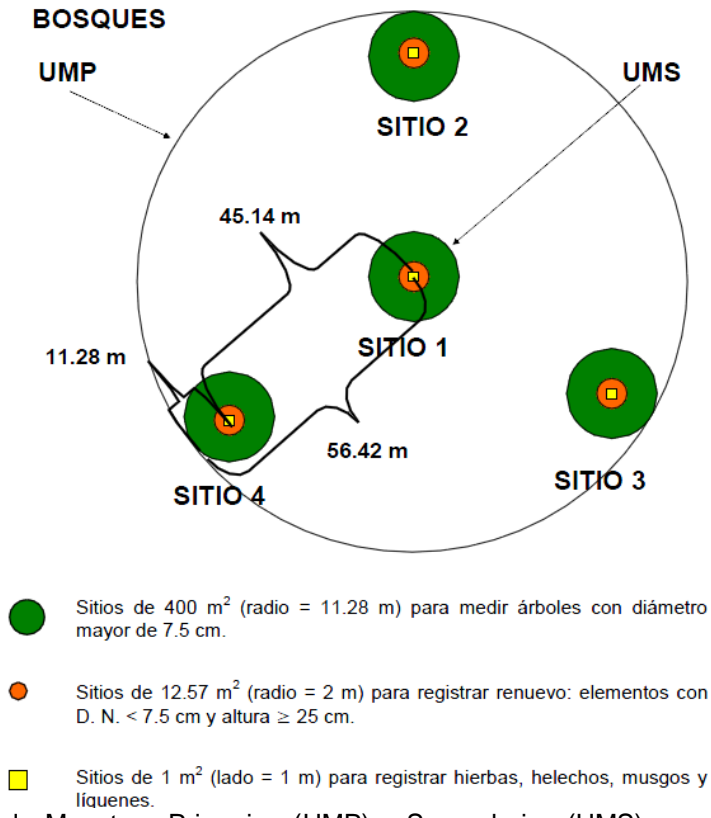


Figura 1.1. Unidades de Muestreo Primarias (UMP) y Secundarias (UMS) para el Inventario Nacional Forestal y de Suelos en bosques y vegetación de zonas áridas. Tomado de CONAFOR (2012b).

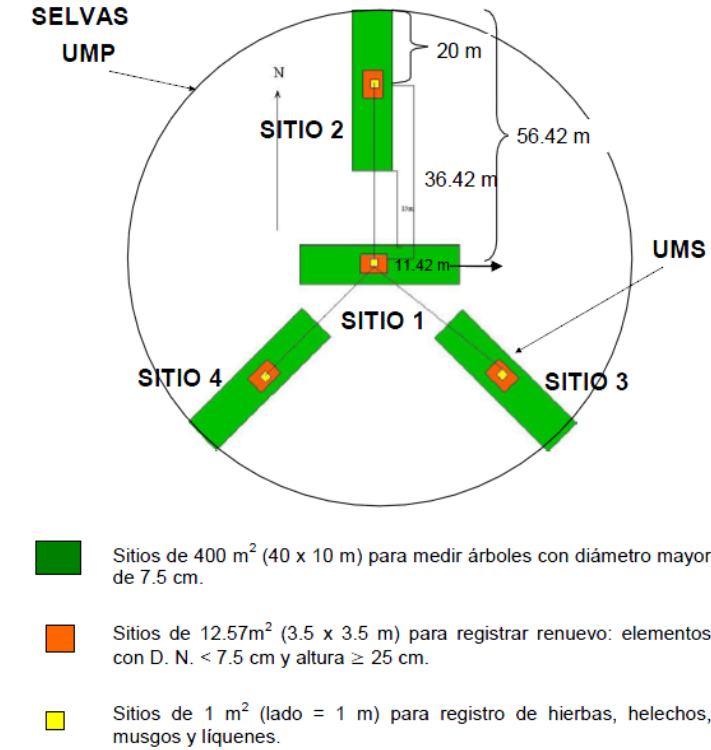


Figura 1.2. Unidades de Muestreo Primarias (UMP) y Secundarias (UMS) para el Inventario Nacional Forestal y de Suelos en selvas. Tomado de CONAFOR (2012b).

El registro de árboles con diámetro normal mayor a 7.5 cm se hace en las UMS circulares de 400 m² (radio 11.28 m) en el caso bosques y vegetación de zonas áridas, y rectangulares, también de 400 m² (de 10 X 40 m), en selvas. Para registrar el repoblado, dentro de cada UMS de 400 m² se traza un círculo concéntrico de 12.56 m² (2 m de radio) en el caso de bosques, y un cuadro de 3.54 x 3.54 m en el caso de selvas. Los datos de las plantas menores no leñosas, se toman en un cuadro de 1 m² localizado en el centro de la UMP (Figuras 1.1 y 1.2) (Velasco *et al.*, 2005; CONAFOR, 2012b).

Las principales variables que se registran en los elementos arbóreos —de bosques, selvas y vegetación de zonas áridas y semiáridas— son nombre científico de la especie, diámetro normal y de copa y altura total, así como presencia/ausencia de daños (bióticos y abióticos); en el caso de las coníferas, adicionalmente se registra la edad de los tres árboles más cercanos al centro de la UMS de 400 m². Sólo en el sitio 3 de cada conglomerado se toman datos para la evaluación de combustibles forestales mediante un muestreo de transectos (CONAFOR, 2012b).

Velasco *et al.* (2004), mediante una simulación de siete arreglos geométricos diferentes de UMS dentro de UMP (circulares de 1 ha) en los que se contabilizó el número de árboles y se determinó el área basal y el volumen; definieron que los mejores diseños, en términos del menor error relativo de muestreo, fueron en las UMS que forman una cruz y en el que las UMS constituyen una T horizontal. Cuando las UMS conforman una Y invertida se presentó el error relativo de muestreo más alto; sin embargo, señalan que en ningún caso éste fue superior al 7%.

De particular importancia resulta el diseño de muestreo y del conglomerado del INFyS, en primera instancia podría considerarse como un muestreo por conglomerados en dos etapas, en donde la unidad de muestreo primaria es una parcela de una hectárea y las unidades de muestreo secundarias son las subparcelas dispuestas geométricamente en forma de Y invertida, por lo que los estimadores tendrían que considerar la varianza entre unidades de muestreo primarias y entre unidades de muestreo secundarias. No obstante, aun cuando un conglomerado comprende dos o

más parcelas disjuntas es tratado como una sola parcela debido a que es atado a un solo punto muestral. En efecto, el conglomerado de subparcelas simplemente sirve para identificar aquellos elementos cuyas zonas de inclusión espacialmente disjuntas incluyen al punto muestral (Gregoire y Valentine, 2008). En estas condiciones los estimadores consideran únicamente la varianza entre conglomerados.

Para cierta población de interés, por ejemplo los bosques de México o del estado de Durango, la estructura espacial y temporal de las mediciones (evaluaciones diferidas en el primer ciclo y anualidades parciales en la segunda medición) conduce al concepto de muestras panelizadas o muestreo de panel rotante. Los datos de panel son aquellos que combinan series de tiempo y secciones transversales; las mismas unidades de las secciones transversales son observadas en el tiempo. Un grupo típico de datos de este tipo consiste de un gran número de unidades de sección transversal observado en pocos puntos en el tiempo (Greene, 2012). En teoría, las variables medidas en los subpaneles del mismo grupo presentan un alto grado de correlación. En sentido amplio, un diseño de muestras panelizadas es una extensión de un muestreo con reemplazamiento parcial en dos ocasiones, en el cual sólo un grupo de parcelas evaluado en la primera ocasión es medido en la segunda. Los inventarios anualizados se caracterizan por proveer estimaciones anuales actualizadas del recurso forestal.

Varios diseños teóricos de muestras panelizadas se pueden generar cuando se planea un sistema de monitoreo a largo plazo (Scott *et al.*, 1999), pero en todos los casos el empleo de estimadores de Mínimos Cuadrados Generalizados es recomendable, ya que al considerar la correlación entre las variables medidas en los subpaneles del mismo grupo se obtienen estimaciones más precisas comparadas con aquellos métodos que ignoran la estructura de la correlación temporal.

El diseño de muestras panelizadas empleado en el contexto del INFyS es similar al diseño usado en la evaluación de recursos forestales de USA coordinado por el Forest Inventory and Analysis (FIA), en donde del 10 al 20% de la muestra original es evaluada anualmente (Reams *et al.*, 2005; Czaplewski *et al.*, 2012).

1.5. MATERIALES Y MÉTODOS

En una primera fase se compararon los estimadores estadísticos MR, RM y FIA. Para este fin se utilizaron datos reales de selva mediana subperennifolia y subcaducifolia de Quintana Roo y Campeche, los cuales se tomaron en campo en 2010 en el proyecto del INFyS. En una segunda fase se compararon estos tres estimadores y el estimador VAN utilizando datos de una población simulada de 9,370,000 árboles. En este caso, los estimadores estudiados fueron derivados analíticamente vía modelos y Mínimos Cuadrados Generalizados de acuerdo a la metodología de Van Deusen (2004, 2006).

Adicionalmente se documentaron y probaron estimadores estadísticos para la estimación de parámetros forestales cuando los datos de campo se obtienen mediante un sistema panelizado. Se estudió el Muestreo con Reemplazamiento Múltiple vía Mínimos Cuadrados Generalizados (MCG) (Van Deusen, 1989; Scott *et al.*, 1999), el estimador de Muestreo Replicado (MRE) y el Estimador Compuesto Univariado (ECU) (Czaplewski, 2010). Las propiedades estadísticas de estos estimadores fueron evaluadas con datos de una pseudo-población forestal. Se consideraron los tamaños de muestra panelizada de 1000 y 1500 conglomerados panelizados.

1.6. CONCLUSIONES

De la comparación MR, RM y FIA con datos reales. El sesgo relativo del estimador de razón siempre fue menor al 10% (sesgo respecto al error estándar), por tanto se consideró despreciable. En las diferentes condiciones estudiadas, los resultados indican que los tres estimadores no son diferentes estadísticamente. El estimador de medias de razón, sin embargo, presenta la ventaja de proporcionar estimaciones por hectárea de manera directa.

De la comparación MR, RM, FIA y VAN con datos simulados. En las tres condiciones de completitud de los conglomerados (100, 88 y 63 % de completitud) y para las dos variables de interés (densidad y volumen), las estimaciones de MR, RM, FIA y VAN resultaron similares y prácticamente insesgadas. En conglomerados de completitud total

la estimación de la varianza fue insesgada independientemente del estimador. Sin embargo, a medida que la muestra tuvo mayor número de conglomerados incompletos, la varianza estimada de FIA se volvió fuertemente sesgada.

De la comparación MCG, MRE y ECU en una pseudo-población forestal artificial. En las dos condiciones estudiadas —tamaños de muestra panelizadas de 1000 y 1500 conglomerados—, el estimador de MCG resultó ligeramente sesgado (5% en algunos casos), sin embargo, su varianza siempre resultó menor en comparación con el estimador MRE. El ECU, por una parte amortiguó el sesgo de MCG, y por la otra, siempre presentó una varianza menor en comparación con los estimadores MCG y MRE. De esta manera, ECU puede considerarse como un estimador insesgado (ó cuasinsesgado) de mínima varianza. Esta característica, hace al estimador combinado atractivo en muchas aplicaciones.

1.7. LITERATURA CITADA

- Bechtold W. A. and C.T. Scott. 2005. The forest inventory and analysis plot design. In: The enhanced Forest Inventory and Analysis program-national sampling design and estimation procedures. W. A. Bechtold and P. L. Patterson (Eds.). General Technical Report SRS-80. Forest Service. United States Department of Agriculture. Southern Research Station, Asheville, North Carolina. USA. pp: 27-42.
- Comisión Nacional Forestal (CONAFOR). 2010. Visión de México sobre REDD+: hacia una estrategia nacional. SEMARNAT. México. 54 p.
- Comisión Nacional Forestal (CONAFOR). 2012a. Memoria documental: Inventario Nacional Forestal y de Suelos. SEMARNAT. México. 47 p.
- Comisión Nacional Forestal (CONAFOR). 2012b. Manual y procedimientos para el muestreo de campo: Re-muestreo 2012. SEMARNAT. México. 136 p.
- Comisión Nacional Forestal (CONAFOR). 2015. Lecciones aprendidas del Inventario Nacional Forestal y de Suelos de México. SEMARNAT. México. 20 p.
- Czaplewski, R. L. 2010. Complex sample survey estimation in static state-space. Gen. Tech. Rep. RMRS-GTR-239. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 124 p.

- Czaplewski R.L., M.T. Thompson and G.G. Moisen. 2012. An efficient estimator to monitor rapidly changing forest conditions. In: Moving from Status to Trends: Forest Inventory and Analysis (FIA) Symposium 2012. R. S. Morin and G.C. Liknes (Comp.). General Technical Report NRS-P-105. Forest Service. United States Department of Agriculture. Northern Research Station. Baltimore, MD. USA. Pp:416-420.
- Diario Oficial de la Federación (DOF). 2014. ACUERDO por el que se determina Información de Interés Nacional el Inventario Nacional Forestal y de Suelos. Tomo DCCXXVIII No. 21. Miércoles 28 de mayo de 2014. México, D.F. 128 p.
- Greene, W.H. 2012. Econometric Analysis. Seventh Edition. Pearson Education, Inc. USA. 1188 p.
- Gregoire T.G. and H.T. Valentine. 2008. Sampling Strategies for Natural Resources and the Environment. Chapman & Hall/CRC. New York, USA. 474 p.
- McRoberts, R.E., W.A. Bechtold, P.L. Patterson, C.T. Scott and G.A. Reams. 2005. The enhanced Forest Inventory and Analysis program of the USDA Forest Service: Historical perspective and announcement of statistical documentation. Journal of Forestry 103(6): 304-308.
- Reams, G. A., W.D. Smith, M.H. Hansen, W.A. Bechtold, F.A. Roesch and G.G. Moisen. 2005. The forest inventory and analysis sampling frame. In: The enhanced Forest Inventory and Analysis program-national sampling design and estimation procedures. W. A. Bechtold and P. L. Patterson (Eds.). General Technical Report SRS-80. Forest Service. United States Department of Agriculture. Southern Research Station, Asheville, North Carolina. USA. pp: 11-26.
- Scott, C.T., M. Köhl and H.J. Schnellbacher. 1999. A comparison of periodic and annual forest surveys. Forest Science. 45(3):433-451.
- Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). 2015. Ley General de Desarrollo Forestal Sustentable. Reforma Publicada DOF. 26-03-2015. México, D. F. México. 80 p.
- Secretaría de Agricultura y Recursos Hidráulicos (SARH). 1992. Inventario nacional forestal de gran visión. México 1991-1992. Reporte Principal. México. 53 p. 18
- Secretaría de Agricultura y Recursos Hidráulicos (SARH). 1994. Inventario nacional forestal periódico 1992-1994. México. 81 p.
- Van Deusen, P.C. 1989. Multiple-occasion partial replacement sampling for growth components. Forest Science. 35(2):388-400.
- Van Deusen, P.C. 2004. Forest inventory estimation with mapped plots. Canadian Journal of Forest Research. 34:493-497.

Van Deusen, P.C. 2006. Comparing mapped plot estimators. In Proceedings of the Sixth Annual Forest Inventory and Analysis Symposium, 21-24 September 2004. Denver, Co, U.S. For. Ser. Gen. Tech. Rep. WO-70.

Velasco B., E., F. Moreno S. y R. Rodríguez P. 2004. Comparación de siete diseños de unidades de muestreo secundarias en inventarios forestales. Revista Ciencia Forestal en México. 27(92):29-51.

Velasco B., E., H. Ramírez M., F. Moreno S. y A. de la Rosa V. 2005. Estimadores de razón para el inventario nacional forestal de México. Revista Ciencia Forestal en México. 28(94):23-43.

CAPÍTULO II

COMPARACIÓN DE ESTIMADORES PARA PARÁMETROS DE INTERÉS FORESTAL EN SELVA MEDIANA DE MÉXICO

2.1. RESUMEN

En México, la Comisión Nacional Forestal coordina el proyecto denominado “Inventario Nacional Forestal y de Suelos (INFiS)”. Considerado éste desde una perspectiva a largo plazo, la fase de campo correspondiente a la primera evaluación inició en el año 2004 y concluyó en el 2007. La actualización de los datos forestales de este proyecto está planeada en periodos de cada cinco años, de tal manera que la segunda evaluación se realizó en 2009-2013. Para el análisis estadístico de los datos obtenidos en 2004-2007 se utilizaron estimadores de razón. Dado que los supuestos teóricos de éstos no se cumplen a plenitud (fuerte correlación lineal) es necesario comparar los resultados del estimador de razón con estimadores alternativos. En el presente trabajo se compararon estadísticamente los estimadores razón de medias (RM), medias de razones (MR) e Inventario Forestal y Análisis (FIA). Se consideraron como variables el área basal ($\text{m}^2 \text{ha}^{-1}$), densidad arbórea (árboles ha^{-1}) y volumen maderable promedio ($\text{m}^3 \text{ha}^{-1}$), las cuales se obtuvieron con datos del 2010 de selva mediana en Quintana Roo y Campeche. Las varianzas se obtuvieron según las expresiones convencionales y mediante bootstrap no paramétrico. El sesgo relativo (sesgo respecto al error estándar) del estimador de razón siempre fue menor al 10%, por tanto se consideró despreciable. En las diferentes condiciones estudiadas, los resultados indican que los tres estimadores no son diferentes estadísticamente. El estimador medias de razón, sin embargo, presenta la ventaja de proporcionar estimaciones por hectárea de manera directa, además, su planteamiento teórico permite analizarlo desde una perspectiva del estimador Horvitz-Thompson. Los intervalos de confianza confirmaron que bootstrap es un método alternativo potente para la estimación de las varianzas.

Palabras clave: Inventario Nacional Forestal de México, razón de medias, media de razones, estimador Horvitz-Thompson, bootstrap, zonas de inclusión espacialmente disjuntas.

ESTIMATORS COMPARISON FOR FOREST PARAMETERS IN MEDIUM RAINFOREST OF MEXICO

2.2. SUMMARY

In Mexico, the National Forestry Commission coordinates the project called "National Forest and Soils Inventory". This long-term perspective project started its first field phase in the year 2004 and concluded it in 2007. The forest data update is planned in five years periods, in such way that the second assessment was conducted from 2009 to 2013. For the statistical analysis of the data obtained in 2004-2007 the estimators of ratio of means were used. However, since the theoretical assumptions of the ratio of means estimator are not fully met (strong linear correlation), the results of this estimator with alternative estimators were compared. In the present work the estimators being compared are: ratio of means (RM), mean of ratios (MR) and Forest Inventory and Analysis (FIA). The basal area (m^2 per hectare), tree density (trees per hectare) and average timber volume (m^3 per hectare), which were obtained with the 2010 data of medium rainforest in Quintana Roo and Campeche, were considered as variables. The variances were obtained according to the conventional expressions and using non-parametric bootstrap. The relative bias (bias with respect to the standard error) of the ratio estimator was always less than 10%, it was therefore considered negligible. In the different studied conditions, the results indicate that the three estimators are not statistically different. The estimator mean of ratios, however, has the advantage of providing estimates per hectare direct way, moreover, his theoretical basis allows you to analyze it from the perspective of the Horvitz-Thompson estimator. The confidence intervals confirmed that bootstrap is a powerful alternative method for the estimation of the variance in cases where incomplete clusters are present.

Key words: National Forest Inventory of Mexico, ratio of means, mean of ratios, Horvitz-Thompson estimator, bootstrap, spatially disjoint inclusion zones

2.3. INTRODUCCIÓN

El diseño de muestreo que se usa en el proyecto del Inventario Nacional Forestal y de Suelos es estratificado sistemático por conglomerados, mismos que se localizan en equidistancias de 5, 10 ó 20 km en función del tipo de ecosistema en que aplique el muestreo. El conglomerado es una parcela circular de una hectárea, en la que se evalúan cuatro unidades de muestreo secundarias dispuestas geométricamente en forma de una “Y” invertida con respecto al norte, los cuales son circulares en el caso de bosques y vegetación de zonas áridas y rectangulares en selvas.

Por varias circunstancias, en campo la información se toma en conglomerados menores a 4 subparcelas, generándose entonces conglomerados de diferentes tamaños. Por otra parte, aun cuando un conglomerado comprende dos o más parcelas espacialmente disjuntas es tratado como una sola parcela debido a que el conglomerado es atado a un solo punto muestral. Una vez que se localiza el sitio 1, las subparcelas satélite quedan perfectamente definidas. Entonces, en la aplicación práctica de los estimadores, la varianza entre conglomerados es de principal interés.

Uno de los estimadores que se pueden aplicar cuando se tienen conglomerados de tamaño diferente es el estimador de razón, el cual considera la variable de interés en el numerador y el área del conglomerado efectivamente muestreada en el denominador. Además de su sencillez, el estimador de razón presentado de esta manera permite obtener estimaciones directamente por hectárea. Otros estimadores como media de razones y Forest Inventory and Analysis también pueden ser prometedores cuando las unidades de muestreo son de tamaño desigual.

Aun cuando la relación entre la variable de interés (volumen, área basal, número de árboles) y la variable auxiliar (área) no es tan fuerte, el estimador de razón se usa en el proyecto de Inventario Nacional Forestal y de Suelos de México, por lo que resulta necesario investigar el comportamiento de estimadores alternativos.

Con datos reales de Quintana Roo y Campeche, obtenidos por el INFyS en 2010, se compararon estadísticamente los estimadores razón de medias, medias de razones y Forest Inventory and Analysis.

2.4. OBJETIVO

El objetivo del presente trabajo fue comparar —usando datos reales del INFyS— los estimadores de MR, RM y FIA al considerar como variables de interés el área basal ($\text{m}^2 \text{ha}^{-1}$), densidad arbórea (árboles ha^{-1}) y volumen fustal ($\text{m}^3 \text{ha}^{-1}$) en selva mediana subperennifolia y subcaducifolia de los estados de Quintana Roo y Campeche.

2.5. REVISIÓN DE LITERATURA

2.5.1. Diseño del conglomerado y zona de inclusión

El diseño de muestreo que se usa en el proyecto del INFyS es estratificado sistemático por conglomerados, los cuales se localizan en equidistancias de 5 x 5 km en bosques y selvas altas y medianas, de 10 x 10 km en selvas bajas y comunidades semiáridas, y de 20 x 20 km en comunidades áridas (CONAFOR, 2012). Pocas comunidades ecológicas tienen un marco de muestreo de unidades elementales, de manera que normalmente no es factible o práctico obtener una muestra de miembros de la comunidad por muestreo aleatorio simple.

El conglomerado o unidad de muestreo primaria (UMP) es una parcela circular de una hectárea (56.42 m de radio), en la que se evalúan cuatro unidades de muestreo secundarias (UMS) o sitios, dispuestos geoméricamente en forma de una “Y” invertida con respecto al norte, los cuales son circulares en el caso de bosques y vegetación de zonas áridas y rectangulares en selvas (Figura 2.1). (Velasco *et al.*, 2005; CONAFOR, 2012). El diseño del conglomerado espacialmente es similar al del programa de Inventario Forestal y Análisis para la evaluación de recursos forestales en Estados Unidos de Norteamérica (Bechtold y Scott, 2005).

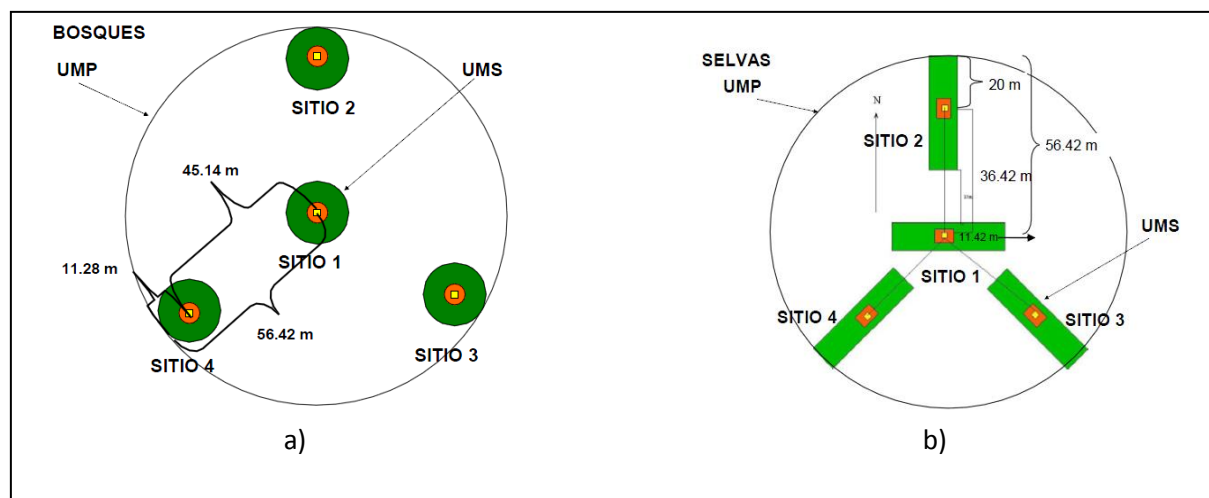


Figura 2.1. Diseño del conglomerado que se utiliza en el Inventario Nacional Forestal y de Suelos: bosques a) y selvas b). Tomado de CONAFOR (2012).

El diseño mostrado en la Figura 2.1 puede conceptualizarse como un conglomerado de subparcelas (circulares o rectangulares) que comprende una subparcela central georeferida en un punto muestral (x_s, z_s) y $n-1$ subparcelas satélite no traslapadas en un arreglo fijo. Este tipo de arreglo de subparcelas define una zona de inclusión disjunta espacialmente alrededor del punto central de cualquier objeto de interés, árboles por ejemplo (Valentine *et al.*, 2006).

Se pueden usar “vectores de dirección” para definir los puntos centrales de las subparcelas satélite relativos al punto muestral (Figura 2.2a, b), en cuyo caso los vectores inversos —vectores de dirección rotados 180° — definen los puntos centrales de las “subzonas satélite” de la zona de inclusión de un árbol relativa al punto central de ese árbol (Figura 2.2c, 2.2d). El árbol está centrado en la “subzona central” de su zona de inclusión. Entonces, una rotación de 180° de un conglomerado de subparcelas alrededor del punto muestral proporciona la forma de la zona de inclusión. Un conglomerado de subparcelas incluye un árbol de interés si el punto muestral cae en cualquier parte de la zona de inclusión de ese árbol (Figura 2.2e, 2.2f) (Valentine *et al.*, 2006; Gregoire y Valentine, 2008). En la Figura 2.2 se presenta la zona de inclusión para árboles en un conglomerado de bosques y en la Figura 2.3 de selvas.

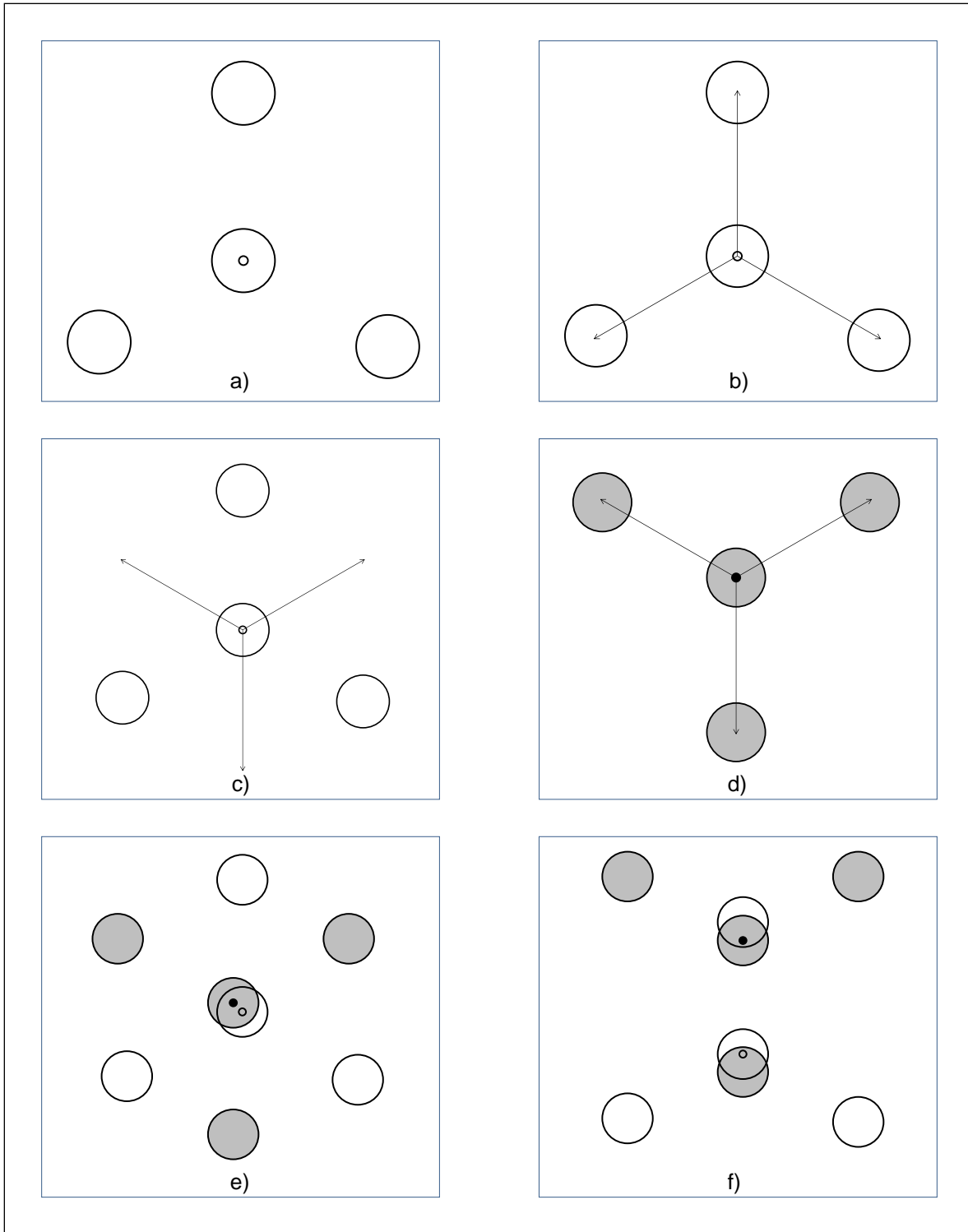


Figura 2.2. Zonas de inclusión de árboles en bosques. a) Un conglomerado de subparcelas circulares (bosques) comprende una subparcela central y subparcelas satélite alrededor de un punto fijo (círculo vacío). b) Vectores de dirección indican la distancia y dirección desde el punto muestral hasta los centros de las parcelas satélite. c) Vectores inversos —vectores de dirección rotados 180° —. d) Punto central de un árbol (círculo relleno) y puntos centrales de las subzonas satélite de la zona de inclusión de ese árbol. Un árbol ocupa un conglomerado de subparcelas si el punto muestral cae en cualquier lugar de la zona de inclusión de ese árbol (e y f).

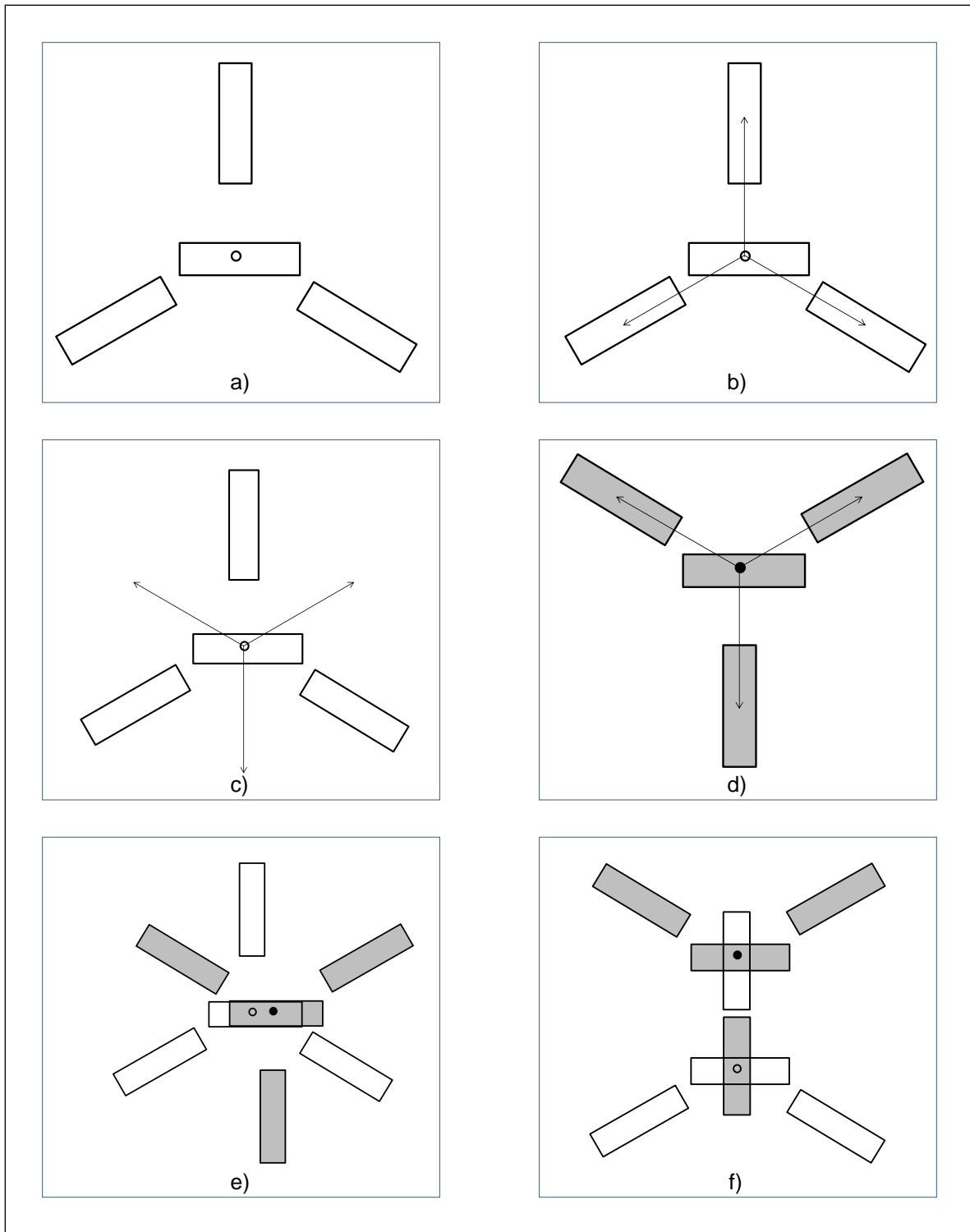


Figura 2.3. Zonas de inclusión de árboles en selvas. a) Un conglomerado de subparcelas rectangulares (selvas) comprende una subparcela central y subparcelas satélite alrededor de un punto fijo (círculo vacío). b) Vectores de dirección indican la distancia y dirección desde el punto muestral hasta los centros de las parcelas satélite. c) Vectores inversos —vectores de dirección rotados 180° —. d) Punto central de un árbol (círculo relleno) y puntos centrales de las subzonas satélite de la zona de inclusión de ese árbol. Un árbol ocupa un conglomerado de subparcelas si el punto muestral cae en cualquier lugar de la zona de inclusión de ese árbol (e y f).

El concepto de zona de inclusión se simplifica cuando se tiene únicamente una parcela de área fija; para una parcela circular, es sencillo verificar que la zona de inclusión es un círculo concéntrico alrededor del árbol del mismo radio que la parcela circular. En este caso, un árbol es seleccionado en la muestra si su zona de inclusión atrapa el punto de muestreo (Schreuder *et al.*, 1993).

La ubicación de los conglomerados en campo siempre está sujeta a las condiciones del terreno y la vegetación presente. En un conglomerado es posible que alguna o varias de las subparcelas sean imposibles de instalar, ya que caerían en barrancas, lagos, terrenos agrícolas, caminos o minas, por citar algunos ejemplos (subparcelas inexistentes). En otras circunstancias, por escala en la cartografía base (Uso de Suelo y Vegetación SERIE IV de INEGI, 1:250 000) existe la posibilidad que una parte de conglomerado caiga en un estrato diferente del definido en la cartografía, por lo que se puede tener la información de las 4 subparcelas, pero en cada estrato se tiene un conglomerado incompleto. En otros casos, por razones de linderos o de barreras topográficas únicamente se puede instalar una parte del conglomerado y entonces serán subparcelas inaccesibles.

La situación anterior conduce a que en campo la información se registre en conglomerados menores a 4 subparcelas, generándose así conglomerados de diferentes tamaños. Por otra parte, aun cuando un conglomerado comprende dos o más parcelas disjuntas es tratado como una sola parcela debido a que el conglomerado es atado a un solo punto muestral. En efecto, el conglomerado de parcelas simplemente sirve para identificar aquellos elementos cuyas zonas de inclusión espacialmente disjuntas incluyen al punto muestral (Gregoire y Valentine, 2008).

Existen por lo menos tres formas diferentes para estimar parámetros agregados de interés forestal cuando se tienen conglomerados de diferente tamaño. Con el fin de presentar los posibles estimadores y sus respectivas varianzas enseguida se detalla la notación.

2.5.2. Notación

Para la presentación de los estimadores es necesario considerar la siguiente notación.

A : área forestal de la región de interés (ha).

U_k : k -ésimo elemento en la población, $k = 1, 2, \dots, N$.

m : número de puntos muestrales.

P_s : s -ésimo conglomerado, $s = 1, 2, \dots, m$.

(x_s, z_s) : coordenadas del conglomerado P_s .

(x_k, z_k) : ubicación geográfica de un elemento de la población, U_k .

a_{sj} : área de cada subparcela (400 m^2) en P_s .

n_s : número de subparcelas evaluadas efectivamente en campo en la parcela s , $n_s = 1, 2, 3, 4$.

a_k : área de la subzona de inclusión del elemento U_k .

$n_s a_k$: área de la zona de inclusión del elemento U_k .

$\pi_k = \frac{n_s a_k}{A}$: probabilidad de inclusión del elemento U_k .

y_k : valor de una variable de interés asociado a U_k (área basal, volumen).

$T_y = \sum_{k=1}^N y_k$: total poblacional de una variable de interés.

N : número de árboles en el área forestal de la región de interés.

t : número de árboles en la subparcela.

2.5.3. Estimadores

Bajo el principio de que los conglomerados pueden ser de tamaño diferente, Velasco *et al.* (2005) propusieron estimadores de razón de medias para analizar datos del Inventario Nacional Forestal de México, en el denominador consideraron la variable de interés (volumen) y en denominador el área del conglomerado como variable auxiliar. Así, el estimador se expresa como

$$y_{ha} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = \frac{\sum_{s=1}^m \sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{\sum_{s=1}^m n_s a_{sj}} \quad (2.1)$$

Donde y_s es la cantidad total de la variable en la parcela muestral s y a_s es el área (ha) de la parcela muestral s efectivamente evaluada en campo.

La varianza estimada de (2.1), según Cochran (1993), es

$$v(y_{ha}) = \frac{1}{a^2} \frac{1}{m(m-1)} \sum_{s=1}^m (y_s - y_{ha} a_s)^2 \quad (2.2)$$

Donde, $\bar{a} = \frac{\sum_{s=1}^m a_s}{m}$

El estimador (2.1) funciona bien si i) la población completa a ser inventariada es teselada por todas las posibles parcelas muestrales (conglomerados no traslapados y sin espacios entre ellos) y ii) las parcelas son seleccionadas con probabilidades iguales. Ambos supuestos normalmente no se cumplen en la práctica de inventarios forestales, y en consecuencia, (2.1) puede ser inapropiado. En su lugar, las parcelas muestrales son seleccionadas como un punto muestral de un universo infinito de posibles puntos centrales de parcelas, y los árboles correspondientes a los rodales, donde parcelas de tamaños grandes son establecidas, son seleccionados con probabilidad más alta que aquellos de parcelas más pequeñas (Smelko y Saborowski, 1999).

En general, el estimador de razón es un estimador sesgado, no obstante existen ciertas circunstancias donde puede ser insesgado (ANEXO 2.1). El estimador de razón (2.1) es más efectivo cuando i) la relación entre y_s y a_s es una línea recta que pasa por el origen y ii) la varianza de y_s alrededor de esta recta es proporcional a a_s .

De manera práctica, el uso de (2.1) se justifica si i) el coeficiente de correlación entre y_s y a_s es mayor a 0.5 y si el tamaño de muestra es mayor a 30 (Scheaffer *et al.*, 2012; Cochran, 1993).

Scott *et al.* (2005), en el marco del programa de Inventario Forestal y Análisis (FIA) de Estados Unidos, propusieron un estimador similar a una media de razones, pero para la estimación en cada parcela muestral s se considera un área ajustada:

$$y_{ha}(s) = \frac{y_s}{a_s} = \frac{\sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{a_0 \bar{p}_0} \quad (2.3)$$

Donde

a_0 : área total usualmente usada para observar la variable de interés en una parcela (es decir, cuatro veces el área de la subparcela).

\bar{p}_0 : proporción media de áreas de las parcelas observadas que caen dentro de la población de interés:

$$\bar{p}_0 = \frac{\sum_{s=1}^m \sum_{j=1}^{n_s} a_{sj}}{a_0 m} \quad (2.4)$$

a_{sj} : área usualmente usada para observar la variable de interés en la subparcela j de la parcela s .

Los valores $y_{ha}(s)$ de (2.3) se usan para la estimación de la media por hectárea:

$$\bar{y}_{ha} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} \quad (2.5)$$

La varianza se estima de la siguiente manera:

$$v(\bar{y}_{ha}) = \frac{\sum_{s=1}^m (y_{ha}(s) - \bar{y}_{ha})^2}{m(m-1)} \quad (2.6)$$

En realidad, el estimador de Scott *et al.* (2005) es un estimador de media de razones que en cada parcela muestral s considera el área del conglomerado completo ajustada por la proporción media de áreas de las parcelas observadas que caen dentro de la población de interés.

El uso de la probabilidad de inclusión de un elemento U_k permite derivar otro estimador desde el punto de vista de muestreo replicado (Gregoire y Valentine, 2008).

La probabilidad de incluir un elemento U_k en cada localización muestral permite estimar

$T_y = \sum_{k=1}^N y_k$ mediante el estimador insesgado de Horvitz-Thompson (HT), es decir,

$$\hat{T}_{y\pi s} = \sum_{U_k \in P_s} \frac{y_k}{\pi_k} \quad (2.7a)$$

$$= \sum_{U_k \in P_s} \frac{y_k}{\frac{n_s a_k}{A}} \quad (2.7b)$$

$$= \sum_{U_k \in P_s} \frac{A y_k}{n_s a_k} \quad (2.7c)$$

$$= A \sum_{U_k \in P_s} \frac{y_k}{n_s a_k} \quad (2.7d)$$

$$= A \sum_{U_k \in P_s} \rho_k = A \rho_s \quad (2.7e)$$

Donde, $\rho_k = y_k / n_s a_k$ es el valor de y_k prorrateado por unidad de área y ρ_s es la suma de todos los valores prorrateados para el conglomerado P_s .

La instalación de múltiples conglomerados independientes P_s en A , seguido por la estimación HT de T_y con los datos de cada P_s , ha sido llamado Muestreo Replicado (Barabesi y Fattorini, 1998; Barabesi y Pisani, 2004). De acuerdo con Schreuder *et al.* (1993) y Gregoire y Valentine (2008), el estimador de T_y basado en un muestreo replicado de m P_s es

$$\hat{T}_{y\pi,rep} = \frac{1}{m} \sum_{s=1}^m \hat{T}_{y\pi s} \quad (2.8a)$$

$$= \frac{1}{m} \sum_{s=1}^m A \rho_s \quad (2.8b)$$

$$= \frac{A}{m} \sum_{s=1}^m \rho_s = A \bar{\rho} \quad (2.8c)$$

Donde $\bar{\rho}$ es el promedio de valores ρ_s entre los m conglomerados.

El estimador (2.8) puede verse como el promedio de m estimadores HT, cada uno basado en un punto muestral.

De acuerdo con Schreuder *et al.* (1993), Cochran (1993) y Gregoire y Valentine, (2008), la varianza de $\hat{T}_{y\pi s}$, al ser éste un estimador HT, está dada por

$$V(\hat{T}_{y\pi s}) = \sum_{k=1}^N y_k^2 \left(\frac{1 - \pi_k}{\pi_k} \right) + 2 \sum_{k=1}^N \sum_{k' > k}^N y_k y_{k'} \left(\frac{\pi_{kk'} - \pi_k \pi_{k'}}{\pi_k \pi_{k'}} \right) \quad (2.9)$$

Donde, $\pi_{kk'} = a_{kk'} / A$ y $a_{kk'}$ es el área de la zona inclusión conjunta para U_k y $U_{k'}$. La zona de inclusión conjunta para cualesquiera dos elementos es el lugar de puntos comunes para a las zonas de inclusión de U_k y $U_{k'}$.

Según Schreuder *et al.* (1993), Cochran (1993) y Thompson (2002), un estimador insesgado de (2.9) es

$$v(\hat{T}_{y\pi s}) = \sum_{k=1}^{t'} y_k^2 \left(\frac{1 - \pi_k}{\pi_k^2} \right) + 2 \sum_{k=1}^{t'} \sum_{k' > k}^{t'} y_k y_{k'} \left(\frac{\pi_{kk'} - \pi_k \pi_{k'}}{\pi_k \pi_{k'} \pi_{kk'}} \right) \quad (2.9')$$

Donde t' es el número de árboles seleccionados en el conglomerado P_s .

Si los m conglomerados han sido establecidos independientemente en A , la varianza de $\hat{T}_{y\pi,rep}$ es (Barabesi y Pisani, 2004)

$$V(\hat{T}_{y\pi,rep}) = \frac{1}{m} V(\hat{T}_{y\pi s}) \quad (2.10)$$

La varianza en (2.9') es algo tediosa para calcular ya que requiere la probabilidad conjunta de seleccionar U_k y $U_{k'}$. Por esta razón Särndal *et al.* (1992-pág. 424), Schreuder *et al.* (1993- pág. 116) y Gregoire y Valentine (2008-pág. 216) indican que $V(\hat{T}_{y\pi,rep})$ se estima insesgadamente de la siguiente manera. Sea

$$v(\hat{T}_{y\pi,rep}) = \frac{1}{m(m-1)} \sum_{s=1}^m \left(\hat{T}_{y\pi s} - \hat{T}_{y\pi,rep} \right)^2 \quad (2.11)$$

La cantidad total de la variable de interés dividida por el área de la población resulta en la cantidad promedio por unidad de área, es decir, $\lambda_y = T_y / A$. Esta cantidad puede ser estimada insesgadamente con datos de la muestra mediante

$$\hat{\lambda}_{y\pi s} = \frac{\hat{T}_{y\pi s}}{A} \quad (2.12)$$

En base a esta expresión, el estimador de λ_y a partir de muestreo replicado de m conglomerados es

$$\hat{\lambda}_{y\pi,rep} = \frac{1}{m} \sum_{s=1}^m \hat{\lambda}_{y\pi s} \quad (2.13a)$$

$$= \frac{1}{m} \sum_{s=1}^m \frac{\hat{T}_{y\pi s}}{A} \quad (2.13b)$$

$$= \frac{1}{Am} \sum_{s=1}^m \hat{T}_{y\pi s} \quad (2.13c)$$

Sustituyendo (2.8a) en (2.13c) se observa que el promedio por unidad de área puede estimarse dividiendo el total estimado entre el área de la región de estudio:

$$\hat{\lambda}_{y\pi,rep} = \frac{\hat{T}_{y\pi,rep}}{A} \quad (2.14)$$

Con base en (2.13c) y (2.7d), este estimador también puede expresarse en términos de ρ_s :

$$\hat{\lambda}_{y\pi,rep} = \frac{1}{Am} \sum_{s=1}^m \hat{T}_{y\pi s} \quad (2.15a)$$

$$= \frac{1}{Am} \sum_{s=1}^m A \sum_{U_k \in P_s} \frac{y_k}{n_s a_k} \quad (2.15b)$$

$$= \frac{1}{m} \sum_{s=1}^m \sum_{U_k \in P_s} \frac{y_k}{n_s a_k} \quad (2.15c)$$

$$= \frac{1}{m} \sum_{s=1}^m \sum_{U_k \in P_s} \rho_k \quad (2.15d)$$

$$= \frac{1}{m} \sum_{s=1}^m \rho_s = \bar{\rho} \quad (2.15e)$$

El resultado (2.15d) hace claro que la estimación prorrateada por unidad de área, ρ_k , no requiere el conocimiento de A ni de la determinación explícita de las probabilidades de inclusión.

Con base en (2.14), la varianza de $\hat{\lambda}_{y\pi,rep}$ es

$$V(\hat{\lambda}_{y\pi,rep}) = \frac{1}{A^2} V(\hat{T}_{y\pi,rep}) \quad (2.16)$$

La cual, de acuerdo con Gregoire y Valentine (2008-pág. 220), puede ser estimada insesgadamente por

$$v(\hat{\lambda}_{y\pi,rep}) = \frac{1}{m(m-1)} \sum_{s=1}^m \left(\hat{\lambda}_{y\pi s} - \hat{\lambda}_{y\pi,rep} \right)^2 \quad (2.17)$$

Si bien es cierto que (2.15) no requiere de la determinación explícita de las probabilidades de inclusión de cada árbol, sí supone que independientemente del número de suparcelas satélite que finalmente integren el conglomerado en campo, las subzonas de inclusión estén completas (no truncadas). En este sentido, (2.15c) implica que para cada elemento, el valor de la variable de interés sea dividido entre el área del conglomerado evaluada efectivamente en campo ($n_s a_k = n_s a_{sj}$), la cual es común para todos los árboles que pertenecen al mismo conglomerado.

El supuesto de que ninguna de las zonas de inclusión esté cerca del límite del área forestal implica que ajustes de las probabilidades de inclusión de los árboles no son requeridos. Esto permite que una probabilidad de inclusión constante sea asignada a todos los árboles contados en las unidades de muestreo. Este supuesto es necesario debido a que es impráctico obtener probabilidades de selección para árboles límite (Schreuder *et al.*, 2003).

En otra perspectiva, y suponiendo que las subzonas de inclusión de los árboles estén completas, el estimador (2.15c) puede considerarse como una media de razones (Smelko y Merganic, 2008). El estimador “media de razones” es insesgado y fue recomendado para muestreo sistemático con parcelas muestrales de tamaño diferente (Smelko y Saborowski, 1999). Por lo tanto, en cada parcela muestral s los datos muestrales y_s necesitan ser calculados a un área igual (1 ha) usando la siguiente expresión (Smelko y Merganic, 2008):

$$y_{ha}(s) = \frac{y_s}{a_s} = \frac{\sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{n_s a_{sj}} \quad (2.18)$$

Donde y_s es la cantidad total de la variable de interés en la parcela muestral s y a_s es el área (ha) de la parcela muestral s efectivamente evaluada en campo.

Estos valores por hectárea $y_{ha}(s)$ son usados para la estimación de la media por hectárea:

$$\bar{y}_{ha} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} \quad (2.19)$$

No existen estimadores para la varianza de (2.19) cuando se utiliza muestreo sistemático, por lo tanto se estima de la siguiente manera:

$$v(\bar{y}_{ha}) = \frac{\sum_{s=1}^m (y_{ha}(s) - \bar{y}_{ha})^2}{m(m-1)} \quad (2.20)$$

la cual ha mostrado ser una varianza conservadora en muchas aplicaciones de muestreo sistemático (Smelko y Saborowski, 1999).

Es claro que el primer nivel de inferencia del estimador (2.15) ó (2.19) es precisamente una hectárea, lo cual permite estimar el total poblacional de manera directa con tan solo conocer el valor de A .

La estimación de la varianza para los estimadores (2.1), (2.5) y (2.19) considera un muestreo aleatorio aun cuando el diseño de muestreo del inventario nacional forestal es sistemático. Sin embargo, en inventarios forestales es razonable suponer que el muestreo sistemático es aproximadamente equivalente al muestreo aleatorio simple (Zarnoch y Bechtold, 2000). Velasco *et al.* (2005) refieren el uso del muestreo aleatorio

en inventarios forestales cuando los datos son colectados en una malla sistemática. Cochran (1993) indica que si la población es aleatoria la fórmula para la varianza en muestreo sistemático es la misma que para una muestra aleatoria simple. Adicionalmente, Cochran (1993, página 266) provee la siguiente justificación para el uso de estimaciones basadas en muestreo aleatorio simple cuando se tienen muestras sistemáticas:

“Considérense todas las $N!$ poblaciones finitas las cuales están formadas por las $N!$ permutaciones de cualquier conjunto de números y_1, y_2, \dots, y_N . Entonces, en el promedio sobre estas poblaciones finitas, $E(V_{sy}) = V_{ran}$. Note que V_{ran} es la misma para todas las permutaciones.”

Este resultado muestra que, si el orden de los objetos en una población finita específica puede considerarse como sacado al azar de entre las $N!$ permutaciones, el muestreo sistemático es, en promedio, equivalente al muestreo aleatorio simple.

Una manera alternativa para estimar la varianza de (2.1), (2.5) y (2.19) es a través del método conocido como “bootstrap” no paramétrico, el cual consiste en lo siguiente (Efron, 1979; Särndal *et al.* 1992; Pérez, 2000):

1. Estimar el parámetro poblacional θ a través de $\hat{\theta}$ con los datos de la muestra original de m observaciones. Se requiere la estimación de $V(\hat{\theta})$, es decir, $v(\hat{\theta})$.
2. A partir de la muestra original obtener un gran número de muestras (B muestras) de tamaño m con igual probabilidad y con reemplazo. Las B muestras seleccionadas de esta manera se llaman “muestras bootstrap”.
3. Para cada muestra “bootstrap” calcular una estimación $\hat{\theta}_b^*$ ($b = 1, \dots, B$) en la misma forma como $\hat{\theta}$ fue obtenida. El estimador (2.1) es una razón, (2.5) y (2.19) son una media.

4. La distribución de $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$ es considerada una estimación de la distribución de muestreo del estimador $\hat{\theta}$ y $V(\hat{\theta})$ es estimada mediante

$$v(\hat{\theta}) = \frac{\sum_{b=1}^B \left(\hat{\theta}_b^* - \hat{\theta} \right)^2}{B-1} \quad (2.21)$$

Donde $\hat{\theta}^* = \frac{\sum_{b=1}^B \hat{\theta}_b^*}{B}$

En cualquier caso, suponiendo una distribución de $\hat{\theta}$ aproximadamente normal o Gaussiana, un intervalo de confianza para θ está dado de la siguiente manera:

$$\hat{\theta} \pm 2 \left(v(\hat{\theta}) \right)^{1/2}$$

En los estados de Quintana Roo y Campeche la selva mediana subperennifolia es la más extensa. Los árboles que crecen en este ecosistema tienen contrafuertes y poseen una gran cantidad de epífitas y bejucos. Estos árboles alcanzan una altura de entre 15 y 25 m, con troncos menos gruesos que los árboles de la selva alta perennifolia, aun cuando se trata prácticamente de las mismas especies. Las palmas forman parte de los estratos, especialmente del bajo y del medio. Las especies características son Ox (ramón), chakah, ya' (zapote, chicozapote), tzalam, yaaxnik, pukte', ja'asche' y caoba. En las riberas de los ríos crece el kuyché (zapote bobo). Se encuentran también helechos y musgos, así como abundantes orquídeas, bromeliáceas y aráceas (Ek, 2011).

En la selva mediana subcaducifolia la altura promedio de los árboles varía de 25 a 30 m. La cobertura vegetal es menor que la observada en las selvas perennifolias y subperennifolias. Los suelos que sustentan a este tipo de vegetación son rocosos y pedregosos con una capa de materia orgánica, formada por la gran cantidad de hojas

que dejan caer los árboles. Las especies que destacan en estos ecosistemas vegetales son Ox (ramón), pich (orejón), jabín, ki, yaaxnik, bari, cascarillo, laurelillo, capulín, pukte', boxchechem, sakchechem, nance, pomolche', caimito, jícaro, bob, pichi'che', ciricote, tsitsilche', katzin, chakah blanco, muk, sakpa', hulub, ta-siste, tzalam, chechem blanco. También abundan gramíneas y ciperáceas tales como jucarillo. Las epífitas son orquídeas de diferentes tipos y colores, piperáceas como el cordoncillo; bromeliáceas, como el xku, y la piñuela y bejucos como el muk (Ek, 2011). En el ANEXO 2.2. se presentan los nombres científicos correspondientes a las especies mencionadas.

Aunada a su alta diversidad, la selva mediana subperennifolia es importante por su capacidad de almacenamiento de carbono. Eaton y Lawrence (2009) reportan que en la selva mediana subperennifolia de Campeche y Quintana Roo el carbono en la biomasa área viva oscila de 4.8 Mg ha^{-1} en un bosque secundario de tres años hasta 73.5 Mg ha^{-1} en un bosque maduro, mientras que el carbono aéreo combinado con el carbono subterráneo varía de 192 Mg ha^{-1} en un bosque secundario de 12 años hasta 469 Mg ha^{-1} en una selva madura.

2.6. MATERIALES Y MÉTODOS

Se estimaron los parámetros de área basal, número de árboles y volumen maderable promedio por hectárea mediante las siguientes expresiones:

Media de razones (Smelko y Merganic, 2008)

$$\bar{y}_{ha} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} \quad (2.19)$$

$$\text{Donde } y_{ha}(s) = \frac{y_s}{a_s} = \frac{\sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{n_s a_{sj}}$$

Razón de medias (Velasco *et al.*, 2005)

$$y_{ha} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} \quad (2.1)$$

$$\text{Donde } y_s = \sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk} ; a_s = n_s a_{sj}$$

Forest Inventory and Analysis (Scott *et al.*, 2005)

$$\bar{y}_{ha} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} \quad (2.5)$$

$$\text{Donde } y_{ha}(s) = \frac{y_s}{a_s} = \frac{\sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{a_0 \bar{p}_0}$$

En cada caso, las varianzas se obtuvieron i) según las expresiones convencionales reportadas en la literatura y 2) mediante procedimientos bootstrap. Para cada muestra de tamaño m se generaron 1000 muestras bootstrap. Además de las estimaciones

puntuales se obtuvieron las estimaciones por intervalo. Para el estimador de razón adicionalmente se estimó el sesgo mediante la siguiente expresión (2.22) (ANEXO 2.1):

$$\hat{B}(y_{ha}) = \frac{1}{m\bar{a}} (y_{ha}s_a^2 - s_{y,a}) \quad (2.22)$$

Donde

$$s_a^2 = \frac{\sum_{s=1}^m (a_s - \bar{a})^2}{m-1} \text{ y } s_{y,a} = \frac{\sum_{s=1}^m (y_s - \bar{y})(a_s - \bar{a})}{m-1}$$

El traslape de los intervalos de confianza y similitud en el error de muestreo se utilizó como criterio para comparar los estimadores estudiados.

Se utilizaron datos de selva mediana subperennifolia y subcaducifolia de los estados de Quintana Roo y Campeche, lo cuales se tomaron en campo en 2010 en el marco del proyecto de Inventario Nacional Forestal y de Suelos. Se usaron las ecuaciones de volumen fustal reportadas por SAG (1976) y SARH (1985); éstas se caracterizan por ser de aplicación a nivel estatal y por grupo botánico. Para el análisis de la información se elaboró un programa en el software Sistema de Análisis Estadístico (SAS, Versión 9.2). En este programa se generó una matriz (mxp); sus columnas correspondieron a la longitud, latitud, área basal, densidad arbórea, volumen maderable fustal y área efectivamente muestreada del conglomerado, y sus filas identificaron a los conglomerados.

En la Figura 2.4 se presenta la distribución espacial de los conglomerados utilizados en el estudio. Para Quintana Roo se usaron $m = 206$ conglomerados de selva mediana subperennifolia y $m = 22$ de selva mediana subcaducifolia, para Campeche $m = 127$ conglomerados de selva mediana subperennifolia y $m = 58$ de selva mediana subcaducifolia, lo cual permitió probar los estimadores en diferentes condiciones y distintos tamaños de muestra.

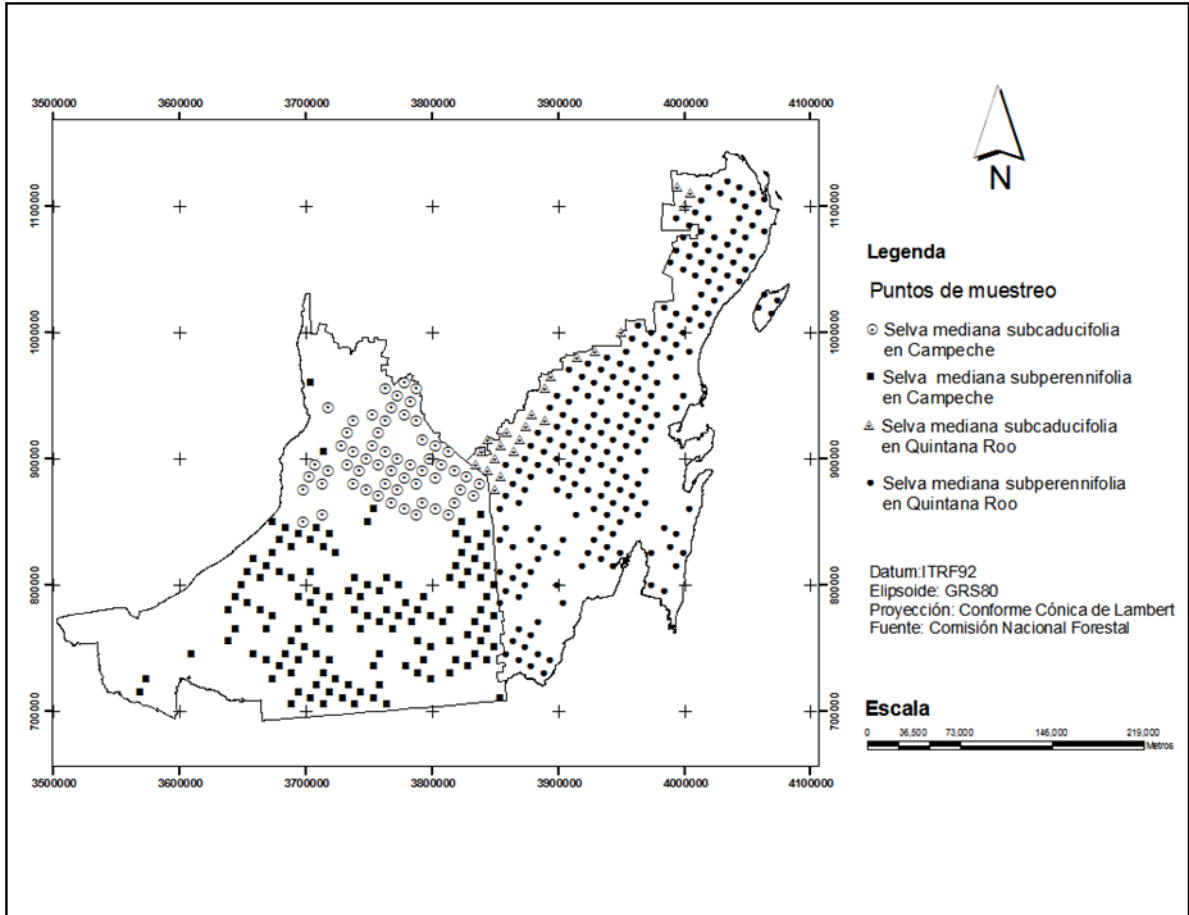


Figura 2.4. Distribución espacial de los conglomerados utilizados en el estudio. Selva mediana subperennifolia y subcaducifolia en Quintana y Campeche, México.

2.7. RESULTADOS Y DISCUSIÓN

La correlación entre las variables dasométricas y el área del conglomerado efectivamente muestreada fue en general bajo (alrededor del 0.42), sólo el número de árboles correspondiente a la selva mediana subcaducifolia de Quintana Roo y de las tres variables de selva mediana subperennifolia de Campeche, el coeficiente de correlación fue ligeramente superior al 0.5 (Cuadro 2.1). En todos los casos las correlaciones resultaron ser significativas al nivel del 5%.

Cuadro 2.1. Coeficientes de correlación de Pearson entre las variables dasométricas y el área del conglomerado efectivamente muestreada.

Población de interés	Conglomerados (m)	Área basal (m^2)	Número de árboles	Volumen (m^3)
Selva mediana subperennifolia de Quintana Roo	206	0.4398 ($<.0001$) [†]	0.4671 ($<.0001$)	0.3936 ($<.0001$)
Selva mediana subcaducifolia de Quintana Roo	22	0.4963 0.0188	0.6581 (0.0009)	0.4352 (0.0429)
Selva mediana subperennifolia de Campeche	127	0.5661 ($<.0001$)	0.5568 ($<.0001$)	0.5221 ($<.0001$)
Selva mediana subcaducifolia de Campeche	58	0.3960 (0.0021)	0.4568 (0.0003)	0.3826 (0.0030)

[†]El valor entre paréntesis indica el valor de p para $H_0: \rho = 0$.

En las Figuras 2.5, 2.6, 2.7 y 2.8 se presentan los diagramas de dispersión correspondiente a las variables de interés y al área del conglomerado efectivamente evaluado en campo. En todos los casos la varianza del área basal, número de árboles y volumen fue proporcional al área del conglomerado. Es interesante notar que mientras en la selva mediana subperennifolia (Quintana Roo y Campeche) los conglomerados pequeños (0.04, 0.08 y 0.12 ha) estuvieron bien representados, en la selva mediana subcaducifolia sucedió lo contrario, así en Quintana Roo ($m = 22$) no hubo conglomerados con área de 0.12 ha, y en Campeche ($m = 58$) solo un conglomerado tuvo una área de 0.04 ha y otro con área de 0.08 ha. Esta última situación puede deberse al tamaño de muestra relativamente bajo. La línea sólida corresponde a la recta de regresión de las variables dasométricas sobre el área del conglomerado (ha).

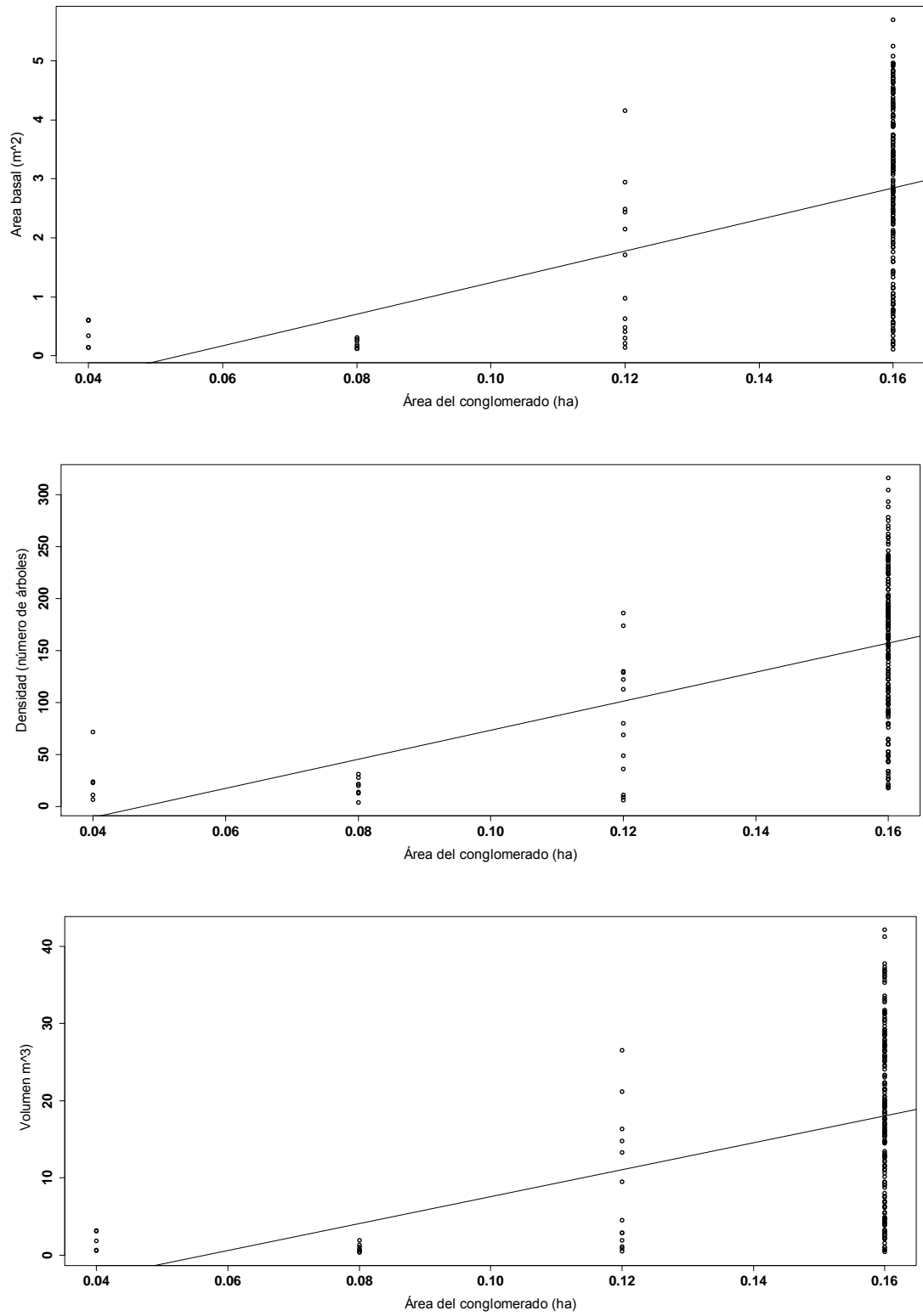


Figura 2.5. Diagramas de dispersión y líneas de regresión de las variables de interés sobre el área del conglomerado (ha) en selva mediana subperennifolia en Quintana Roo.

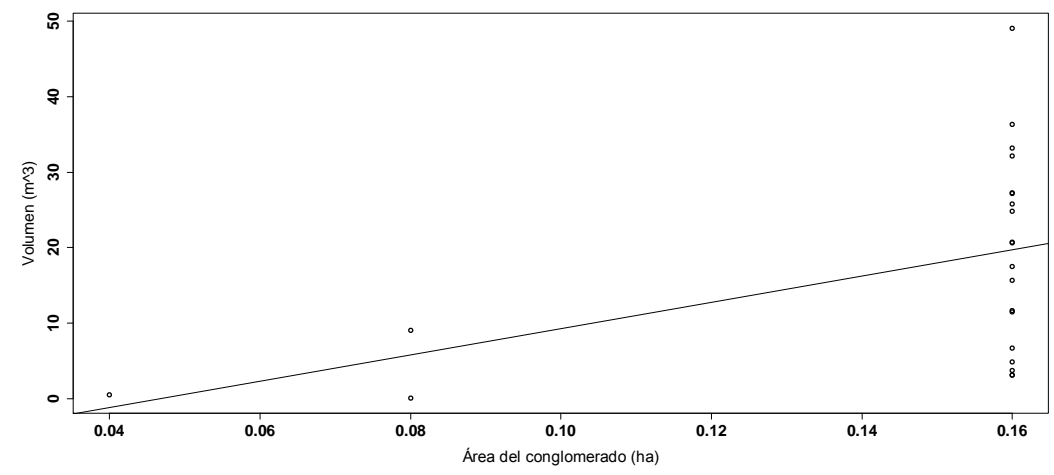
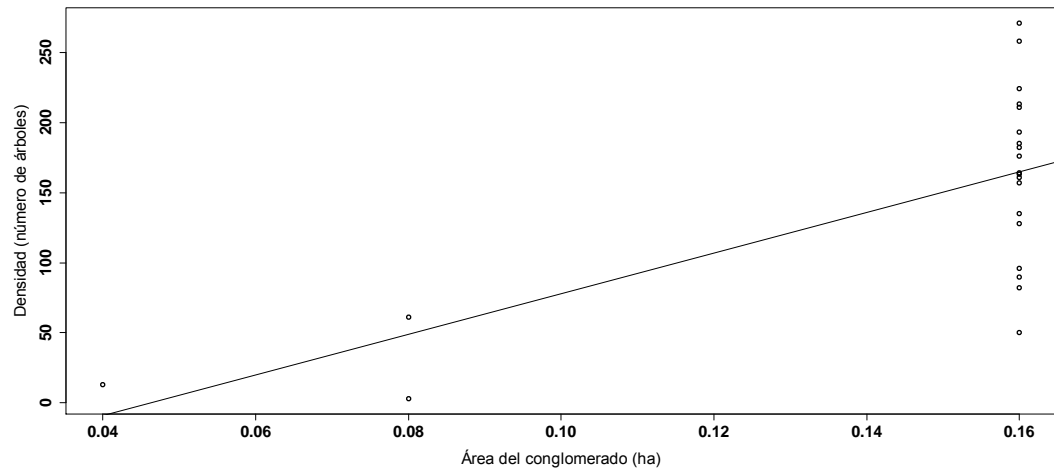
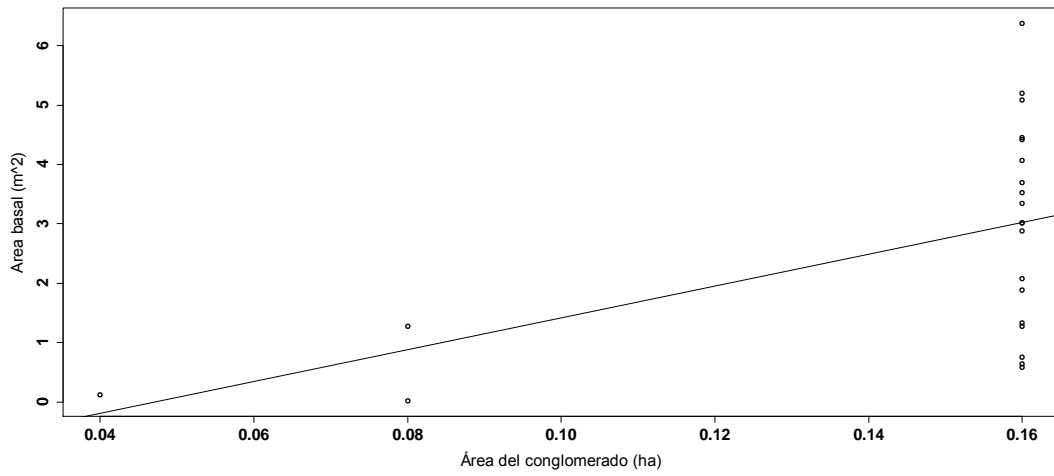


Figura 2.6. Diagramas de dispersión y líneas de regresión de las variables de interés sobre el área del conglomerado (ha) en selva mediana subcaducifolia en Quintana Roo.

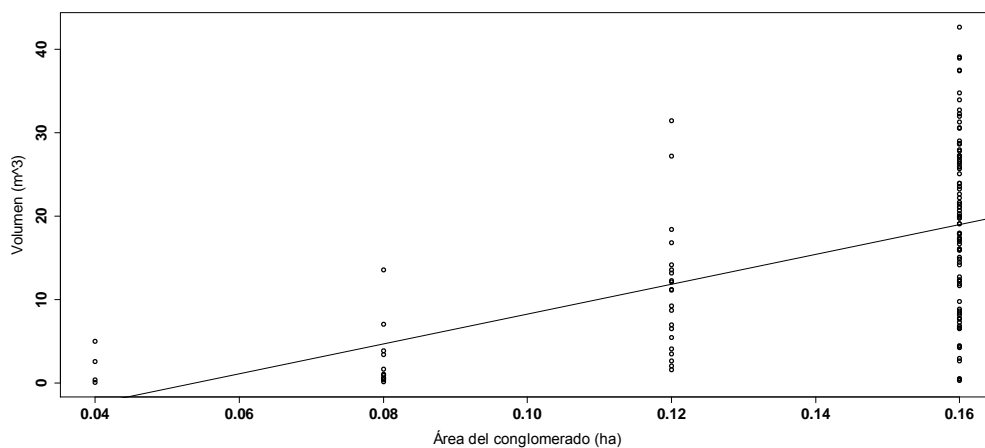
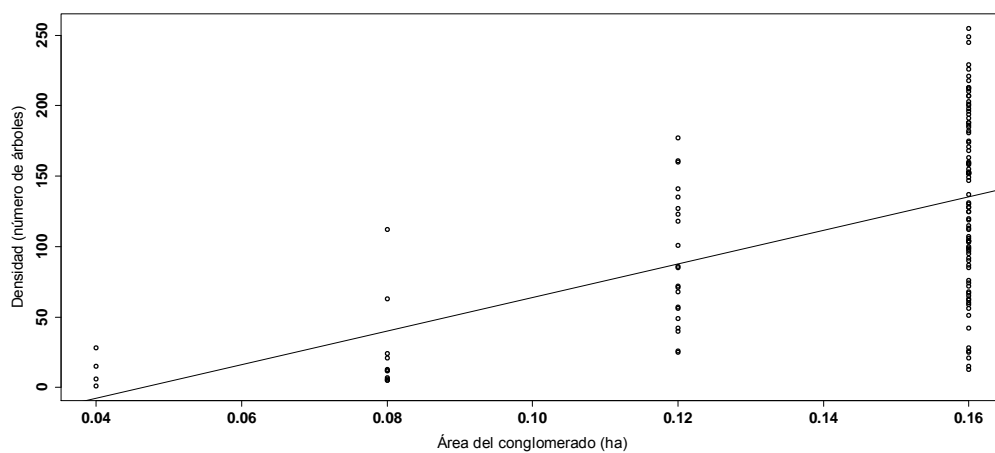
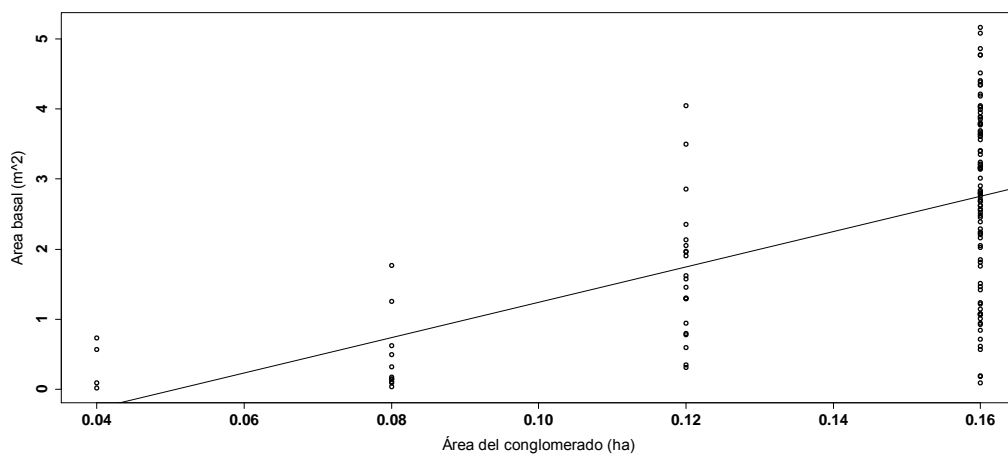


Figura 2.7. Diagramas de dispersión y líneas de regresión de las variables de interés sobre el área del conglomerado (ha) en selva mediana subperennifolia en Campeche.

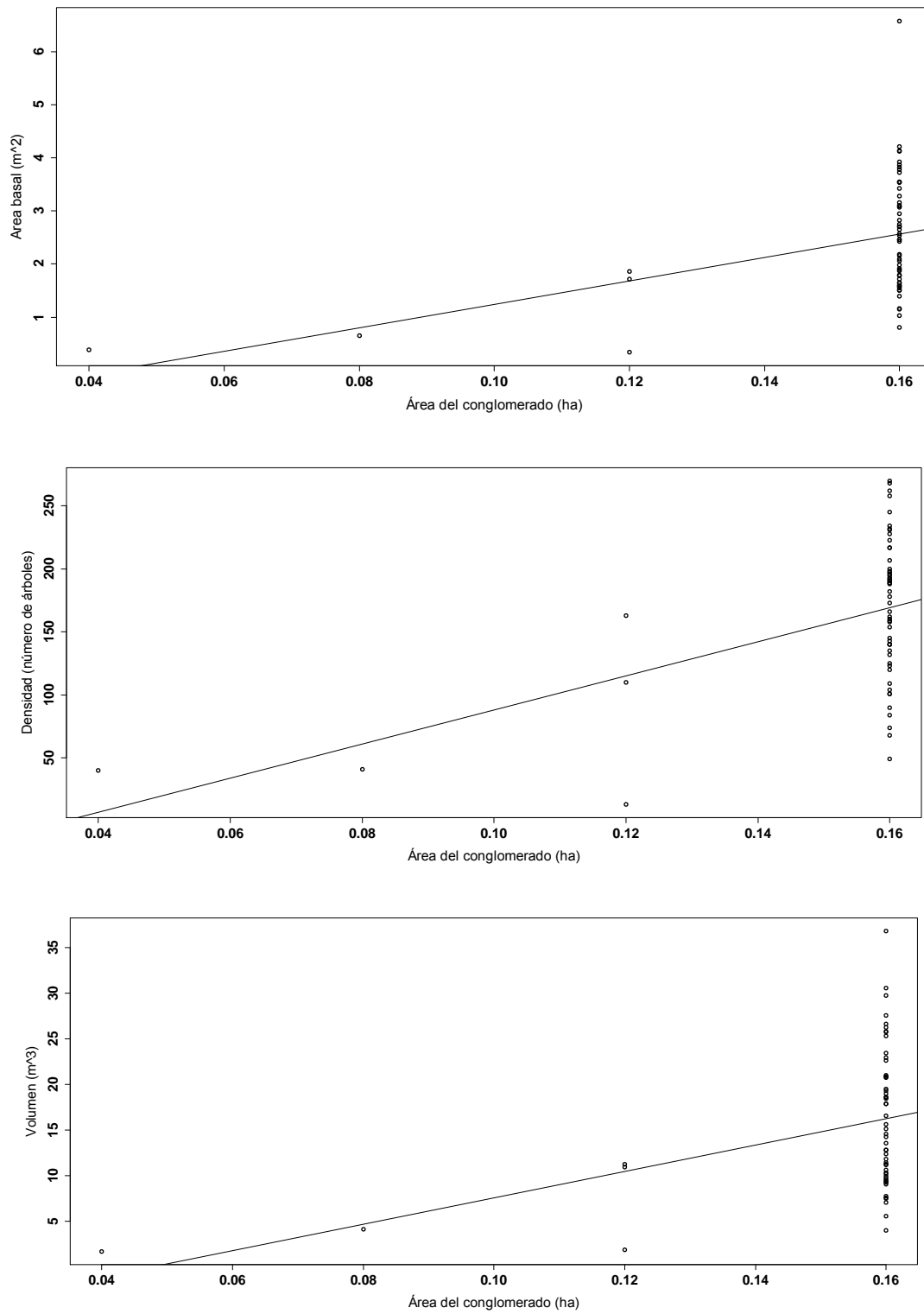


Figura 2.8. Diagramas de dispersión y líneas de regresión de las variables de interés sobre el área del conglomerado (ha) en selva mediana subcaducifolia en Campeche.

En el Cuadro 2.2 se indican los estadísticos de ajuste y los coeficientes estimados de las regresiones correspondientes a las variables dasométricas sobre el área del conglomerado. Tanto para el área basal, número de árboles y volumen de la selva mediana subperennifolia de Quintana Roo y Campeche el valor de p de la prueba de t indica que el intercepto es significativo al 5%, es decir, se rechaza $H_0: \beta_0 = 0$. Sólo para las variables forestales de la selva mediana subcaducifolia, en las dos entidades federativas, no se rechaza $H_0: \beta_0 = 0$. Dada la dispersión de las observaciones, en ningún caso el coeficiente de determinación superó al 0.5, y los cuadrados medios del error fueron altos.

La situación anterior pone en desventaja al estimador de razón de medias, mismo que para ser efectivo requiere un coeficiente de correlación mayor al 0.5 entre la variable de interés y la variable auxiliar, y la relación entre ambas debe ser una línea recta que pase por el origen. Sin embargo, debe tomarse en cuenta que en tres de los cuatros escenarios estudiados el número de observaciones rebasa en gran medida el tamaño de muestra sugerido, el cual debe ser por lo menos de 30.

En los Cuadros 2.3 y 2.4 se presentan los resultados del análisis estadístico correspondiente a la selva mediana subperennifolia y subcaducifolia de Quintana Roo y Campeche. En las dos comunidades vegetales de Quintana Roo y Campeche y para los tres variables forestales (área basal, densidad y volumen por hectárea) se nota que los intervalos de confianza (función de la varianza) obtenidos por el método convencional y por bootstrap, para los tres estimadores evaluados, son similares. En todos los casos los promedios de las muestras bootstrap son parecidos a las estimaciones puntuales de la muestra original. Los valores puntuales por hectárea calculados mediante los tres estimadores —por método convencional y bootstrap—, en cada población, también son parecidos. El sesgo del estimador de razón, a pesar de ser negativo, siempre fue cercano a cero; además, el sesgo relativo (sesgo respecto al error estándar) siempre fue menor a 10%, inclusive donde se tuvo un tamaño de muestra de solo 22 observaciones. Por lo que, el sesgo del estimador RM es prácticamente despreciable.

Cuadro 2.2. Estadísticos de ajuste y coeficientes de regresión estimados para modelos lineales simples de variables dasométricas (y) sobre área de conglomerado (a). Modelo $y = \beta_0 + \beta_1 a$.

Población de interés	Variable dependiente	CME	R ²	Parámetros	Estimación	Error estándar	Valor de t	Valor de p
Selva mediana subperennifolia de Quintana Roo	Área basal	1.8205	0.1934	β_0	-1.4420	0.5896	-2.45	0.0153
				β_1	26.8105	3.8332	6.99	<.0001
	Número de árboles	4243.2368	0.2182	β_0	-66.4505	28.4647	-2.33	0.0205
				β_1	1396.3171	185.0609	7.55	<.0001
	Volumen	100.3279	0.1550	β_0	-9.8413	4.3769	-2.25	0.0256
				β_1	174.0431	28.4562	6.12	<.0001
Selva mediana subcaducifolia de Quintana Roo	Área basal	2.60139	0.2463	β_0	-1.2662	1.5828	-0.80	0.4331
				β_1	26.8191	10.4909	2.56	0.0188
	Número de árboles	3248.7535	0.4331	β_0	-67.2277	55.9357	-1.20	0.2435
				β_1	1449.0769	370.7388	3.91	0.0009
	Volumen	152.4613	0.1894	β_0	-8.0826	12.1174	-0.67	0.5124
				β_1	173.6342	80.3137	2.16	0.0429
Selva mediana subperennifolia de Campeche	Área basal	1.3298	0.3204	β_0	-1.2750	0.4792	-2.66	0.0088
				β_1	25.1930	3.2815	7.68	<.0001
	Número de árboles	3108.1960	0.3101	β_0	-55.0969	23.1700	-2.38	0.0189
				β_1	1189.1451	158.6495	7.50	<.0001
	Volumen	83.9232	0.2726	β_0	-9.5505	3.8073	-2.51	0.0134
				β_1	178.4356	26.0691	6.84	<.0001
Selva mediana subcaducifolia de Campeche	Área basal	1.1153	0.1568	β_0	-0.9703	1.0658	-0.91	0.3665
				β_1	22.0765	6.8408	3.23	0.0021
	Número de árboles	2965.0215	0.2087	β_0	-47.4259	54.9539	-0.86	0.3918
				β_1	1355.3241	352.7045	3.84	0.0003
	Volumen	51.9419	0.1464	β_0	-6.9119	7.2735	-0.95	0.3461
				β_1	144.6655	46.6827	3.10	0.0030

Cuadro 2.3. Estimaciones puntuales y por intervalo de parámetros forestales en selva mediana subperennifolia y subcaducifolia en Quintana Roo bajo diferentes estimadores.

Población de interés	Variable	Estimador	Estimación de la varianza	Estimación puntual	Límite inferior	Límite superior	Error de muestreo (%)	Valor de p [§]	Sesgo	
Selva mediana subperennifolia (m = 206)	Área basal (m ² ha ⁻¹)	Media de razones	Convencional	16.8333	15.5559	18.1107	7.58828	>0.15,>0.25,>0.25		
			Bootstrap	16.8393 [¶]	15.5620	18.1166	7.58537			
		Razón de medias	Convencional	17.3136	16.0599	18.5672	7.24071			-0.0012 (0.0019) ^{¶¶}
			Bootstrap	17.3314	16.0700	18.5928	7.27826			>0.15,>0.25,>0.25
		FIA*	Convencional	17.3136	15.9382	18.6889	7.94365			
			Bootstrap	17.3144	15.9411	18.6877	7.93170			>0.15,>0.25,>0.25
	Densidad (árboles ha ⁻¹)	Media de razones	Convencional	937.571	875.163	999.98	6.65631	>0.15,>0.25,>0.25		
			Bootstrap	938.037	874.517	1001.56	6.77164			
		Razón de medias	Convencional	958.696	898.251	1019.14	6.30486			-0.0557 (0.0018)
			Bootstrap	959.797	898.068	1021.53	6.43146			>0.15,>0.25,>0.25
		FIA	Convencional	958.695	891.253	1026.14	7.03476			
			Bootstrap	958.854	890.370	1027.34	7.14229			>0.15,>0.25,>0.25
Volumen (m ³ ha ⁻¹)	Media de razones	Convencional	105.896	96.5405	115.251	8.83447	>0.15,>0.25,>0.25			
		Bootstrap	105.943	96.5643	115.321	8.85232				
	Razón de medias	Convencional	109.231	99.9458	118.517	8.50065			-0.0082 (0.0018)	
		Bootstrap	109.355	99.9778	118.733	8.57533			>0.15,>0.25,>0.25	
	FIA	Convencional	109.231	99.2562	119.206	9.13187				
		Bootstrap	109.249	99.2555	119.242	9.14715			>0.15,>0.25,>0.25	
Selva mediana subcaducifolia (m = 22)	Área basal (m ² ha ⁻¹)	Media de razones	Convencional	17.2392	12.5455	21.9330	27.2273	>0.15,>0.25,>0.25		
			Bootstrap	17.1812	12.5434	21.8191	26.9938			
		Razón de medias	Convencional	18.2215	13.5883	22.8546	25.4269			-0.0203 (.0088)
			Bootstrap	18.1651	13.5567	22.7736	25.3697			>0.15,>0.25,>0.25
		FIA	Convencional	18.2216	12.9722	23.4709	28.8083			
			Bootstrap	18.1244	12.9543	23.2946	28.5259			>0.15,>0.25,>0.25
	Densidad (árboles ha ⁻¹)	Media de razones	Convencional	942.898	761.088	1124.71	19.2820	0.05,0.02,0.02		
			Bootstrap	942.111	760.086	1124.14	19.3210			
		Razón de medias	Convencional	992.593	825.550	1159.63	16.8289			-1.0768 (0.0129)
			Bootstrap	992.037	823.930	1160.14	16.9457			0.06,0.05,0.06
		FIA	Convencional	992.597	778.705	1206.49	21.5487			
			Bootstrap	989.910	776.076	1203.74	21.6014			0.03,<0.01,<0.05
Volumen (m ³ ha ⁻¹)	Media de razones	Convencional	112.347	77.2999	147.393	31.1951	0.11,0.23,0.18			
		Bootstrap	111.923	76.6802	147.167	31.4887				
	Razón de medias	Convencional	118.752	83.4590	154.046	29.7201			-0.1295 (0.0073)	
		Bootstrap	118.333	82.5182	154.148	30.2660			>0.15,>0.25,0.17	
	FIA	Convencional	118.753	80.0014	157.504	32.6321				
		Bootstrap	118.059	79.2804	156.837	32.8467			0.10,0.10,0.09	

*Forest Inventory and Analysis. [¶] Promedio de las medias bootstrap ($\hat{\theta}$). [§] El valor de p corresponde a las pruebas de normalidad de Kolmogorov-Smirnov, Cramer-von Mises y Anderson-Darling (en ese orden) de las distribución de las muestras bootstrap. ^{¶¶} El valor en paréntesis de la columna de sesgo es el sesgo relativo.

Cuadro 2.4. Estimaciones puntuales y por intervalo de parámetros forestales en selva mediana subperennifolia y subcaducifolia en Campeche bajo diferentes estimadores.

Población de interés	Variable	Estimador	Estimación de la varianza	Estimación puntual	Límite inferior	Límite superior	Error de muestreo (%)	Valor de p [§]	Sesgo
Selva mediana subperennifolia (m = 127)	Área basal (m ² ha ⁻¹)	Media de razones	Convencional	15.5389	14.0312	17.0466	9.7029		
			Bootstrap	15.4383[¶]	13.8653	17.0112	10.1884	>0.15,>0.25, >0.25	
		Razón de medias	Convencional	16.2565	14.7860	17.7269	9.0453		-0.0034 (0.0046) [¶]
			Bootstrap	16.1498	14.6240	17.6757	9.4480	>0.15,>0.25, >0.25	
		FIA*	Convencional	16.2565	14.5234	17.9896	10.6608		
			Bootstrap	16.1365	14.3370	17.9361	11.1520	>0.15,>0.25, >0.25	
	Densidad (árboles ha ⁻¹)	Media de razones	Convencional	768.701	695.448	841.954	9.5295		
			Bootstrap	766.667	693.346	839.987	9.5635	>0.15,>0.25, >0.25	
		Razón de medias	Convencional	802.980	732.291	873.669	8.8034		-0.1464 (0.0041)
			Bootstrap	800.891	730.991	870.790	8.7277	>0.15,0.14, 0.19	
		FIA	Convencional	802.982	719.825	886.140	10.3561		
			Bootstrap	800.209	716.660	883.758	10.4409	>0.15,0.11, 0.17	
Volumen (m ³ ha ⁻¹)	Media de razones	Convencional	106.096	94.3206	117.871	11.0988			
		Bootstrap	105.263	93.1551	117.371	11.5025	>0.15,>0.25, >0.25		
	Razón de medias	Convencional	111.498	99.8524	123.143	10.4443		-0.0254 (0.0043)	
		Bootstrap	110.602	98.6776	122.526	10.7810	>0.15,>0.25, >0.25		
	FIA	Convencional	111.498	98.1901	124.806	11.9355			
		Bootstrap	110.512	96.8654	124.159	12.3485	>0.15,>0.25, >0.25		
Selva mediana subcaducifolia (m = 58)	Área basal (m ² ha ⁻¹)	Media de razones	Convencional	15.5847	13.8011	17.3684	11.4447		
			Bootstrap	15.6413	13.8281	17.4544	11.5921	<0.01, 0.02, 0.02	
		Razón de medias	Convencional	15.7953	14.0025	17.5881	11.3504		-0.0019 (0.0021)
			Bootstrap	15.8527	14.0249	17.6806	11.5304	0.09,0.05, 0.04	
		FIA	Convencional	15.7953	13.8574	17.7332	12.2689		
			Bootstrap	15.8529	13.8834	17.8224	12.4235	<0.01,0.02,0.01	
	Densidad (árboles ha ⁻¹)	Media de razones	Convencional	1039.80	946.313	1133.28	8.9907		
			Bootstrap	1041.68	944.028	1139.34	9.3747	>0.15,>0.25, >0.25	
		Razón de medias	Convencional	1048.33	955.957	1140.69	8.8111		-0.0927 (0.0020)
			Bootstrap	1050.20	953.174	1147.23	9.2388	>0.15,>0.25, >0.25	
		FIA	Convencional	1048.32	945.184	1151.46	9.8385		
			Bootstrap	1050.18	942.897	1157.46	10.2157	0.11,0.15, 0.17	
Volumen (m ³ ha ⁻¹)	Media de razones	Convencional	98.324	86.1031	110.546	12.4295			
		Bootstrap	98.720	86.3561	111.084	12.5242	>0.15,0.07, 0.05		
	Razón de medias	Convencional	99.923	87.6805	112.166	12.2523		-0.0135 (0.0022)	
		Bootstrap	100.327	87.8713	112.782	12.4147	>0.15,0.13, 0.10		
	FIA	Convencional	99.923	86.7794	113.067	13.1539			
		Bootstrap	100.327	87.0430	113.611	13.2408	0.11,0.07, 0.05		

*Forest Inventory and Analysis. [¶] Promedio de las medias bootstrap ($\hat{\theta}$). [§] El valor de p corresponde a las pruebas de normalidad de Kolmogorov-Smirnov, Cramer-von Mises y Anderson-Darling (en ese orden) de las distribución de las muestras bootstrap. [¶] El valor en paréntesis de la columna de sesgo es el sesgo relativo.

Resulta interesante observar que en todos los casos el estimador de razón de medias, a pesar de que en general no existe una fuerte correlación entre la variable de interés y la variable auxiliar (Cuadro 2.1), es ligeramente más preciso (menor error de muestreo), seguido del estimador media de razones y del FIA; al existir traslape de los intervalos de confianza al 95%, puede considerarse que los tres estimadores son de igual precisión respecto al promedio (tienen medias que comparten intervalos de confianza similares). Adicionalmente se aprecia que para todas las variables, en cada una de las poblaciones estudiadas, los errores de muestreo de los estimadores estudiados presentan alto grado de similitud.

En las dos poblaciones estudiadas de cada estado y para las tres variables analizadas, el estimador FIA tuvo una precisión ligeramente más baja que los otros dos, de manera consistente se observa que presenta mayor error de muestreo incluso en donde el tamaño de muestra es superior a 100. Esto posiblemente puede deberse a que el denominador de la expresión que permite obtener los valores por hectárea considera un área ajustada común, el área del conglomerado originalmente deseable (0.16 ha) es ajustada por una proporción media de todas las áreas efectivamente muestreadas en campo.

Los resultados anteriores son congruentes con otros estudios realizados sobre la comparación de estimadores media de razones y razón de medias en el ámbito forestal. Al respecto, Smelko y Saborowski (1999) con datos de la región forestal del norte de Slovakia, 32 parcelas muestrales de diferente tamaño (desde 100 hasta 1000 m² en rangos de 100) levantadas en 1991 y 1995, estudiaron los estimadores media de razones (método A) y razón de medias (método B) en la estimación del número de árboles por ha y del volumen maderable por ha. Las correlaciones entre el número de árboles y el área de las parcelas fueron bajas (0.323 y 0.278 para 1991 y 1995, respectivamente), mientras que las correlaciones entre el volumen y el área de las parcelas fueron moderadas (0.82 ambas en ocasiones); en este caso segundo no se observó una relación de línea recta que pase por el origen. En los dos años de estudio el método A dio menor volumen por hectárea que el B (1991: A 423.3 y B 453.9 m³ ha⁻¹; 1995: A 443.5 y B 470 m³ ha⁻¹) y también errores estándar más pequeños que el B.

Sucedió lo contrario al volumen para el número de árboles por hectárea. Para las dos ocasiones y las dos variables las diferencias de los errores estándar en ambos métodos no superaron el 2.7% (Cuadro 2.5).

Cuadro 2.5. Estimación del número de árboles y volumen por hectárea en Orova, Slovakia (Smelko y Saborowski, 1999).

Variables	Método [†]	Año 1991			Año 1995		
		Estimación puntual	Error estándar absoluto	Error estándar relativo	Estimación puntual	Error estándar absoluto	Error estándar relativo
Densidad (Árboles ha ⁻¹)	A	602.3	61.3	10.2	555.8	51.2	9.2
	B	485.8	37.2	7.6	456.1	34.2	7.5
Volumen (m ³ ha ⁻¹)	A	423.3	27.9	6.6	443.5	27.5	6.2
	B	453.9	33.3	7.3	470.0	34.1	7.2

[†]El método A es el estimador media de razones y el método B es el estimador razón de medias.

Smelko y Merganic (2008) evaluaron el estimador de razón de medias y media de razones con datos del Inventario Nacional Forestal y Monitoreo de Slovakia. Para el volumen maderable comercial reportan los siguientes resultados por hectárea: $y_{ha,RM} = 266.2m^3$, $(v(y_{ha,RM}))^{0.5} = 5.15m^3$, $y_{ha,MR} = 263.9m^3$ y $(v(y_{ha,MR}))^{0.5} = 5.16m^3$. Esto conduce a diferencias en valores promedio por ha de $2.3 m^3$ y en error estándar de $0.01 m^3$. Entonces, bajo las condiciones estudiadas en Slovakia los dos métodos son prácticamente iguales en cuanto a precisión. No obstante, el estimador media de razones tiene como ventaja la obtención inmediata de estimaciones por hectárea, lo cual permite evaluar la variabilidad espacial de la variable de interés en el área de estudio, además la fórmula de la varianza es relativamente más sencilla que el estimador de razón de medias.

El efecto del tamaño de muestra en cada comunidad vegetal se refleja en la distribución de probabilidad de las muestras bootstrap; así, en la selva mediana subcaducifolia de Quintana Roo con $m = 22$ observaciones, en los tres estimadores la prueba de normalidad es aceptada de manera no contundente para el parámetro de densidad ($\alpha = 0.05$). En la selva mediana subcaducifolia de Campeche con $m = 58$ observaciones, en los estimadores media de razones y FIA para el área basal se rechaza la normalidad ($\alpha = 0.05$). Cuando el tamaño de muestra rebasa las 100 observaciones (selva mediana subperennifolia) en todos los casos la normalidad de las muestras bootstrap se obtiene sin mayor dificultad.

En las Figuras 2.9 y 2.10 se presenta la distribución de las muestras bootstrap para el área basal, densidad arbórea y volumen por hectárea resultante del estimador media de razones de selva mediana subperennifolia de Quintana Roo y Campeche, respectivamente; se observa que en efecto las muestras bootstrap correspondientes a las variables forestales estudiadas se ajustan a una distribución normal, comportamiento congruente con el valor de p de la prueba de normalidad. Las estimaciones puntuales y los límites inferior y superior de los intervalos de confianza al 95% (valores resaltados en los Cuadros 2.3 y 2.4) son congruentes con la siguiente propiedad de la distribución Gaussiana: en una distribución normal el 95% de las observaciones se encuentran dentro de dos desviaciones estándar por arriba y por debajo de la media (Figura 2.9 y 2.10).

Adicionalmente el tamaño de muestra tiene un efecto notable en el error de muestreo, así por ejemplo para el volumen estimado por FIA, cuando $m = 206$ el error de muestreo máximo es de 9.15%, cuando $m = 127$ el error máximo de muestreo es de 12.35%, cuando $m = 58$ el error máximo de muestreo es de 13.24% y cuando $m = 22$ el error máximo de muestreo asciende hasta 32.85%.

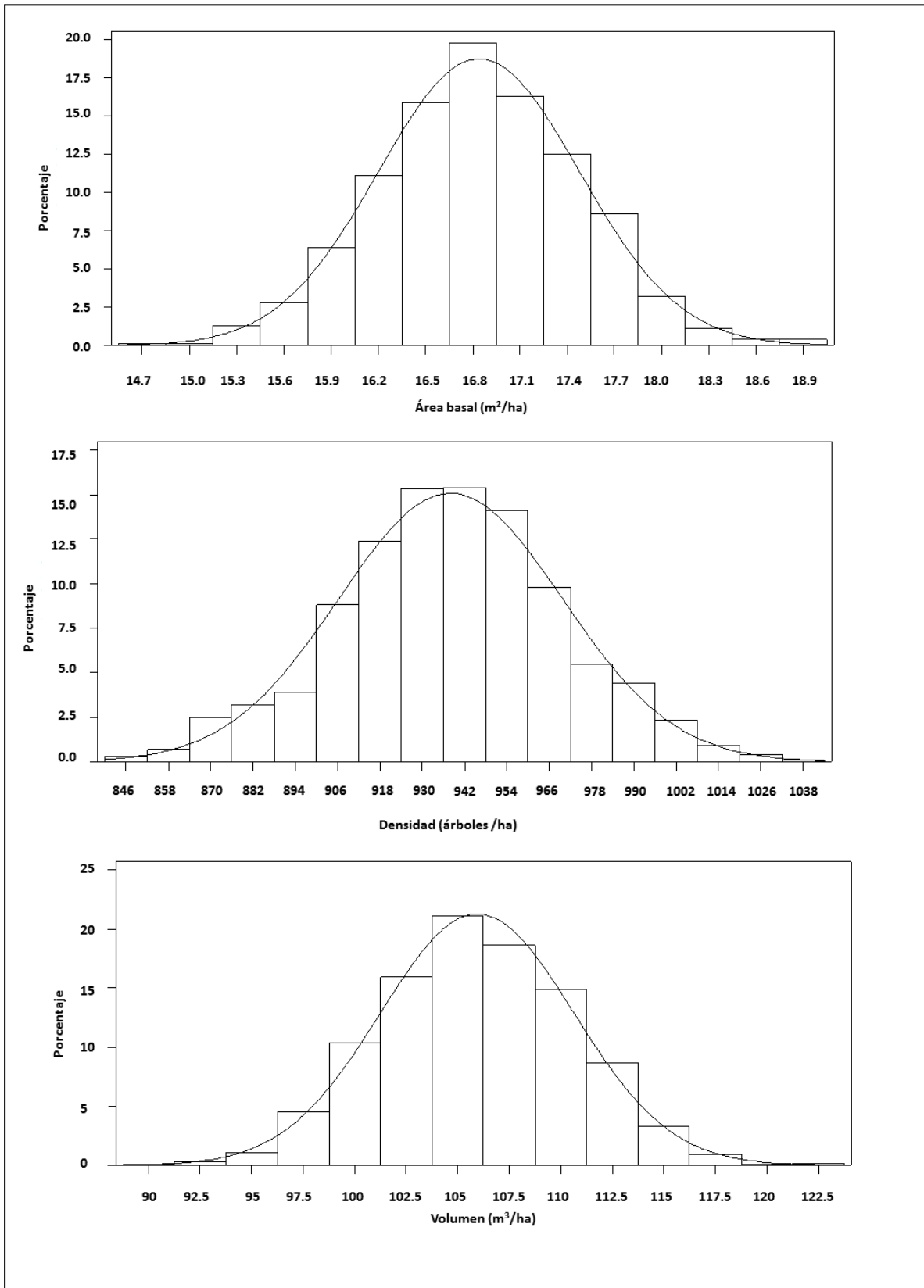


Figura 2.9. Distribución de las muestras bootstrap de selva mediana subperennifolia de Quintana Roo. Método media de razones.

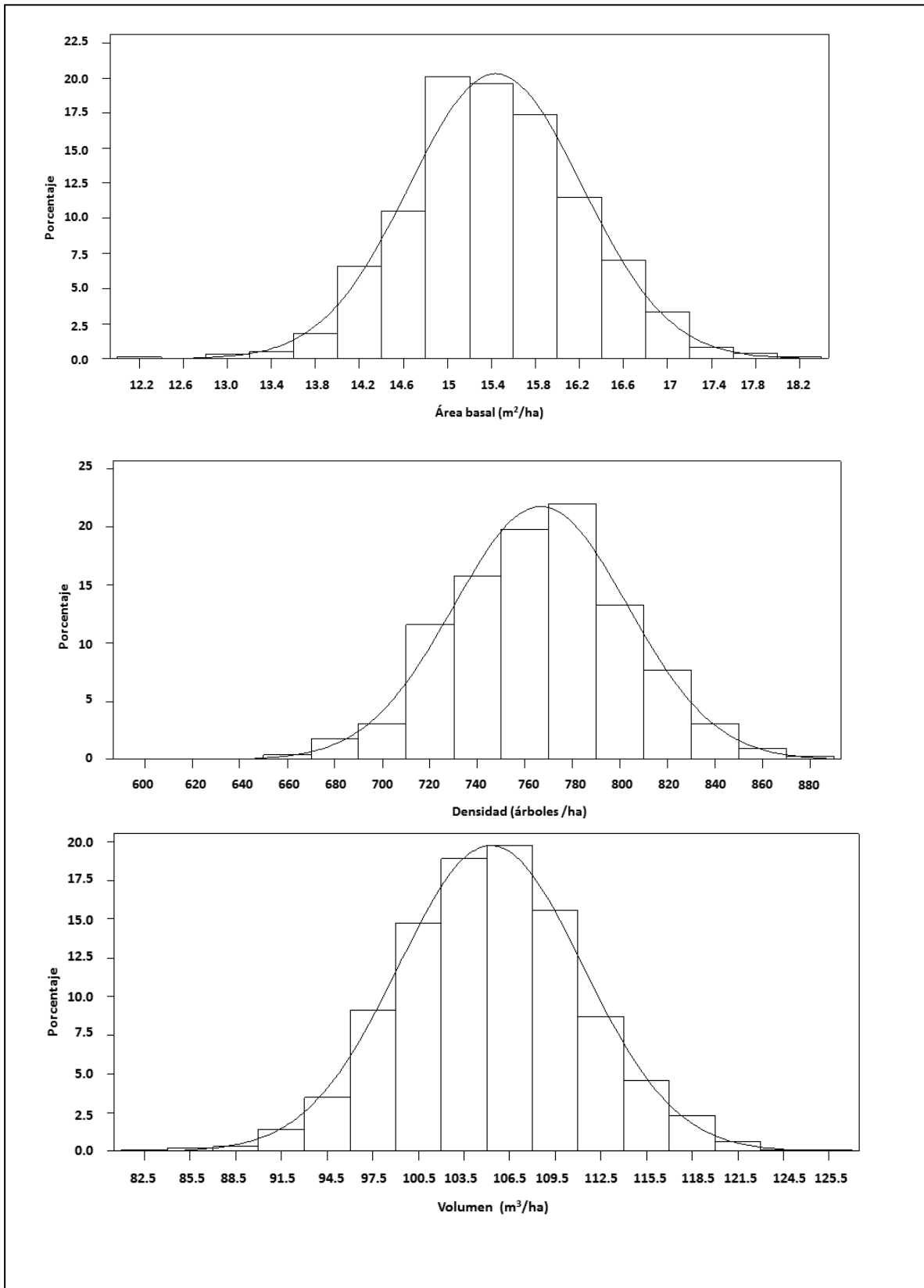


Figura 2.10. Distribución de las muestras bootstrap de selva mediana subperennifolia de Campeche. Método media de razones.

2.8. CONCLUSIONES

Aun cuando el grado de correlación entre las variables dasométricas y el área del conglomerado efectivamente muestreada en campo en general fue bajo (alrededor del 0.42), y en algunos casos se rechazó la hipótesis de que el intercepto es cero, el estimador de razón se justifica porque en tres de las cuatro poblaciones evaluadas el número de observaciones rebasa en gran medida el tamaño de muestra sugerido, el cual debe ser por lo menos de 30 observaciones.

En cada una de las situaciones ensayadas y para cada variable estudiada se observó traslape de los intervalos de confianza al 95% y similitud en los errores de muestreo correspondientes a los estimadores evaluados; por tanto, se concluye que los estimadores evaluados son equivalentes, aunque el estimador de razón es ligeramente más preciso. No obstante, el estimador de media de razones se prefiere porque su derivación teórica considera la estructura del conglomerado (probabilidades de inclusión) y la estimación directa de valores por hectárea (para los m conglomerados) permite evaluar la variabilidad espacial de la variable bajo estudio.

El tamaño de muestra tuvo un efecto significativo en la precisión de los estimadores evaluados en este estudio. Se observó que en tamaños de muestra menores a 30, los errores de muestreo ascienden hasta el 30% en la estimación del volumen maderable.

En todos los casos, el estimador FIA presentó un error de muestreo relativo ligeramente mayor comparado con MR y RM. De manera consistente, el estimador de media de razones proporciona valores ligeramente más bajos (conservadores) en comparación con los otros dos estimadores.

El sesgo del estimador de razón, a pesar de ser negativo, fue pequeño en todas las condiciones estudiadas; el sesgo relativo siempre fue menor a 10%, inclusive cuando se tuvo un tamaño de muestra de 22 observaciones únicamente. Por estas razones se consideró prácticamente despreciable.

El método bootstrap, procedimiento alternativo para la estimación de las varianzas, proporciona resultados tan confiables como aquellos obtenidos por los métodos convencionales reportados en la literatura. Bootstrap tiene la ventaja de ser intuitivamente sencillo y de fácil implementación desde el punto vista computacional. La normalidad de las muestras bootstrap se logra sin dificultad para tamaños de muestra superiores a 100. En este caso, el modelo normal ajustado permite apreciar directamente el límite inferior y superior de los intervalos de confianza, cualquier que sea al parámetro de interés.

2.9. LITERATURA CITADA

- Barabesi, L. and L. Fattorini 1998. The use of replicated plot, line and point sampling for estimating species abundance and ecological diversity. *Environmental and Ecological Statistics* 5: 353-370.
- Barabesi, L. and C. Pisani 2004. Steady – state ranked set sampling for replicated environmental designs. *Environmetrics* 15: 45-46.
- Bechtold, W. A. and C. T. Scott. 2005. The Forest Inventory and Analysis Plot Design. In: The enhanced Forest Inventory and Analysis program-national sampling design and estimation procedures. Eds. W. A. Bechtold and P. L. Patterson. General Technical Report SRS-80. Forest Service. United States Department of Agriculture. Southern Research Station, Asheville, North Carolina. USA. pp:27-42.
- Cochran, W. G. 1993. *Técnicas de Muestreo*. Compañía Editorial Continental. México. 513 p.
- Comisión Nacional Forestal (CONAFOR). 2012. Manual y procedimientos para el muestreo de campo: Re-muestreo 2012. SEMARNAT. México. 136 p.
- Eaton, J.M. and D. Lawrence. 2009. Loss of carbon sequestration potential after several decades of shifting cultivation in the Southern Yucatán. *Forest Ecology and Management* 258: 949-958.
- Ek D., A. 2011. Vegetación. In: *Riqueza Biológica de Quintana Roo. Un análisis para su conservación*. Tomo I. C. Pozo, N. Armijo C. y S. Calmé (Editoras). El Colegio de la Frontera Sur (Ecosur), Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (Conabio), Gobierno del Estado de Quintana Roo y Programa de Pequeñas Donaciones (ppd). México, D. F. pp: 62-77.
- Efron, B. 1979. Bootstrap methods: another look at the Jackknife. *The Annals of Statistics* 7(1):1-26.
- Gregoire, T.G. and H.T. Valentine. 2008. *Sampling Strategies for Natural Resources and the Environment*. Chapman & Hall/CRC. New York, USA. 474 p.
- Pérez L., C. 2000. *Técnicas de Muestreo Estadístico. Teoría, práctica y aplicaciones informáticas*. Alfaomega Grupo Editor. México. 603 p.
- Särndal, C. E., B. Swensson and J. Wretman. 1992. *Model Assisted Survey Sampling*. Springer-Verlag. New York, USA. 694 p.
- Scheaffer, R. L., W. Mendenhall, R. L. Ott and K.G. Gerow. 2012. *Elementary Survey Sampling*. Seventh Edition. Brooks/Cole Cengage Learning. Boston. USA. 436 p.

- Schreuder, H.T., T. G. Gregoire and G. B. Wood. 1993. Sampling Methods for Multiresource Forest Inventory. John Wiley & Sons, Inc. New York, USA. 446 p.
- Schreuder, H. T., Williams, M. S., Aguirre-Bravo, C. and Patterson, P. L. 2003. Statistical strategy for inventorying and monitoring the ecosystem resources of the Mexican States of Jalisco and Colima at multiple scales and resolution levels. Gen. Tech. Rep. RMRS-GTR-107. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 15 p.
- Scott, C.T., W. A. Bechtold, G. A. Reams, W. D. Smith, J. A. Westfall, M. H. Hansen and G.G. Moisen. 2005. Sample-based estimators used by the Forest Inventory and Analysis national information management system. In: The enhanced Forest Inventory and Analysis program-national sampling design and estimation procedures. W. A. Bechtold and P. L. Patterson (Eds.). General Technical Report SRS-80. Forest Service. United States Department of Agriculture. Southern Research Station, Asheville, North Carolina. USA. pp: 43-67.
- Secretaría de Agricultura y Ganadería (SAG). 1976. Inventario forestal del estado de Quintana Roo. Publicación Número 41. México. 44 p.
- Secretaría de Agricultura y Recursos Hidráulicos (SARH). 1985. Inventario forestal del estado de Campeche. Publicación Especial Número 56. México. 119 p.
- Smelko, S. and J. Saborowski. 1999. Evaluation of variable size sampling plots for monitoring of forest condition. *Journal of Forest Science* 45(8): 341-347.
- Smelko, S. and J. Merganic. 2008. Some methodological aspects of the National Forest Inventory and Monitoring in Slovakia. *Journal of Forest Science* 54 (10):476-483.
- Thompson, S.K. 2002. Sampling. John Wiley & Sons. New York, USA. 367 p.
- Valentine, H. T., M. J. Ducey, J. H. Gove, A. Lanz and D. L.R. Affleck. 2006. Corrections for cluster-plot slop. *Forest Science* 52(1):55-66.
- Velasco B., E., H. Ramírez M., F. Moreno S. y A. de la Rosa V. 2005. Estimadores de razón para el inventario nacional forestal de México. *Revista Ciencia Forestal en México*. 28(94):23-43.
- Zarnoch, S. J. and W. A. Bechtold. 2000. Estimating mapped-plot forest attributes with ratios of means. *Canadian Journal of Forest Research*. 30:688-697.

ANEXO 2.1. ANÁLISIS DEL SESGO DEL ESTIMADOR DE RAZÓN EN MUESTREO ALEATORIO SIMPLE (adaptación de Pérez, 2000)

Sean (y_s, a_s) los diferentes pares de valores relativos a la variable de interés (área basal, densidad, volumen) y a la variable auxiliar (área), respectivamente. Se denomina la razón poblacional Y_{ha} de y y a al cociente entre las medias o totales poblaciones de y y a . Se tiene

$$Y_{ha} = \frac{\sum_{s=1}^M y_s}{\sum_{s=1}^M a_s} = \frac{Y}{A} = \frac{Y/M}{A/M} = \frac{\bar{Y}}{\bar{A}} = \frac{\sum_{s=1}^M y_s / M}{\sum_{s=1}^M a_s / M}$$

Se denomina la razón muestral y_{ha} de y y a al cociente entre las medias o totales muestrales de y y a . Se tiene

$$y_{ha} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = \frac{y}{a} = \frac{y/m}{a/m} = \frac{\bar{y}}{\bar{a}} = \frac{\sum_{s=1}^m y_s / m}{\sum_{s=1}^m a_s / m}$$

En la estimación por razón se trata de estimar la razón poblacional a través de la razón muestral, pero en general y_{ha} no es un estimador insesgado de Y_{ha} , ya que **la esperanza de un cociente de variables aleatorias no es igual al cociente de las esperanzas**:

$$E(y_{ha}) = E\left(\frac{\bar{y}}{\bar{a}}\right) \neq \frac{E(\bar{y})}{E(\bar{a})} = \frac{\bar{Y}}{\bar{A}} = Y_{ha}$$

No obstante, existen ciertas circunstancias donde el estimador de razón puede ser insesgado. El **valor exacto del sesgo** se puede hallar considerando la covarianza de y_{ha} y \bar{a} .

$$\begin{aligned} \text{cov}(y_{ha}, \bar{a}) &= E\{(y_{ha} - E[y_{ha}])(\bar{a} - E[\bar{a}])\} = E\{(y_{ha}\bar{a} - y_{ha}E[\bar{a}] - E[y_{ha}]\bar{a} + E[y_{ha}]E[\bar{a}])\} \\ &= E(y_{ha}\bar{a}) - E(y_{ha})E(\bar{a}) - E(y_{ha})E(\bar{a}) + E(y_{ha})E(\bar{a}) = E(y_{ha}\bar{a}) - E(y_{ha})E(\bar{a}) \\ &= E\left(\frac{\bar{y}}{\bar{a}}\bar{a}\right) - E(y_{ha})\bar{A} = E(\bar{y}) - \bar{A}E(y_{ha}) = \bar{Y} - \bar{A}E(y_{ha}) = \frac{\bar{A}}{A}\bar{Y} - \bar{A}E(y_{ha}) \\ &= \bar{A}Y_{ha} - \bar{A}E(y_{ha}) = \bar{A}(Y_{ha} - E(y_{ha})) = -\bar{A}B(y_{ha}) \end{aligned}$$

En la última expresión se consideró que $B(y_{ha}) = E(y_{ha}) - Y_{ha}$, y en consecuencia, $-B(y_{ha}) = -E(y_{ha}) + Y_{ha}$

Por lo tanto,

$$\begin{aligned} \text{cov}(y_{ha}, \bar{a}) &= -\bar{A}B(y_{ha}) \Rightarrow \bar{A}B(y_{ha}) = -\text{cov}(y_{ha}, \bar{a}) \Rightarrow B(y_{ha}) = \frac{-\text{cov}(y_{ha}, \bar{a})}{\bar{A}} \\ &= \frac{-\rho_{(y_{ha}, \bar{a})} \sigma_{y_{ha}} \sigma_{\bar{a}}}{E(\bar{a})} = -\rho_{(y_{ha}, \bar{a})} \sigma_{y_{ha}} \frac{\sigma_{\bar{a}}}{E(\bar{a})} = -\rho_{(y_{ha}, \bar{a})} \sigma_{y_{ha}} Cv(\bar{a}) \end{aligned}$$

Así,

$$B(y_{ha}) = -\rho_{(y_{ha}, \bar{a})} \sigma_{y_{ha}} Cv(\bar{a})$$

Por lo tanto, $B(y_{ha}) = 0, \Leftrightarrow, y_{ha}$ y \bar{a} son variables no correlacionadas, con lo que se tiene la primera condición de insesgidez. Además se cumple que

$$\begin{aligned} B(y_{ha}) &= -\rho_{(y_{ha}, \bar{a})} \sigma_{y_{ha}} Cv(\bar{a}) \Rightarrow \frac{B(y_{ha})}{\sigma_{y_{ha}}} = -\rho_{(y_{ha}, \bar{a})} Cv(\bar{a}) \\ \Rightarrow \left| \frac{B(y_{ha})}{\sigma_{y_{ha}}} \right| &= \left| \rho_{(y_{ha}, \bar{a})} \right| Cv(\bar{a}) \leq Cv(\bar{a}) \end{aligned}$$

Por lo que el sesgo relativo (cociente entre el sesgo del estimador de la razón y su desviación típica) está acotado por el coeficiente de variación de \bar{a} . Para que el sesgo del estimador de razón sea despreciable bastará con que el coeficiente de variación de la media muestral de la variable auxiliar sea menor que 1/10.

Una segunda condición para la insesgidez se refiere a la relación lineal entre la variable de interés y la variable auxiliar. “Si la recta de regresión de la variable de interés sobre la variable auxiliar pasa por el origen entonces el estimador de razón es insesgado”. En este caso se cumple que $y_s = ka_s$.

$$Y_{ha} = \frac{Y}{A} = \frac{\sum_{s=1}^M y_s}{\sum_{s=1}^M a_s} = \frac{\sum_{s=1}^M ka_s}{\sum_{s=1}^M a_s} = \frac{k \sum_{s=1}^M a_s}{\sum_{s=1}^M a_s} = k$$

$$y_{ha} = \frac{y}{a} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = \frac{\sum_{s=1}^m k a_s}{\sum_{s=1}^m a_s} = \frac{k \sum_{s=1}^m a_s}{\sum_{s=1}^m a_s} = k$$

Por lo que,

$$B(y_{ha}) = E(y_{ha}) - Y_{ha} = E(y_{ha} - Y_{ha}) = E(k - k) = 0$$

Mediante el método general de linealización se puede obtener el **sesgo aproximado** del estimador de razón, para lo cual se considera en principio la diferencia entre el estimador y_{ha} y la razón poblacional Y_{ha} .

$$y_{ha} - Y_{ha} = \frac{\bar{y}}{\bar{a}} - Y_{ha} = \frac{\bar{y} - Y_{ha}\bar{a}}{\bar{a}} = \frac{\bar{y} - Y_{ha}\bar{a}}{\bar{A}} \frac{\bar{A}}{\bar{a}} = \frac{\bar{y} - Y_{ha}\bar{a}}{\bar{A}} \frac{\bar{A}}{\bar{A} + \bar{a} - \bar{A}} = \frac{\bar{y} - Y_{ha}\bar{a}}{\bar{A}} \frac{1}{1 + \frac{\bar{a} - \bar{A}}{\bar{A}}}$$

Note que $r = \left| \frac{\bar{a} - \bar{A}}{\bar{A}} \right| < 1$ para $\bar{A} \neq 0$. Además, se sabe que si

$$|r| < 1 \Rightarrow \sum_{n=0}^{\infty} (-1)^n (r)^n = \sum_{n=0}^{\infty} (-r)^n = \frac{1}{1 - (-r)} = \frac{1}{1 + r}. \text{ Entonces,}$$

$$\begin{aligned} \frac{1}{1 + \frac{\bar{a} - \bar{A}}{\bar{A}}} &= \sum_{n=0}^{\infty} \left[-\left(\frac{\bar{a} - \bar{A}}{\bar{A}} \right) \right]^n = \left[-\left(\frac{\bar{a} - \bar{A}}{\bar{A}} \right) \right]^0 + \left[-\left(\frac{\bar{a} - \bar{A}}{\bar{A}} \right) \right]^1 + \left[-\left(\frac{\bar{a} - \bar{A}}{\bar{A}} \right) \right]^2 + \dots \\ &= 1 - \left(\frac{\bar{a} - \bar{A}}{\bar{A}} \right) + \left(\frac{\bar{a} - \bar{A}}{\bar{A}} \right)^2 - \dots = 1 - \left(\frac{\bar{a} - \bar{A}}{\bar{A}} \right) \end{aligned}$$

Por tanto,

$$y_{ha} - Y_{ha} = \frac{\bar{y} - Y_{ha}\bar{a}}{\bar{A}} \left[1 - \left(\frac{\bar{a} - \bar{A}}{\bar{A}} \right) \right] = \left[\frac{\bar{y} - Y_{ha}\bar{a}}{\bar{A}} - \frac{(\bar{y} - Y_{ha}\bar{a})(\bar{a} - \bar{A})}{\bar{A}^2} \right]$$

La expresión asintótica del sesgo se obtiene tomando la esperanza a la expresión anterior:

$$\begin{aligned} B(y_{ha}) &= E(y_{ha}) - Y_{ha} = E(y_{ha} - Y_{ha}) = E \left[\frac{\bar{y} - Y_{ha}\bar{a}}{\bar{A}} - \frac{(\bar{y} - Y_{ha}\bar{a})(\bar{a} - \bar{A})}{\bar{A}^2} \right] \\ &= E \left[\frac{\bar{y} - Y_{ha}\bar{a}}{\bar{A}} \right] - E \left[\frac{(\bar{y} - Y_{ha}\bar{a})(\bar{a} - \bar{A})}{\bar{A}^2} \right] = \frac{E(\bar{y} - Y_{ha}\bar{a})}{\bar{A}} - \frac{E[(\bar{y} - Y_{ha}\bar{a})(\bar{a} - \bar{A})]}{\bar{A}^2} \end{aligned}$$

$$= -\left(\frac{\text{cov}(\bar{y}, \bar{a}) - Y_{ha} \text{var}(\bar{a})}{\bar{A}^2}\right)$$

En la última expresión se consideró lo siguiente:

$$E(\bar{y} - Y_{ha} \bar{a}) = E(\bar{y}) - Y_{ha} E(\bar{a}) = \bar{Y} - Y_{ha} \bar{A} = \bar{Y} - \frac{\bar{Y}}{\bar{A}} \bar{A} = \bar{Y} - \bar{Y} = 0$$

Y

$$\begin{aligned} E[(\bar{y} - Y_{ha} \bar{a})(\bar{a} - \bar{A})] &= E[\bar{y}\bar{a} - \bar{y}\bar{A} - Y_{ha} \bar{a}\bar{a} + Y_{ha} \bar{a}\bar{A}] \\ &= E(\bar{y}\bar{a}) - E(\bar{y}\bar{A}) - E(Y_{ha} \bar{a}\bar{a}) + E(Y_{ha} \bar{a}\bar{A}) \\ &= E(\bar{y}\bar{a}) - \bar{A}E(\bar{y}) - Y_{ha} E(\bar{a}^2) + Y_{ha} \bar{A}E(\bar{a}) \\ &= E(\bar{y}\bar{a}) - E(\bar{a})E(\bar{y}) - Y_{ha} E(\bar{a}^2) + Y_{ha} E(\bar{a})E(\bar{a}) \\ &= E(\bar{y}\bar{a}) - E(\bar{a})E(\bar{y}) - Y_{ha} E(\bar{a}^2) + Y_{ha} [E(\bar{a})]^2 \\ &= [E(\bar{y}\bar{a}) - E(\bar{a})E(\bar{y})] - Y_{ha} [E(\bar{a}^2) - [E(\bar{a})]^2] \\ &= \text{cov}(\bar{y}, \bar{a}) - Y_{ha} \text{var}(\bar{a}) \end{aligned}$$

En esta última expresión se consideró que $\text{var}(\hat{\theta}) = E(\hat{\theta}^2) - [E(\hat{\theta})]^2$.

Por tanto, la expresión asintótica del sesgo es

$$B(y_{ha}) = \frac{Y_{ha} \text{var}(\bar{a}) - \text{cov}(\bar{y}, \bar{a})}{\bar{A}^2} = \frac{Y_{ha} \frac{S_a^2}{m} - \frac{S_{y,a}}{m}}{\bar{A}^2} = \frac{1}{m\bar{A}^2} (Y_{ha} S_a^2 - S_{y,a})$$

El cual puede estimarse insesgadamente por

$$\hat{B}(y_{ha}) = \frac{1}{m\bar{a}^2} (y_{ha} S_a^2 - s_{y,a})$$

Donde

$$\begin{aligned} S_a^2 &= \frac{\sum_{s=1}^m (a_s - \bar{a})^2}{m-1} = \frac{\sum_{s=1}^m (a_s^2 - 2a_s \bar{a} + \bar{a}^2)}{m-1} = \frac{\sum_{s=1}^m a_s^2 - 2\bar{a} \sum_{s=1}^m a_s + m\bar{a}^2}{m-1} \\ S_{y,a} &= \frac{\sum_{s=1}^m (y_s - \bar{y})(a_s - \bar{a})}{m-1} = \frac{\sum_{s=1}^m (y_s a_s - y_s \bar{a} - \bar{y} a_s + \bar{y} \bar{a})}{m-1} = \frac{\sum_{s=1}^m y_s a_s - \bar{a} \sum_{s=1}^m y_s - \bar{y} \sum_{s=1}^m a_s + m\bar{y}\bar{a}}{m-1} \end{aligned}$$

Por lo tanto sí $\hat{B}(y_{ha}) / \hat{\sigma}_{y_{ha}} \leq .010$ el sesgo será despreciable.

**ANEXO 2.2. LISTA DE ESPECIES REPRESENTATIVAS DE SELVA MEDIANA
SUBPERENNIFOLIA Y SUBCADUCIFOLIA**

Nombre común	Nombre científico
ya' (zapote, chicozapote)	<i>Manilkara zapota</i>
ja'asche'	<i>Alseis yucatanensis</i>
caoba	<i>Swietenia macrophylla</i>
kuyché (zapote bobo)	<i>Pachira aquatica</i>
Ox (Ramón)	<i>Brosimum alicastrum</i>
Pich (orejón)	<i>Enterolobium cyclocarpum</i>
Jabín	<i>Piscidia piscipula</i>
Ki	<i>Agave sp</i>
Yaaxnik	<i>Vitex gaumeri</i>
Bari	<i>Calophyllum brasiliense</i>
Cascarillo	<i>Matayba oppositifolia</i>
Laurelillo	<i>Nectandra coriácea</i>
Pukte'	<i>Bucidas buceras</i>
Boxchechem	<i>Metopium brownei</i>
Sakchechem	<i>Cameraria latifolia</i>
Nance	<i>Byrsonima crassifolia</i>
Pomolche'	<i>Jatropha gaumeri</i>
Caimito	<i>Chrysophyllum mexicanum</i>
Jícaro	<i>Crescentia cujete</i>
Bob	<i>Coccoloba acapulcensis</i>
Pichi'che'	<i>Psidium sartorianum</i>
Ciricote	<i>Cordia dodecandra</i>
Tsitsilche'	<i>Gymnopodium floribundum</i>
Katzin	<i>Mimosa bahamensis</i>
Chakah blanco	<i>Dendropanax arboreus</i>
Muk	<i>Dalbergia glabra</i>
Sakpa'	<i>Byrsonima bucidaefolia</i>
Hulub	<i>Bravaisia tubiflora</i>
Ta-Siste	<i>Acoelorrhaphe wrightii</i>
Tzalam	<i>Lysiloma bahamensis</i>
<i>Chechem blanco</i>	<i>Sebastiania adenophora</i>
Jucarillo	<i>Bucida spinosa</i>
Cordoncillo	<i>Piper auritum</i>
Xku	<i>Aechmea bracteata</i>

CAPÍTULO III

COMPARACIÓN DE ESTIMADORES PARA PARÁMETROS FORESTALES EN UNA POBLACIÓN SIMULADA

3.1 RESUMEN

El diseño de muestreo del Inventario Nacional Forestal y de Suelos (INFiS) de México está integrado por conglomerados de una hectárea donde, de manera ideal se miden siempre cuatro subparcelas. No obstante, son muchas las razones que en la realidad impiden que éste sea siempre el caso, por lo que es frecuente tener información de conglomerados con menos de cuatro subparcelas. Por lo que resulta importante estudiar desde un punto de vista teórico el comportamiento de los diferentes estimadores que se pueden ser de utilidad para la estimación de parámetros forestales agregados cuando se tienen conglomerados de diferente tamaño. En base a la densidad arbórea estimada (árboles ha^{-1}) y a la distribución probabilística de volúmenes (lognormal) ajustada con datos del 2010 de selva mediana de Quintana Roo, se generó una población artificial de 9,370,000 árboles en una superficie de 10,000 ha en Quintana Roo, México. Se diseñaron dos redes sistemáticas, cada una de 81 puntos, los cuales estuvieron a equidistancias de 1 km; en cada punto se simuló un conglomerado de subparcelas dispuestas en forma de una Y invertida. Para comparar diseños de parcelas en selva, en un primer escenario se consideraron subparcelas rectangulares, y en un segundo, subparcelas circulares. Para densidad y volumen, una prueba de t pareada bivariada indicó que no existen diferencias significativas entre los valores generados por las subparcelas circulares y rectangulares. De manera adicional se generaron 86 redes sistemáticas considerando únicamente conglomerados de subparcelas circulares. Cada una de las 88 muestras se trató en tres condiciones diferentes en función del número de conglomerados completos: a) completitud total (81 conglomerados completos), b) completitud al 88% (71 conglomerados completos) y c) completitud al 63% (51 conglomerados completos). En cada condición y repetición se estimó el número de árboles y volumen maderable por hectárea mediante los estimadores media de razones (MR), razón de medias (RM), Forest Inventory and Analysis (FIA) y Van Deusen (VAN). Para determinar el mejor estimador en cada condición estudiada se usó el sesgo relativo de la estimación y el sesgo relativo de la varianza. En las tres condiciones estudiadas y para las dos variables de interés, las estimaciones de MR, RM, FIA y VAN resultaron similares y prácticamente insesgadas. En conglomerados completos al 100% la estimación de la varianza fue insesgada independientemente del estimador. Sin embargo, a medida que la muestra tuvo mayor número de conglomerados incompletos, la varianza de FIA se volvió fuertemente sesgada.

Palabras clave: razón de medias, media de razones, estimador Van Deusen, simulación, distribución espacial de árboles, distribución lognormal.

ESTIMATORS COMPARISON FOR FOREST PARAMETERS IN A SIMULATED POPULATION

3.2. SUMMARY

The sampling unit of the National Forest and Soils Inventory (INFyS) of Mexico consists of one hectare clusters that ideally always contains four measured subplots. However, there are many reasons in reality that prevent that this is always the case, so it is common to have information of clusters with fewer than four subplots. Therefore is important to study from a theoretical point of view the behavior of the different estimators that can be useful for the estimation of aggregate forest parameters when the clusters are by definition of different size (unequal number of subplots). Based on the estimated trees per hectare and the tree volume an artificial population of 9,370,000 trees in an area of 10,000 ha was created. The distribution used to simulate the population follow a log-normal distribution (for volume) and the data used as reference came from the INFyS 2010 data of medium rainforest in Quintana Roo, Mexico. Two systematical sampling grids were designed, each of 81 points, which were to 1 km equidistances; in every point was simulated a cluster of subplots arranged in the shape of an inverted Y. To compare sub-plots design used in the tropical rainforest all the subplots were considered first to be rectangular, then to be circular. For density and volume, a bivariate paired t-test indicated that there are not significant differences among the values generated by the circular and rectangular subplots. Additionally, 86 systematic grids were generated considering only clusters of circular subplots. Each of the 88 samples were treated in three different conditions depending on the number of complete clusters: (a) total completeness (81 complete clusters), (b) completeness to 88% (71 complete clusters) and c) completeness to 63% (51 complete clusters). In each condition the number of trees and timber volume per hectare was estimated by Mean of Ratios (MR), Ration of Means (RM), Forest Inventory and Analysis (FIA) and Van Deusen (VAN) estimators. To determine the best estimator in each condition we estimate the relative mean bias and relative variance bias. In the three conditions and for the two variables of interest, estimates MR, RM, FIA and VAN were similar and practically unbiased. In complete clusters (100%) variance estimates were unbiased independently of the estimator. However, as the sample included larger proportion of incomplete clusters, the variance of FIA became strongly biased.

Key words: ratio of means, mean of ratios, Van Deusen estimator, simulation, tree spatial distribution, lognormal distribution.

3.3. INTRODUCCIÓN

Si bien es cierto que el diseño del conglomerado del INFyS incluye teóricamente cuatro subparcelas, bajo ciertas circunstancias existe la posibilidad que en campo la información se obtenga en conglomerados menores a cuatro subparcelas, generándose así conglomerados de tamaños distintos (conglomerados completos e incompletos). Aun cuando un conglomerado comprende dos o más subparcelas disjuntas es tratado como una sola parcela debido a que es atado a un solo punto muestral, y en consecuencia, la varianza de los estimadores considera únicamente la varianza entre conglomerados.

El diseño del conglomerado del INFyS es parecido al diseño del “plot” que se utiliza en el programa de Inventario Forestal y Análisis (FIA) en los Estados Unidos de Norteamérica (Bechtold y Scott, 2005), por lo que ambos pueden ser estudiados de manera similar.

En estas condiciones resulta importante estudiar desde un punto de vista teórico el comportamiento de los diferentes estimadores que se pueden ser de utilidad para la estimación de parámetros forestales agregados cuando se tienen conglomerados —de subparcelas espacialmente disjuntas —de diferente tamaño. Una opción para realizar esta tarea es simular una población de árboles en base a parámetros ya estimados en estudios previos; el sesgo relativo de la estimación y el sesgo relativo de la varianza más pequeños pueden ser los criterios para definir el mejor estimador (Van Deusen, 2006). Esta alternativa, aunque requiere de gran trabajo computacional, ha sido preferida en el ámbito científico por su flexibilidad en cuanto a las múltiples maneras de llevar a cabo el muestreo una vez que la población ha sido generada; puede estudiarse la eficiencia del muestreo sistemático o aleatorio, y diversos tipos y tamaños de unidades de muestreo.

Así, al considerar que la cantidad de la variable de interés en el conglomerado es proporcional al área del mismo, existen al menos cuatro formas diferentes de estimar los parámetros agregados de interés forestal: razón de medias, medias de razones, Forest Inventory and Analysis y Van Deusen. En el presente trabajo estos estimadores se

derivaron mediante el enfoque de inferencia basada en modelos, además, fueron comparados usando datos de una población forestal artificial generada en el Sureste de México.

3.4. OBJETIVO

El objetivo del presente trabajo fue evaluar el comportamiento del estimador MR, RM, FIA y VAN, en tres condiciones diferentes, relacionadas al grado de completitud del conglomerado, en una población artificial de 9,370,000 árboles en selva mediana subperennifolia de Quintana Roo, al estimar la densidad arbórea y el volumen maderable por hectárea. Los conglomerados fueron similares al diseño de la parcela que se utiliza en el Inventario Nacional Forestal y de Suelos de México.

3.5. REVISIÓN DE LITERATURA

3.5.1. Derivación de estimadores

Mediante el enfoque de inferencia basada en modelos, a partir del modelo general propuesto por Van Deusen (2007), y considerando diferentes supuestos de la varianza del error se obtienen los estimadores FIA, VAN y MR. Se analiza la similitud del estimador RM con VAN. La relación entre los cuatro métodos de estimación también es considerada analíticamente desde el punto de vista del estimador *per se* cómo de sus varianzas estimadas.

3.5.1.1. Notación

μ : Media de la variable y por unidad de área. Ejemplo, volumen en m^3ha^{-1} .

$y_s^p = \frac{y_s}{a_0}$: Medición de la variable y en la parcela s expandida a una hectárea.

y_s : Medición de la variable y en la parcela s

$a_s^p = \frac{a_s}{a_0}$: Proporción del área de la parcela o conglomerado s que pertenece al tipo

forestal de interés, $0 \leq a_s^p \leq 1$.

a_s : Área de la parcela s (ha) efectivamente muestreada en campo.

a_0 : Área de la parcela s ideal para medir la variable y , es decir, 0.16 ha.

m : Número de parcelas que pertenecen al tipo forestal de interés.

$\bar{a}^p = \frac{\sum_{s=1}^m a_s^p}{m}$: Proporción media de área que pertenece al tipo forestal de interés.

Cuando $a_s = a_0 \Rightarrow a_s^p = \frac{a_s}{a_0} = \frac{a_0}{a_0} = 1$, y $\bar{a}^p = \frac{\sum_{s=1}^m a_s^p}{m} = \frac{\sum_{s=1}^m 1}{m} = \frac{m}{m} = 1$

3.5.1.2. Modelo general

Este modelo fue propuesto por Van Deusen (2007) y se expresa como

$$y_s^p = f(a_s^p)\mu + \varepsilon_s \quad (3.1)$$

Donde, $\varepsilon_s \sim N(0, f(a_s^p)\sigma^2)$ y $E(\varepsilon_s \varepsilon_s') = f(a_s^p)\sigma^2 \quad \forall \quad \varepsilon_s = \varepsilon_s' \quad \text{y} \quad E(\varepsilon_s \varepsilon_s') = 0$ de otro modo.

Reescribiendo la ecuación anterior en forma matricial (Van Deusen, 2004), se tiene

$$Y = A\mu + \varepsilon \quad (3.2)$$

Donde

Y : Vector de mediciones expandidas a la parcela, y_s^p .

A : Vector de los valores correspondientes a $f(a_s^p)$.

ε : Vector de términos aleatorios.

La matriz de varianzas y covarianzas del vector aleatorio es $E(\varepsilon \varepsilon') = \Sigma = \sigma^2 f(a_s^p)$.

La estimación insesgada para la media μ obtenida mediante mínimos cuadrados generalizados (Van Deusen, 2004) es

$$\hat{\mu} = [A' \Sigma^{-1} A]^{-1} A' \Sigma^{-1} Y \quad (3.3)$$

y

$$V(\hat{\mu}) = [A' \Sigma^{-1} A]^{-1} \quad (3.4)$$

3.5.1.3. Modelo Forest Inventory and Analysis

$$y_s^p = \bar{a}^{-p} \mu + \varepsilon 1_s \quad (3.5)$$

Donde $E(\varepsilon 1_s) = 0$ y $E(\varepsilon 1_s \varepsilon 1_s') = (\bar{a}^{-p})^2 \sigma^2 \quad \forall \quad \varepsilon 1_s = \varepsilon 1_s'$

Este modelo establece que el valor expandido por hectárea es igual a la media por unidad de área, μ , multiplicada por la proporción promedio de una parcela, \bar{a}^{-p} , que está en el tipo forestal de interés. La varianza del valor expandido por parcela es proporcional a la fracción promedio al cuadrado en el tipo forestal de interés.

Al considerar el modelo (3.5) y las expresiones (3.3) y (3.4), se tienen los siguientes estimadores para la media y la varianza:

$$\hat{\mu}^{FIA} = \hat{\mu} = \frac{\sum_{s=1}^m \tilde{y}_s}{m} \quad (3.6)$$

$$v(\hat{\mu}^{FIA}) = \frac{\sum_{s=1}^m \left(\tilde{y}_s - \hat{\mu} \right)^2}{m(m-1)} = \frac{\sum_{s=1}^m \tilde{y}_s^2 - m \hat{\mu}^2}{m(m-1)} \quad (3.7)$$

Donde, $\tilde{y}_s = \frac{y_s^p}{\bar{a}^{-p}}$

Note que,

$$\tilde{y}_s = \frac{y_s^p}{\bar{a}^{-p}} = \frac{y_s^p}{\sum_{s=1}^m a_s^p} = \frac{\frac{y_s}{a_0}}{\sum_{s=1}^m \frac{a_s}{a_0}} = \frac{\frac{y_s}{a_0}}{\frac{1}{a_0} \sum_{s=1}^m a_s} = \frac{\frac{y_s}{a_0} m}{\sum_{s=1}^m a_s} = \frac{y_s}{\frac{\sum_{s=1}^m a_s}{a_0 m}} = y_{ha}(s)$$

Entonces, las expresiones (3.6) y (3.7) expresadas en las unidades originales de medición son, respectivamente,

$$y_{ha}^{-FIA} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} \quad (3.8)$$

Donde

$$y_{ha}(s) = \frac{y_s}{\frac{\sum_{s=1}^m a_s}{a_0 m}} \quad (3.8a)$$

y

$$v(y_{ha}^{-FIA}) = \frac{\sum_{s=1}^m \left(y_{ha}(s) - y_{ha}^{-FIA} \right)^2}{m(m-1)} \quad (3.9)$$

Las expresiones (3.8) y (3.9) son los estimadores de la media y la varianza de la media que utiliza el FIA (Scott *et al.*, 2005). En (3.8a) $a_0 m = \sum_{s=1}^m a_0$ es el área total deseable a muestrear asumiendo conglomerados completos y $\sum_{s=1}^m a_s$ es el área total efectivamente muestreada. Entonces, $\sum_{s=1}^m a_s / a_0 m \leq 1$. Así, el denominador de (3.8a) es el área del conglomerado completo ajustada por la proporción de las parcelas que pertenecen en el tipo forestal de interés. En estas condiciones (3.8a) es un valor por hectárea. Las expresiones (3.8) y (3.9) son los estimadores estándares de la media y la varianza que podrían aplicarse al muestreo aleatorio simple sin un factor de corrección finita (Cochran, 1993 y Thompson, 2002).

3.5.1.4. Modelo Van Deusen

$$y_s^p = a_s^p \mu + \varepsilon 2_s \quad (3.10)$$

Donde: $E(\varepsilon 2_s) = 0$ y $E(\varepsilon 2_s \varepsilon 2_s') = a_s^p \sigma^2 \quad \forall \quad \varepsilon 2_s = \varepsilon 2_s'$

Es razonable asumir que la $\text{var}(\varepsilon 2_s) = a_s^p \sigma^2$ debido a que la varianza debe ser cero cuando $a_s^p = 0$ y debe incrementarse conforme a_s^p se incremente.

La ecuación (3.10) establece que la cantidad y_s^p en el tipo forestal de interés está relacionada con la proporción a_s^p de la parcela que cae en el tipo forestal de interés. La

varianza de y_s^p es también proporcional a a_s^p , lo cual parece intuitivamente razonable.

Al considerar el modelo (3.10) y las expresiones (3.3) y (3.4), se tienen los siguientes estimadores para la media y la varianza (Van Deusen, 2004):

$$\hat{\mu}^{VAN} = \frac{\sum_{s=1}^m y_s^p}{\sum_{s=1}^m a_s^p} \quad (3.11)$$

y

$$v(\hat{\mu}^{VAN}) = \frac{1}{\sum_{s=1}^m a_s^p} \left(\frac{\sum_{s=1}^m (y_s^p - a_s^p \hat{\mu}^{VAN})^2}{\sum_{s=1}^m a_s^p - 1} \right) \quad (3.12)$$

La ecuación (3.11) puede expresarse de la siguiente manera:

$$\hat{\mu}^{VAN} = \frac{\sum_{s=1}^m y_s^p}{\sum_{s=1}^m a_s^p} = \frac{\sum_{s=1}^m \frac{y_s}{a_0}}{\sum_{s=1}^m \frac{a_s}{a_0}} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = y_{ha} \quad (3.13)$$

Similarmente, la varianza en (3.12)

$$\begin{aligned}
v(\hat{\mu}^{VAN}) &= \frac{1}{\sum_{s=1}^m a_s^p} \left(\frac{\sum_{s=1}^m (y_s^p - a_s^p \hat{\mu})^2}{\sum_{s=1}^m a_s^p - 1} \right) = \frac{1}{\sum_{s=1}^m \frac{a_s}{a_0}} \left(\frac{\sum_{s=1}^m \left(\frac{y_s}{a_0} - \frac{a_s}{a_0} y_{ha} \right)^2}{\sum_{s=1}^m \frac{a_s}{a_0} - 1} \right) \\
&= \frac{1}{\sum_{s=1}^m \frac{a_s}{a_0}} \left(\frac{\sum_{s=1}^m \frac{y_s^2}{a_0^2} - 2y_{ha} \sum_{s=1}^m \frac{a_s y_s}{a_0^2} + y_{ha}^2 \sum_{s=1}^m \frac{a_s^2}{a_0^2}}{\sum_{s=1}^m \frac{a_s}{a_0} - 1} \right) = \frac{1}{\frac{1}{a_0} \sum_{s=1}^m a_s} \left(\frac{\frac{1}{a_0^2} \sum_{s=1}^m y_s^2 - \frac{2y_{ha}}{a_0^2} \sum_{s=1}^m a_s y_s + \frac{y_{ha}^2}{a_0^2} \sum_{s=1}^m a_s^2}{\frac{1}{a_0} (\sum_{s=1}^m a_s - a_0)} \right) \\
&= \frac{\frac{1}{a_0^2} \sum_{s=1}^m a_s}{\frac{1}{a_0^2} \sum_{s=1}^m a_s} \left(\frac{\sum_{s=1}^m y_s^2 - 2y_{ha} \sum_{s=1}^m a_s y_s + y_{ha}^2 \sum_{s=1}^m a_s^2}{(\sum_{s=1}^m a_s - a_0)} \right) = \frac{1}{\sum_{s=1}^m a_s} \left(\frac{\sum_{s=1}^m y_s^2 - 2y_{ha} \sum_{s=1}^m a_s y_s + y_{ha}^2 \sum_{s=1}^m a_s^2}{(\sum_{s=1}^m a_s - a_0)} \right) \\
&= \frac{\sum_{s=1}^m (y_s - a_s y_{ha})^2}{\sum_{s=1}^m a_s (\sum_{s=1}^m a_s - a_0)} = v(y_{ha}^{VAN}) \tag{3.14}
\end{aligned}$$

3.5.1.5. Modelo media de razones

Si en el modelo (3.10) se considera que la varianza del error es una función de la proporción a_s^p al cuadrado se tiene lo siguiente:

$$y_s^p = a_s^p \mu + \varepsilon 3_s \tag{3.15}$$

Donde $E(\varepsilon 3_s) = 0$ y $E(\varepsilon 3_s \varepsilon 3_s') = (a_s^p)^2 \sigma^2 \quad \forall \quad \varepsilon 3_s = \varepsilon 3_s'$

Al considerar el modelo (3.15) y las expresiones (3.3) y (3.4) se obtienen siguientes estimadores para la media y la varianza:

$$\hat{\mu}^{MR} = \frac{1}{m} \sum_{s=1}^m \frac{y_s^p}{a_s^p} \tag{3.16}$$

y

$$v(\hat{\mu}^{MR}) = \frac{\sum_{s=1}^m \left(\dot{y}_s - \hat{\mu} \right)^2}{m(m-1)} = \frac{\sum_{s=1}^m \dot{y}_s^2 - m \hat{\mu}^2}{m(m-1)} \quad (3.17)$$

Donde

$$\dot{y}_s = \frac{y_s^p}{a_s^p}$$

En las unidades originales, las expresiones (3.16) y (3.17) son, respectivamente,

$$\hat{\mu}^{MR} = \frac{1}{m} \sum_{s=1}^m \frac{y_s^p}{a_s^p} = \frac{1}{m} \sum_{s=1}^m \frac{\frac{y_s}{a_0}}{\frac{a_s}{a_0}} = \frac{1}{m} \sum_{s=1}^m \frac{y_s}{a_s} = \bar{y}_{ha} \quad (3.18)$$

y

$$v(\bar{y}_{ha}) = \frac{\sum_{s=1}^m (y_{ha} - \bar{y}_{ha})^2}{m(m-1)} \quad (3.19)$$

ya que, $\dot{y}_s = \frac{y_s^p}{a_s^p} = \frac{y_s / a_0}{a_s / a_0} = \frac{y_s}{a_s} = y_{ha}$

En el ANEXO 3.1 se presenta la derivación de los estimadores FIA, MR y VAN, así como sus respectivas varianzas.

3.5.1.6. Estimador razón de medias

El estimador de razón de medias tiene la misma estructura que el estimador Van Deusen, ecuación (3.13), el cual también puede ser usado cuando se presentan conglomerados incompletos, a saber:

$$y_{ha}^{RM} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} \quad (3.20)$$

La varianza estimada de (3.20), considerado éste como estimador de razón, según Cochran (1993), es

$$v(y_{ha}^{RM}) = \frac{\sum_{s=1}^m (y_s - y_{ha} a_s)^2}{a^2 m(m-1)} \quad (3.21)$$

Donde, $\bar{a} = \frac{\sum_{s=1}^m a_s}{m}$

Este estimador fue recomendado para la estimación de parámetros forestales del Inventario Nacional Forestal de México (Velasco *et al.*, 2005).

El denominador de la expresión (3.21) puede expresarse en términos de las áreas efectivamente muestreadas considerando que $\bar{a} = \sum_{s=1}^m a_s / m \Rightarrow m = \sum_{s=1}^m a_s / \bar{a}$, así

$$\begin{aligned} v(y_{ha}^{RM}) &= \frac{\sum_{s=1}^m (y_s - y_{ha} a_s)^2}{a^2 m(m-1)} = \frac{\sum_{s=1}^m (y_s - y_{ha} a_s)^2}{a^2 \left(\frac{\sum_{s=1}^m a_s}{a} \left(\left(\frac{\sum_{s=1}^m a_s}{a} \right) - 1 \right) \right)} = \frac{\sum_{s=1}^m (y_s - y_{ha} a_s)^2}{\bar{a} \left(\sum_{s=1}^m a_s \right) \left(\left(\frac{\sum_{s=1}^m a_s}{a} \right) - 1 \right)} \\ &= \frac{\sum_{s=1}^m (y_s - y_{ha} a_s)^2}{\sum_{s=1}^m a_s (\sum_{s=1}^m a_s - \bar{a})} \end{aligned} \quad (3.21a)$$

Una inspección detallada de las expresiones (3.21a) y (3.14) —varianza estimada de Van Deusen— permite apreciar que ambas solo difieren en el denominador por el término que se sustrae del área total efectivamente muestreada, mientras que la varianza del estimador de razón considera un promedio, en Van Deusen se resta el área de un conglomerado completo, es decir, 0.16 ha. Si los conglomerados de la muestra son completos, se tiene que $\bar{a} = \sum_{s=1}^m a_0 / m = a_0 m / m = a_0$, y en consecuencia, el estimador

de razón y Van Deusen conducen a los mismos resultados. De hecho, $a_0=0.16$ es el máximo valor que se puede considerar, luego entonces, $\bar{a} \leq a_0$, en estas condiciones la varianza estimada del estimador de razón siempre será ligeramente más pequeña que Van Deusen.

La ecuación (3.21) se puede obtener multiplicando la ecuación (3.14) por un factor que involucra \bar{a} (ANEXO 3.2):

$$v(y_{ha}^{RM}) = \left(\frac{m\bar{a} - a_0}{a(m-1)} \right) v(y_{ha}^{VAN}) \quad (3.22)$$

En la expresión anterior nuevamente se observa que cuando $\bar{a} = a_0$, $v(y_{ha}^{RM}) = v(y_{ha}^{VAN})$. Una situación extrema ocurre cuando $\bar{a} = (1/2)a_0$ y $m = 20$, resultando en la razón de la ecuación (3.21) sobre (3.14) de 0.9473 (Van Deusen, 2005).

3.5.1.7. Relación entre estimadores

Cuando los conglomerados están completos, es decir, $a_s = a_0$, los cuatro estimadores coinciden:

Estimador FIA (ecuación 3.8)

$$\bar{y}_{ha} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} = \frac{a_0 \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s}}{m} = \frac{a_0 \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_0}}{m} = \frac{\sum_{s=1}^m y_s}{m} = \frac{\sum_{s=1}^m y_s}{a_0 m} \quad (3.23)$$

Estimador Van Deusen (Ecuación 3.13)

$$y_{ha} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_0} = \frac{\sum_{s=1}^m y_s}{a_0 m} \quad (3.24)$$

Media de razones (ecuación 3.18)

$$\bar{y}_{ha} = \frac{1}{m} \sum_{s=1}^m \frac{y_s}{a_s} = \frac{1}{m} \sum_{s=1}^m \frac{y_s}{a_0} = \frac{\sum_{s=1}^m y_s}{a_0 m} \quad (3.25)$$

Razón de medias (ecuación 3.20)

$$y_{ha} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_0} = \frac{\sum_{s=1}^m y_s}{a_0 m} \quad (3.26)$$

También, cuando $a_s = a_0$ los cuatro estimadores presentan las mismas varianzas estimadas:

Estimador FIA (ecuación 3.9)

$$\begin{aligned} v(\bar{y}_{ha}) &= \frac{\sum_{s=1}^m (y_{ha}(s) - \bar{y}_{ha})^2}{m(m-1)} = \frac{\sum_{s=1}^m \left(\frac{y_s}{a_0} - \frac{\sum_{s=1}^m y_s}{a_0 m} \right)^2}{m(m-1)} = \frac{\sum_{s=1}^m \left(\left(\frac{1}{a_0} \right) \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right) \right)^2}{m(m-1)} \\ &= \frac{\left(\frac{1}{a_0} \right)^2 \sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{m(m-1)} = \frac{\sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{a_0^2 m(m-1)} \end{aligned} \quad (3.27)$$

Estimador Van Deusen (ecuación 3.14)

$$\begin{aligned}
 v(y_{ha}) &= \frac{\sum_{s=1}^m (y_s - a_s y_{ha})^2}{\sum_{s=1}^m a_s \left(\sum_{s=1}^m a_s - a_0 \right)} = \frac{\sum_{s=1}^m \left(y_s - a_0 \frac{\sum_{s=1}^m y_s}{a_0 m} \right)^2}{\sum_{s=1}^m a_0 \left(\sum_{s=1}^m a_0 - a_0 \right)} = \frac{\sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{a_0 m (a_0 m - a_0)} = \frac{\sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{a_0 m a_0 (m-1)} \\
 &= \frac{\sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{a_0^2 m (m-1)} \tag{3.28}
 \end{aligned}$$

Estimador media de razones (ecuación 3.19)

$$\begin{aligned}
 v(\bar{y}_{ha}) &= \frac{\sum_{s=1}^m (y_{ha} - \bar{y}_{ha})^2}{m(m-1)} = \frac{\sum_{s=1}^m \left(\frac{y_s}{a_s} - \frac{\sum_{s=1}^m y_s}{a_0 m} \right)^2}{m(m-1)} = \frac{\sum_{s=1}^m \left(\frac{y_s}{a_0} - \frac{\sum_{s=1}^m y_s}{a_0 m} \right)^2}{m(m-1)} = \frac{\sum_{s=1}^m \left(\frac{1}{a_0} \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right) \right)^2}{m(m-1)} \\
 &= \frac{\left(\frac{1}{a_0} \right)^2 \sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{m(m-1)} = \frac{\sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{a_0^2 m (m-1)} \tag{3.29}
 \end{aligned}$$

Estimador razón de medias (ecuación 3.21)

$$\begin{aligned}
 v(y_{ha}) &= \frac{\sum_{s=1}^m (y_s - y_{ha} a_s)^2}{a^2 m (m-1)} = \frac{\sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{a_0 m} a_0 \right)^2}{\left(\frac{\sum_{s=1}^m a_0}{m} \right)^2 m (m-1)} = \frac{\sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{\frac{a_0^2 m^2}{m^2} m (m-1)} = \frac{\sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{a_0^2 m (m-1)} \tag{3.30}
 \end{aligned}$$

Con base a las ecuaciones (3.23) a (3.26) queda de manifiesto que la estimación por hectárea cuando los conglomerados son completos se obtiene promediando valores por conglomerado y el resultado se expande por hectárea mediante $1/a_0$. El factor de expansión, al ser una constante, también aparece en la varianza, expresiones (3.27) a (3.30). Esta situación es de particular importancia porque las fórmulas se reducen a los estimadores del muestreo aleatorio simple. Los resultados anteriores, media y varianza de la media cuando los conglomerados son completos, también se pueden obtener al considerar el siguiente modelo:

$$y_s^p = \mu + \varepsilon 3_s, E(\varepsilon 3_s) = 0, E(\varepsilon 3_s, \varepsilon 3'_s) = \sigma^2 \quad \forall \quad \varepsilon 3_s = \varepsilon 3'_s$$

Esta situación se describe en el ANEXO 3.1.

Los estimadores FIA, Van Deusen y RM son iguales, incluso si los conglomerados son incompletos:

Considere que $\bar{a} = \frac{\sum_{s=1}^m a_s}{m} \Rightarrow \bar{a}m = \sum_{s=1}^m a_s$

Estimador FIA (ecuación 3.8)

$$\bar{y}_{ha} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} = \frac{a_0 \frac{\sum_{s=1}^m y_s}{a_0 m}}{m} = \frac{\sum_{s=1}^m \frac{y_s}{am}}{m} = \frac{\sum_{s=1}^m \frac{y_s}{a}}{m} = \frac{\sum_{s=1}^m y_s}{am} \quad (3.31)$$

Estimador Van Deusen (Ecuación 3.13)

$$y_{ha} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = \frac{\sum_{s=1}^m y_s}{am} \quad (3.32)$$

Razón de medias (ecuación 3.20)

$$y_{ha} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = \frac{\sum_{s=1}^m y_s}{am} \quad (3.33)$$

3.6. MATERIALES Y MÉTODOS

Una forma práctica de comparar los estimadores es utilizar directamente datos existentes de parcelas de inventario, y mediante los intervalos de confianza o error de muestreo estimado definir el mejor de ellos, en términos de menor varianza o mayor precisión (1/varianza), suponiendo insesgadez. Una segunda opción es mediante una enumeración completa de los árboles que integran una masa forestal, realizar el muestreo y evaluar el comportamiento de las estimaciones respecto al valor verdadero. Una tercera opción es simular una población de árboles en base a parámetros ya estimados en estudios previos (información *a priori*); el sesgo relativo de la estimación y el sesgo relativo de la varianza más pequeños pueden ser los criterios para definir el mejor estimador (Van Deusen, 2006).

La última alternativa, aunque requiere de gran trabajo computacional, ha sido preferida en el ámbito científico por su flexibilidad en cuanto a las múltiples maneras de llevar a cabo el muestreo una vez que la población ha sido generada; puede estudiarse la eficiencia del muestreo sistemático o aleatorio, y diversos tipos y tamaños de unidades de muestreo. Los Sistemas de Información Geográfica y de Análisis Estadístico son las principales herramientas para simular poblaciones en superficies relativamente grandes.

En el Capítulo II del proyecto de investigación se compararon estimadores utilizando directamente los datos de las parcelas de inventario, por lo que en el presente trabajo se utilizó la tercera opción, es decir, se generó una población hipotética de árboles cuyo tamaño fue de $N = 9,370,000$ en la parte central del estado de Quintana, México.

Asumiendo que los árboles en una selva mediana subperennifolia presentan un patrón espacial aleatorio, y por lo tanto, la misma probabilidad de encontrarse en cualquier lugar del terreno, en el Sistema de Análisis Estadístico (SAS v9) se generó

una población aleatoria de 9,370,000 puntos. Los números en X se obtuvieron mediante el código $\text{Int}((\text{MaxValue}-\text{MinValue}+1)*\text{ranuni})+\text{MinValue}$, en donde, MaxValue y MinValue corresponden a los valores extremos en X (en metros) de la región de estudio, y ranuni , es para generar datos con distribución uniforme. Un código similar se aplicó para generar los valores en Y. De esta manera, la posición geográfica de los puntos se definió mediante las coordenadas (X,Y). En la generación de los puntos aleatorios (9,370,000), los valores extremos de X y Y se definieron de tal manera que el área de estudio tuviera una superficie de 10,000 hectáreas, y en consecuencia, el número promedio de árboles por hectárea fuera 937.

El volumen de cada árbol se generó en SAS suponiendo una distribución probabilística lognormal mediante la expresión $\text{VOLUMEN} = \text{EXP}(-2.67507 + 0.906635 * \text{RANNOR}(100))$, en donde, 2.67507 y 0.906635, son respectivamente, el parámetro de escala y de forma de una distribución lognormal ajustada con datos reales en la región de estudio. Tanto el cálculo del número de árboles por hectárea y el ajuste de la distribución lognormal, se llevó a cabo con datos obtenidos en selva mediana subperennifolia en Quintana Roo en 2010 por parte del INFyS. Se utilizó un tamaño de muestra de 206 conglomerados, de los cuales 25 estuvieron incompletos: cinco presentaron un sitio, siete tenían dos sitios y 13 contaron con tres sitios.

La metodología utilizada en esta investigación es similar a la empleada en otros trabajos. Condes y Martínez (1998) simularon parcelas al considerar como base los sitios del Segundo Inventario Forestal Nacional de España en lo que se refiere al número de árboles por hectárea y a la distribución diamétrica de los árboles que la componen. Stamatellos y Panourgias (2005) utilizaron una muestra aleatoria de 229 parcelas de 0.05 ha (20 x 25 m) para simular la distribución espacial de *Abies borisii regis* en la Universidad Forestal de Pertouli, Trikala, Grecia, mediante distribuciones discretas de probabilidad. La distribución Neyman Tipo A, una distribución típica de poblaciones agregadas, tuvo un buen ajuste a los datos, mientras que la simulación de las localizaciones de árboles fue basada en procesos de conglomerados Poisson.

Por lo anterior, los valores verdaderos, es decir, los parámetros, de la población generada de árboles fueron los siguientes: $T_y = \sum_{k=1}^N y_k = 9,370,000$, número total de árboles; $\lambda_y = T_y / A = 937$, número de árboles por hectárea; $T_y = \sum_{k=1}^N y_k = 973,745.53$, volumen total y $\lambda_y = T_y / A = 97.37$, volumen por hectárea.

En primer lugar se diseñaron en SAS dos redes sistemáticas, cada una de 81 puntos, los cuales estuvieron a equidistancias de 1 km; cada punto representó el centro (sitio 1) de un conglomerado de subparcelas. Este tamaño de muestra justifica el uso de la distribución normal en los intervalos de confianza para la media. Mediante las funciones trigonométricas seno y coseno se generaron los centros de las unidades de muestreo periféricas del conglomerado (sitio 2, 3 y 4). En cada punto de la malla de muestreo se generaron de manera independiente conglomerados con subparcelas circulares y rectangulares.

Las coordenadas de puntos aleatorios y de centros correspondientes a los cuatro sitios de conglomerados se capturaron en el Sistema de Información Geográfica ArcMap (v9.3). En la Figura 3.1 se presenta la distribución espacial de los puntos aleatorios que integran la población hipotética de los 9,370,000 árboles. Las subparcelas circulares se generaron través de la herramienta “buffer” de Arc Toolbox-ArcMap, para tal efecto que consideró una distancia de 11.28 m. Las subparcelas rectangulares se trazaron directamente como polígonos; se tomó en cuenta un ancho de 10 m y un largo de 40 m. Los conglomerados trazados de esta manera fueron similares al diseño de la unidad de muestreo que se utiliza en el INFyS (Figura 3.2 y 3.3).

El muestreo de los puntos (árboles) utilizando conglomerados de subparcelas, circulares o rectangulares dispuestas espacialmente en forma de una Y invertida, se realizó mediante la herramienta “Intersect” de Arc Toolbox-ArcMap (Figura 3.4 y 3.5). Esta forma de proceder permitió conservar los identificadores de los conglomerados y de los sitios.



Figura 3.1. Distribución aleatoria de árboles (puntos) en una población hipotética en Quintana, Roo, México, en una superficie de 10,000 hectáreas. Los valores extremos de X son 3,894,331.69 m y 3,904,331.69 m y de Y son 889,927.12 m y 899,927.12 m (Proyección Conforme Cónica de Lambert, Datum ITRF92, Elipsoide GRS80).

De los 81 conglomerados completos se modificaron diez de ellos con el fin de simular conglomerados incompletos; esto se realizó de manera proporcional al número de conglomerados con menos de cuatro sitios que presentó la muestra de 206 conglomerados, misma que sirvió de base para simular la población de 9,370,000 puntos distribuidos aleatoriamente en el terreno. Tomando como base esta condición y triplicando el número de conglomerados con uno, dos y tres subparcelas, adicionalmente se generaron 30 conglomerados incompletos. Entonces, el estudio se llevó a cabo en las siguientes condiciones: completitud total (c0) incluye 81 conglomerados completos, completitud al 88 % (c1) comprende 71 conglomerados completos y diez incompletos (dos con una subparcela, tres con dos subparcelas y cinco con tres subparcelas) y completitud al 63% (c2) considera 51 conglomerados completos y 30 incompletos (seis con una subparcela, nueve con dos subparcelas y 15 con tres subparcelas).

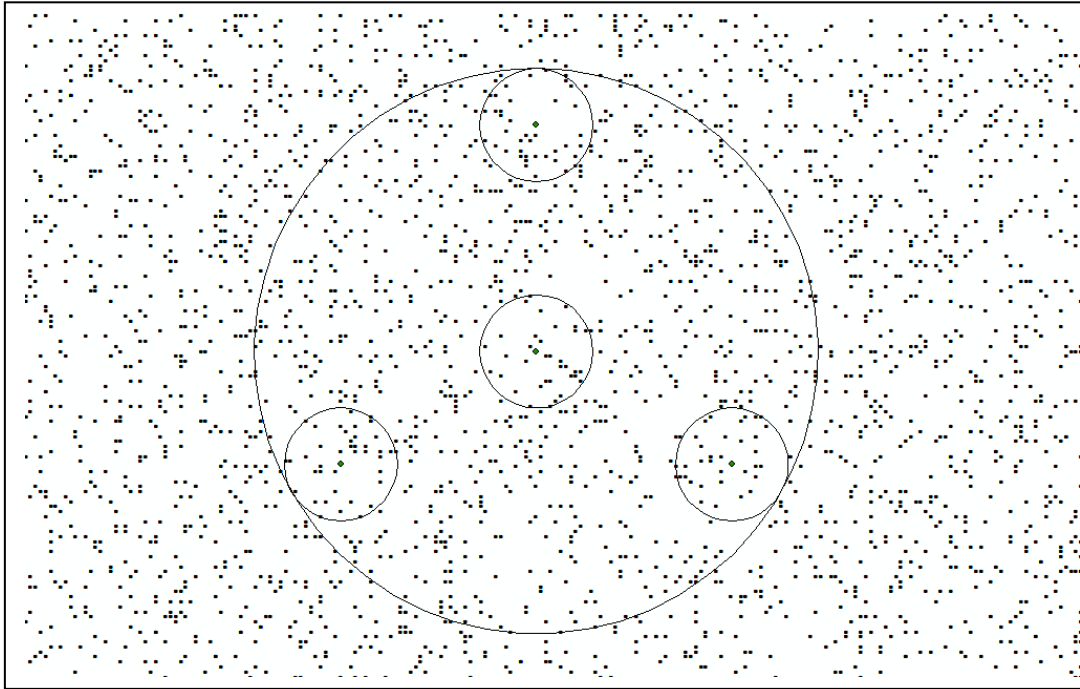


Figura 3.2. Distribución aleatoria de árboles (puntos) en una población hipotética en Quintana, Roo, México y conglomerado (círculo grande) con subparcelas circulares (círculos chicos).

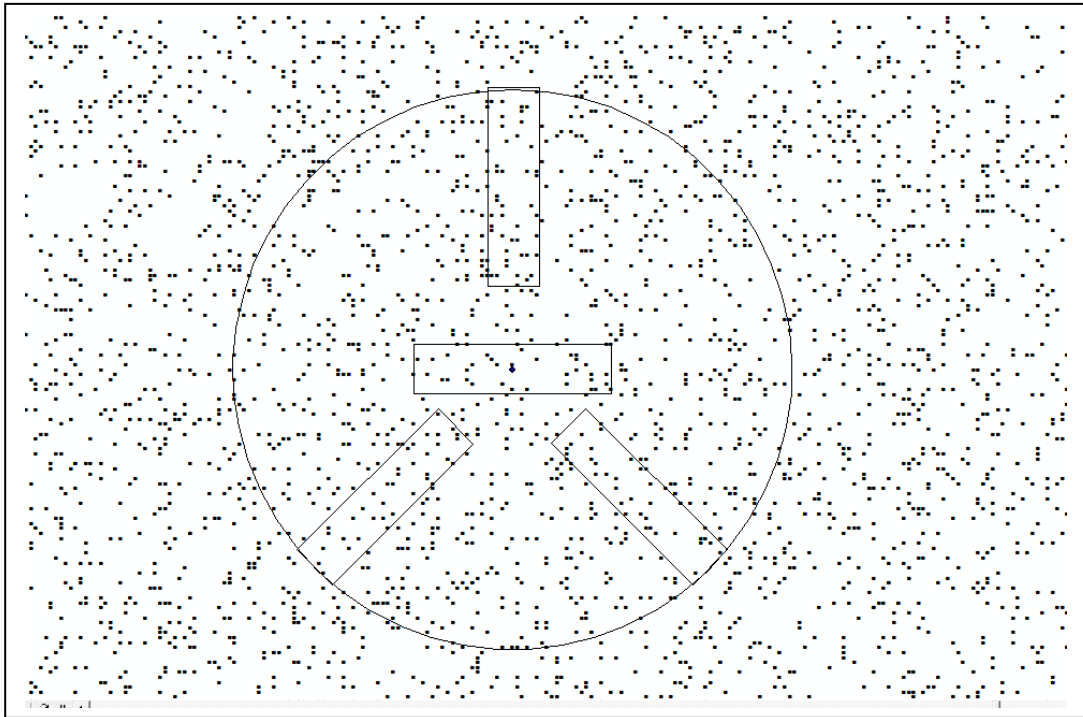


Figura 3.3. Distribución aleatoria de árboles (puntos) en una población hipotética en Quintana, Roo, México y conglomerado (círculo grande) con subparcelas rectangulares.

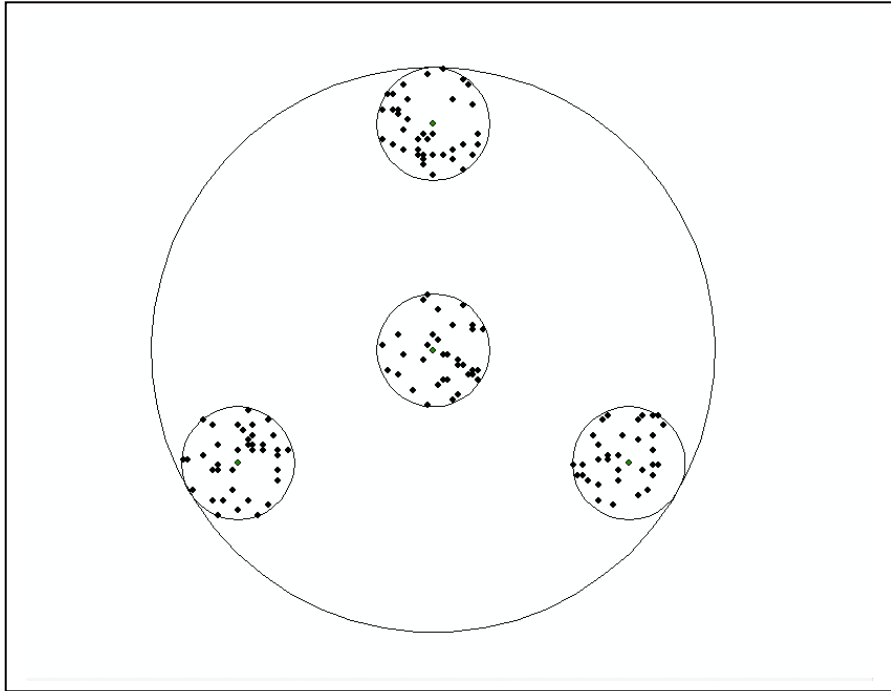


Figura 3.4. Muestreo de árboles en una población hipotética en Quintana, Roo, México, cuando el conglomerado es con subparcelas circulares.

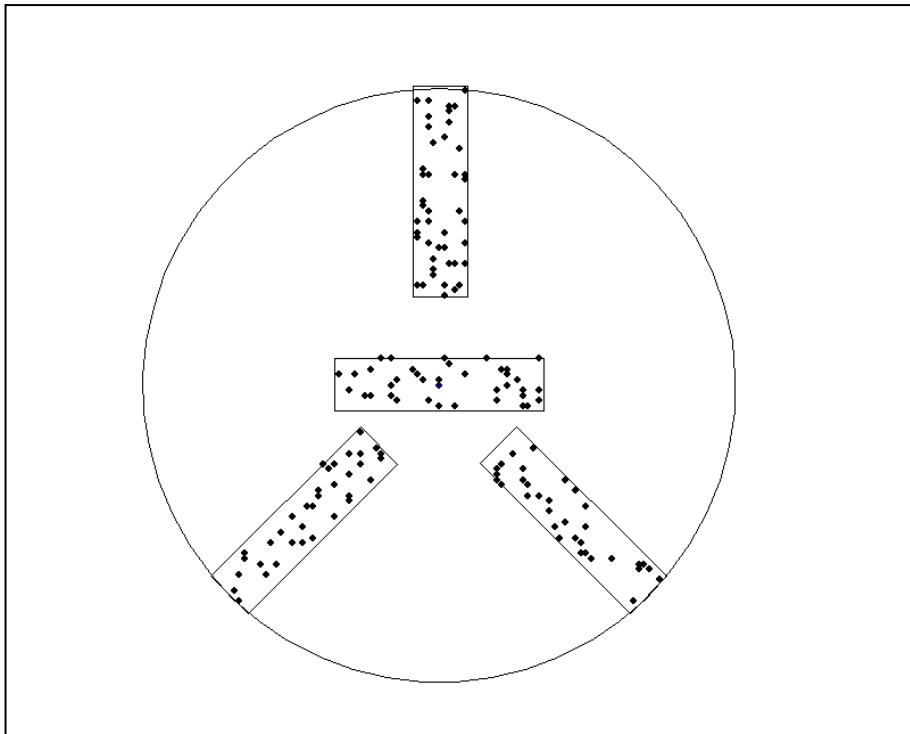


Figura 3.5. Muestreo de árboles en una población hipotética en Quintana, Roo, México, cuando el conglomerado es con subparcelas rectangulares.

Con el fin de evaluar las posibles diferencias entre las poblaciones generadas con conglomerados de subparcelas circulares y conglomerados de subparcelas rectangulares, en cada una de las dos redes sistemáticas se generaron pares de valores por hectárea —estimaciones a nivel conglomerado—; el primer valor correspondió a la estimación en base al muestreo en subparcelas circulares y el segundo en rectangulares. Mediante una prueba estadística de t pareada univariada (Montgomery y Runger, 2003) se compararon las poblaciones generadas a través de los dos tipos de subparcelas. Esto se realizó para las variables densidad y volumen en cada una de las condiciones del estudio. Adicionalmente se llevó a cabo una prueba de t pareada bivariada, en la cual la diferencia de densidad entre subparcelas circulares y rectangulares y la diferencia de volumen entre subparcelas circulares y rectangulares se analizaron simultáneamente mediante un MANOVA implementado en paquete R.

En segundo lugar, se generaron de manera adicional 86 redes sistemáticas considerando únicamente conglomerados de subparcelas circulares, por lo que finalmente se estudiaron 88 muestras de subparcelas circulares. En cada una de las 88 muestras se estimó el número de árboles y volumen maderable por hectárea mediante los estimadores media de razones (MR), razón de medias (RM), Forest Inventory and Analysis (FIA) y Van Deusen (VAN). Las varianzas estimadas (\hat{V}_{bck}) se obtuvieron de acuerdo a las expresiones teóricas esbozadas en la sección “derivación de estimadores”. El análisis conjunto de las 88 muestras permitió observar el comportamiento de los estimadores estudiados bajo el concepto de muestreo repetido. Para determinar el mejor estimador (k) en cada condición estudiada (c) se usó el sesgo relativo de la estimación y el sesgo relativo de la varianza. Estos sesgos se obtuvieron según las siguientes expresiones (Van Deusen, 2006):

Sesgo relativo de la estimación (SRE):

$$SRE = \frac{\left(\hat{\theta}_{ck} - \theta_{ck} \right)}{\theta_{ck}}$$

Donde

$k = 1,2,3,4$: estimadores

$c = 1,2,3$: condiciones

$$\hat{\theta}_{ck} = \frac{\sum_{b=1}^{88} \hat{\theta}_{bck}}{88} : \text{promedio de las estimaciones}$$

$\theta_{ck} = \theta = 937$: parámetro de la densidad (árboles ha^{-1})

$\theta_{ck} = \theta = 97.37$: parámetro de volumen ($\text{m}^3 \text{ha}^{-1}$)

Sesgo relativo de la varianza (SRV):

$$SRV = \frac{\left(\hat{V}_{ck} - V_{ck} \right)}{V_{ck}}$$

$$\hat{V}_{ck} = \frac{\sum_{b=1}^{88} \hat{V}_{bck}}{88} : \text{Promedio de varianzas estimadas}$$

$$V_{ck} = \frac{\sum_{b=1}^{88} \left(\hat{\theta}_{bck} - \hat{\theta}_{ck} \right)^2}{88} : \text{Varianza verdadera}$$

En las expresiones del sesgo relativo un valor cercano a 0 significa que la fórmula es insesgada, un valor de 1 muestra que la fórmula predice dos veces la varianza que debería ser, y un valor de 2 indica que la fórmula sobrestima por factor de 3.

3.7. RESULTADOS Y DISCUSIÓN

3.7.1. Comparación de conglomerados de subparcelas circulares y rectangulares

En el Cuadro 3.1 se presentan los valores de t y p correspondientes a las pruebas de t pareada univariada, es decir, donde la variable densidad y volumen se analizan por separado. A excepción de la variable densidad en la condición dos de la muestra dos, en todos los casos el valor de p supera el 0.05, por lo que con esta prueba se puede concluir que las poblaciones generadas por conglomerados de subparcelas circulares y rectangulares no son estadísticamente diferentes al considerar un nivel de significancia del 5%.

En el Cuadro 3.2 se muestran los valores de F (estadístico de Wilks) y p de las pruebas de t pareadas bivariadas. En todos los casos se observa que los valores de p superan en gran magnitud al 0.05, por tanto, con un nivel de significancia del 5% puede afirmarse que existen evidencias suficientes para pensar que los valores generados por las subparcelas circulares y rectangulares, correspondientes a las variables densidad y volumen, no son estadísticamente diferentes. Aun cuando las pruebas de t pareada univariada y la bivariada conducen al mismo resultado, la segunda se justifica al considerar que las variables diferencia de densidad (entre subparcelas circulares y rectangulares) y diferencia de volumen (entre subparcelas circulares y rectangulares) están altamente correlacionadas de manera positiva. En todas las condiciones estudiadas y en las dos muestras evaluadas las correlaciones resultaron ser significativas y mayores al 0.55.

Cuadro 3.1. Valores de t y p de la prueba de t pareada univariada para comparar poblaciones de conglomerados de subparcelas circulares y rectangulares.

Muestra	Condición	Variable	Valor de t	Valor de p
1	c0	Densidad	-1.22	0.2250
		Volumen	-0.55	0.5862
	c1	Densidad	-0.65	0.5206
		Volumen	-0.18	0.8615
	c2	Densidad	-0.70	0.4832
		Volumen	-0.31	0.7539
2	c0	Densidad	-1.30	0.1985
		Volumen	-0.89	0.3784
	c1	Densidad	-1.34	0.1853
		Volumen	-1.15	0.2555
	c2	Densidad	-2.20	0.0307
		Volumen	-1.96	0.0536

Cuadro 3.2. Valores de F y p de la prueba de t pareada bivariada (densidad y volumen simultáneamente) para comparar poblaciones de conglomerados de subparcelas circulares y rectangulares.

Muestra	Condición	Variable	Correlación entre diferencias*	Valor de F**	Valor de p
1	c0	Densidad y volumen	0.7192(<.0001)	0.4567	0.6351
	c1	Densidad y volumen	0.7244 (<.0001)	0.3527	0.7039
	c2	Densidad y volumen	0.7209 (<.0001)	0.2043	0.8157
2	c0	Densidad y volumen	0.6231(<.0001)	1.4688	0.2365
	c1	Densidad y volumen	0.5965 (<.0001)	1.8570	0.1630
	c2	Densidad y volumen	0.7247 (<.0001)	0.8575	0.4282

*Correlación de la diferencia de densidad y diferencia de volumen. En ambos casos la diferencia se obtuvo entre los valores de subparcelas circulares y rectangulares. **El valor de F corresponde la prueba de Wilks.

Al no existir diferencias significativas entre las poblaciones generadas mediante conglomerados de subparcelas circulares y rectangulares, permitió en este trabajo de investigación estudiar el comportamiento de los cuatro estimadores considerando únicamente subparcelas circulares, las cuales presentan mayor facilidad de trazo en ArcMap en comparación con las rectangulares. Así, en el primer caso solo se requiere especificar el radio del círculo a partir de la coordenada central, mientras que en el segundo, es necesario definir los vértices y los lados del rectángulo. En campo, la delimitación de las parcelas rectangulares también es complicada. Aun cuando algunos países usan unidades muestrales rectangulares, Indonesia por ejemplo, varios países como Suecia, Alemania, Estados Unidos utilizan parcelas circulares (Kangas y Maltamo, 2009).

3.7.2. Comparación de estimadores por variable y condición

En el Cuadro 3.3 se presentan los resultados del análisis estadístico relacionado a la comparación de los estimadores MR, RM, FIA y VAN en cada una de las tres condiciones estudiadas.

Cuadro 3.3. Sesgos relativos de la estimación y sesgos relativos de la varianza para cuatro estimadores estudiados en una población forestal artificial.

Estimación y sesgo/condición	Densidad (árboles ha ⁻¹)				Volumen (m ³ ha ⁻¹)			
	MR	RM	FIA	VAN	MR	RM	FIA	VAN
Estimación/c0	939.94	939.94	939.94	939.94	97.76	97.76	97.76	97.76
Parámetro/c0	937.00	937.00	937.00	937.00	97.37	97.37	97.37	97.37
Sesgo relativo de estimación/c0	0.0031	0.0031	0.0031	0.0031	0.004	0.004	0.004	0.004
Varianza estimada/c0	72.99	72.99	72.99	72.99	1.79	1.79	1.79	1.79
Varianza verdadera/c0	72.92	72.92	72.92	72.92	1.82	1.82	1.82	1.82
Sesgo relativo de varianza/c0	0.0009	0.0009	0.0009	0.0009	-0.0186	-0.0186	-0.0186	-0.0186
Estimación/c1	939.26	939.48	939.48	939.48	97.60	97.69	97.69	97.69
Parámetro/c1	937.00	937.00	937.00	937.00	97.37	97.37	97.37	97.37
Sesgo relativo de estimación/c1	0.0024	0.0026	0.0026	0.0026	0.0024	0.0033	0.0033	0.0033
Varianza estimada/c1	82.39	76.02	375.50	76.08	2.03	1.88	5.18	1.88
Varianza verdadera/c1	83.62	80.44	80.44	80.44	2.22	2.04	2.04	2.04
Sesgo relativo de varianza/c1	-0.0148	-0.0549	3.6682	-0.0542	-0.0824	-0.0781	1.5432	-0.0775
Estimación/c2	938.59	938.99	939.00	938.99	97.65	97.61	97.61	97.61
Parámetro/c2	937.00	937.00	937.00	937.00	97.37	97.37	97.37	97.37
Sesgo relativo de estimación/c2	0.0017	0.0021	0.0021	0.0021	0.0029	0.0024	0.0024	0.0024
Varianza estimada/c2	102.24	87.16	965.45	87.37	2.50	2.11	11.51	2.12
Varianza verdadera/c2	96.51	84.80	84.81	84.80	2.69	2.19	2.19	2.19
Sesgo relativo de varianza/c2	0.0594	0.0278	10.3843	0.0302	-0.0714	-0.0341	4.2576	-0.0319

MR: Media de razones, RM: Razón de medias, FIA: Forest Inventory and Analysis, VAN: Van Deusen

Enseguida se indican los resultados para cada condición y variable de interés.

3.7.2.1. Condición c0 y variable densidad (árboles ha⁻¹)

En c0 y para densidad, las estimaciones puntuales de los cuatro estimadores fueron las mismas y en promedio estiman 940 árboles ha⁻¹. La estimación más baja fue de 912 y la más alta de 959 árboles ha⁻¹. El sesgo relativo promedio fue de 0.0031, mientras que el mínimo y el máximo resultaron ser de -0.0261 y 0.0240, respectivamente (Figura 3.6). Por lo tanto, las estimaciones puntuales de los cuatro estimadores son prácticamente insesgadas para la variable densidad cuando todos los conglomerados presentan las cuatro subparcelas circulares.

Las varianzas estimadas de los cuatro estimadores también fueron las mismas, y en consecuencia, los sesgos relativos de la varianza son idénticos y en promedio fueron de 0.0009. El sesgo relativo menor de la varianza fue -0.3436 y el mayor fue de 0.3028 (Figura 3.7). Entonces, las estimaciones puntuales y las varianzas estimadas de los cuatro estimadores, para la variable densidad cuando todos los conglomerados son completos, se consideran insesgadas.

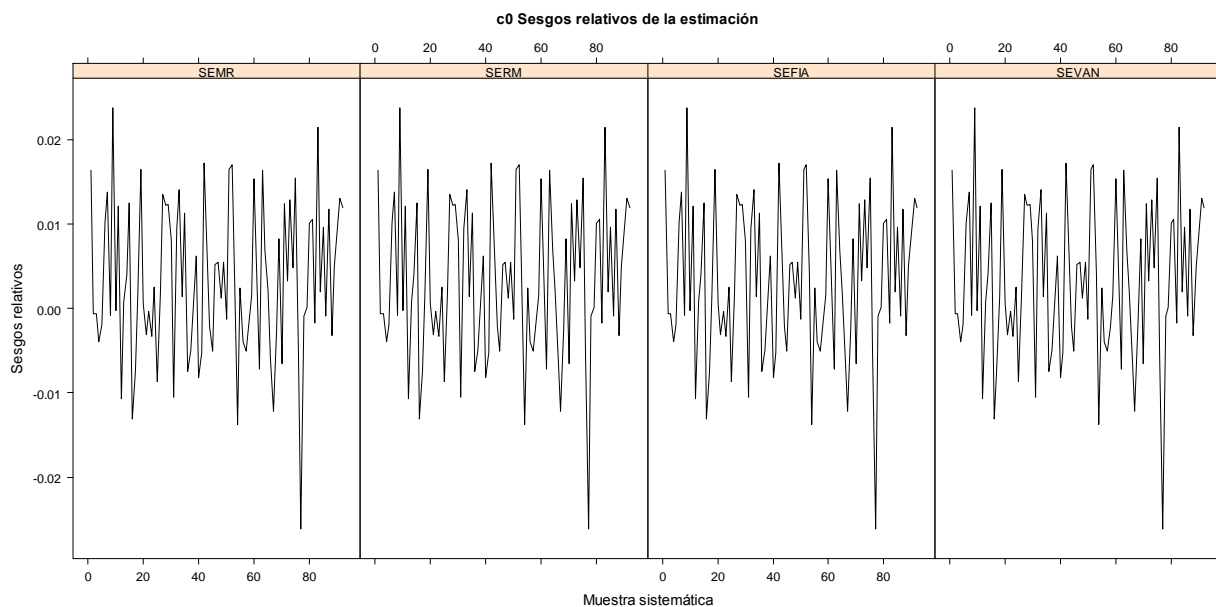


Figura 3.6. Sesgos relativos de la estimación. Variable densidad en condición c0. SEMR: sesgo relativo de la estimación para Media de Razones. SERM: sesgo relativo de la estimación para Razón de Medias. SEFIA: sesgo relativo de la estimación para Forest Inventory and Analysis. SEVAN: sesgo relativo de la estimación para Van Deusen.

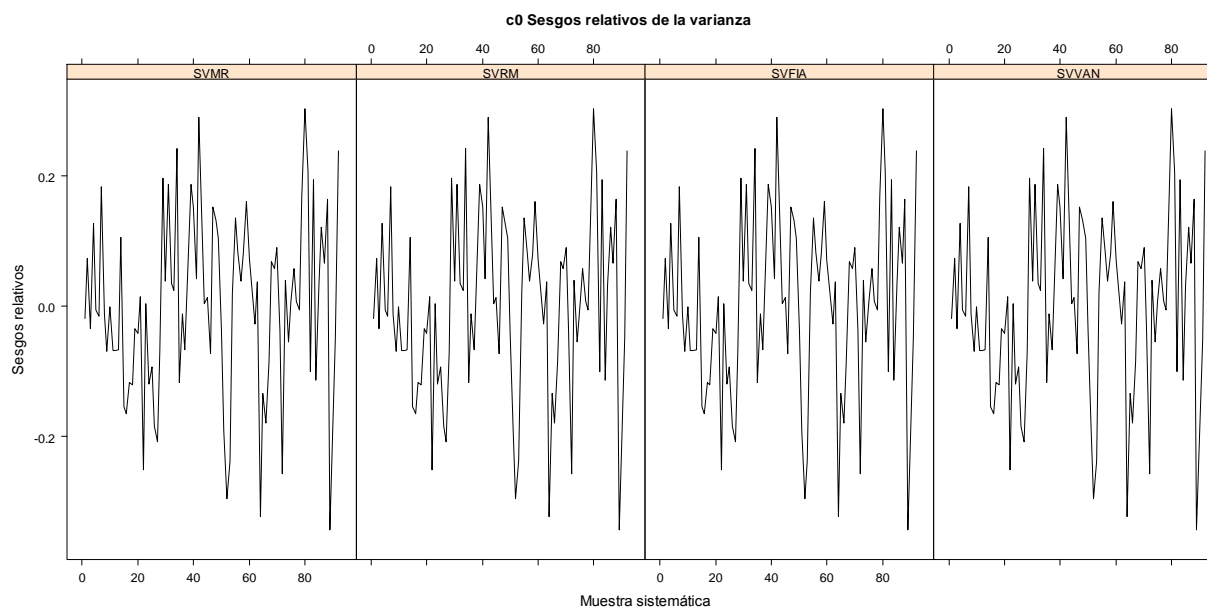


Figura 3.7. Sesgos relativos de la varianza. Variable densidad en condición c0. SVMR: sesgo relativo de la varianza para Media de Razones. SVRM: sesgo relativo de la varianza para Razón de Medias. SVFIA: sesgo relativo de la varianza para Forest Inventory and Analysis. SVVAN: sesgo relativo de la varianza para Van Deusen.

3.7.2.2. Condición c1 y variable densidad (árboles ha⁻¹)

En c1 y para densidad, las estimaciones puntuales de RM, FIA y VAN son iguales y varían ligeramente con el estimador MR. En promedio, los cuatros estimadores estiman 939 árboles ha⁻¹. La estimación más baja fue de 910 y la más alta de 960 (árboles ha⁻¹), ambas correspondientes a RM, FIA y VAN. El sesgo relativo promedio de la estimación para MR fue de 0.0024 y para RM, FIA y VAN fue de 0.0026. El sesgo relativo mínimo fue de -0.0290 y el máximo de 0.0246, ambos de RM, FIA y VAN (Figura 3.8). Por tanto, los cuatro estimadores se consideran prácticamente insesgados para la variable densidad (árboles ha⁻¹).

El sesgo relativo promedio de la varianza para RM y VAN fue de -0.054, y para para MR fue de -0.0148. El valor mínimo fue de -0.4148 y el máximo de 0.3868, ambos correspondientes a MR (Figura 3.9). Entonces, la varianza estimada en tres estos casos se puede considerar insesgada. Sin embargo, el sesgo promedio relativo de la varianza para FIA fue de 3.6681, por lo que la varianza estimada es cuatro veces más grande que la varianza verdadera, de manera tal que la varianza estimada para FIA es

fuertemente sesgada cuando diez de los 81 conglomerados están incompletos en diferente magnitud.

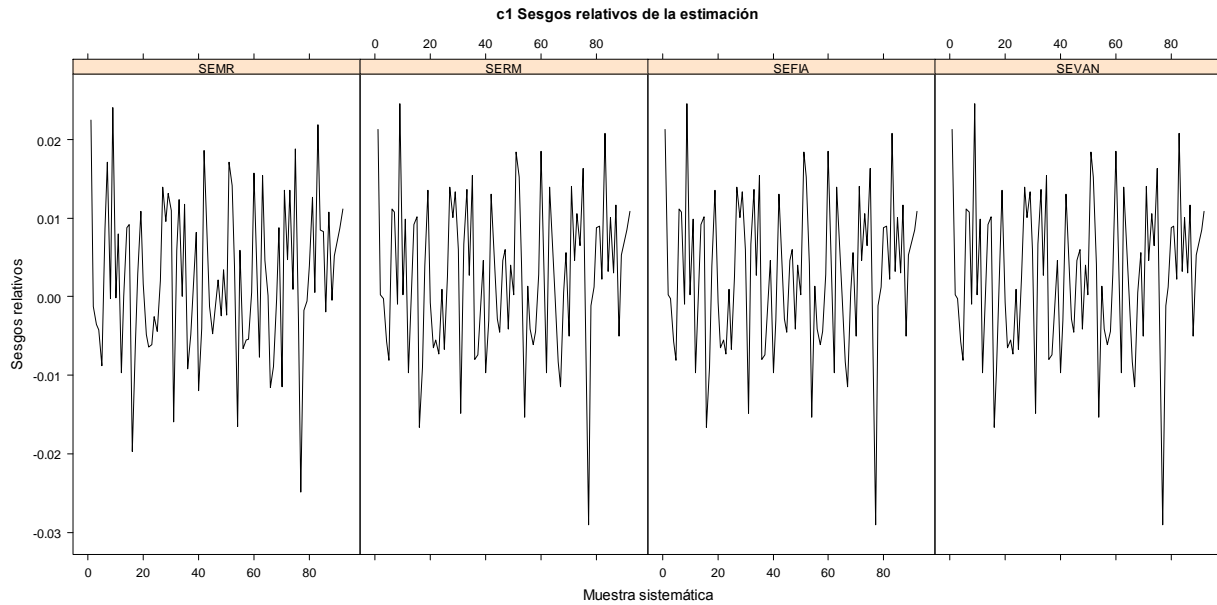


Figura 3.8. Sesgos relativos de la estimación. Variable densidad en condición c1.

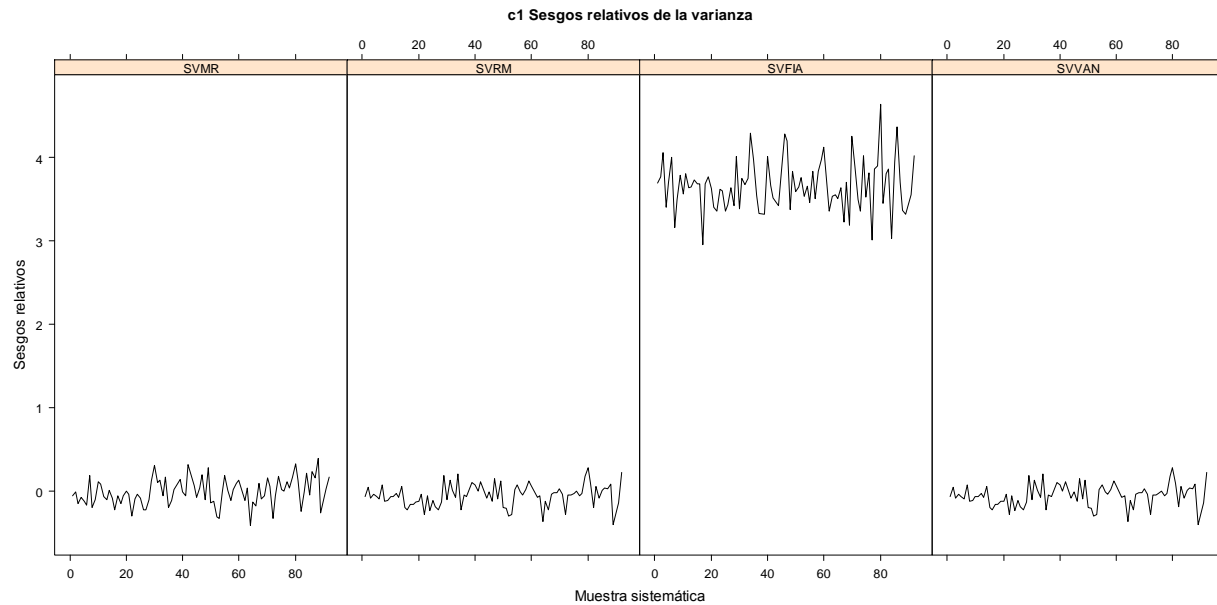


Figura 3.9. Sesgos relativos de la varianza. Variable densidad en condición c1.

3.7.2.3. Condición c2 y variable densidad (árboles ha⁻¹)

En c2 y densidad, las estimaciones puntuales de RM, FIA y VAN son iguales, y difieren ligeramente de MR, de hecho, en promedio los cuatros estimadores estiman 939 árboles ha⁻¹. La estimación mínima fue de 918 árboles ha⁻¹ correspondiente a RM, FIA y VAN. La estimación máxima fue de 967 correspondiente a MR. El sesgo relativo promedio de estimación para RM, FIA y VAN fue de 0.0021, mientras que para MR fue de 0.0017. El sesgo relativo mínimo fue de -0.0199 correspondiente a RM, FIA y VAN. El sesgo relativo máximo fue de 0.0318 para MR (Figura 3.10). En estas condiciones, los cuatro estimadores se consideran insesgados.

Los sesgos relativos promedio de la varianza fueron de 0.0594, 0.0278 y 0.0302, respectivamente, para MR, RM y VAN. El sesgo relativo mínimo fue de -0.3347 para RM. El sesgo relativo máximo fue de 0.6678 para MR. Por lo tanto, las varianzas estimadas de MR, RM y VAN son insesgadas (Figura 3.11). No obstante, el sesgo relativo promedio de la varianza para FIA fue de 10.3843, por lo que la varianza estimada es once veces mayor que la varianza verdadera. Entonces, el estimador de la varianza de FIA es totalmente sesgado cuando se tienen conglomerados fuertemente incompletos y se estima densidad.

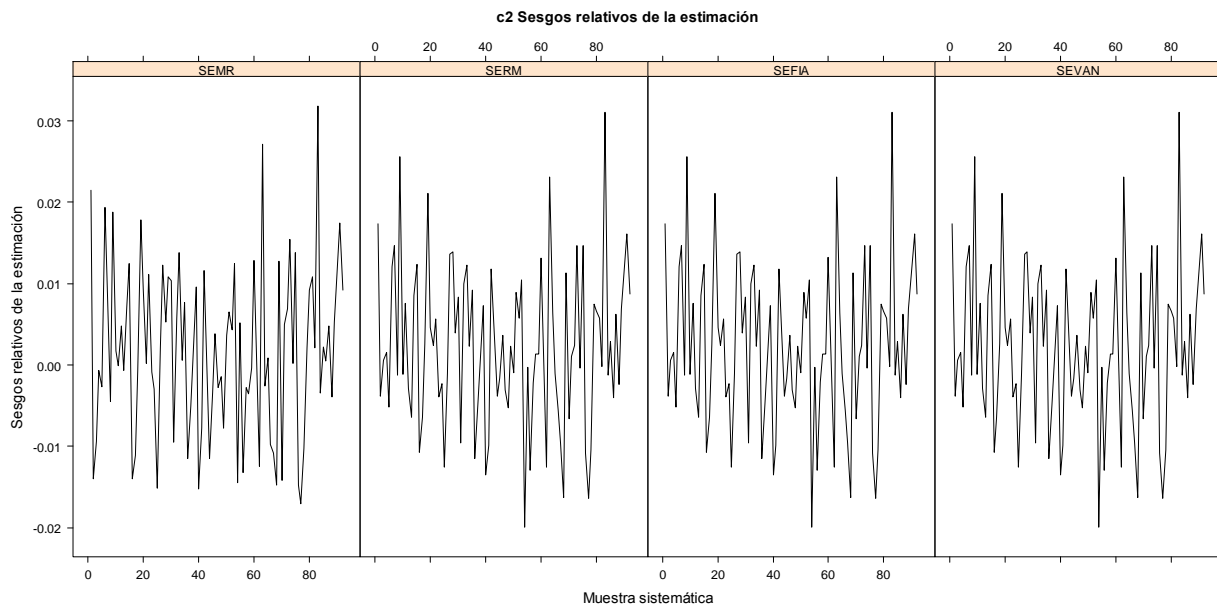


Figura 3.10. Sesgos relativos de la estimación. Variable densidad en condición c2.

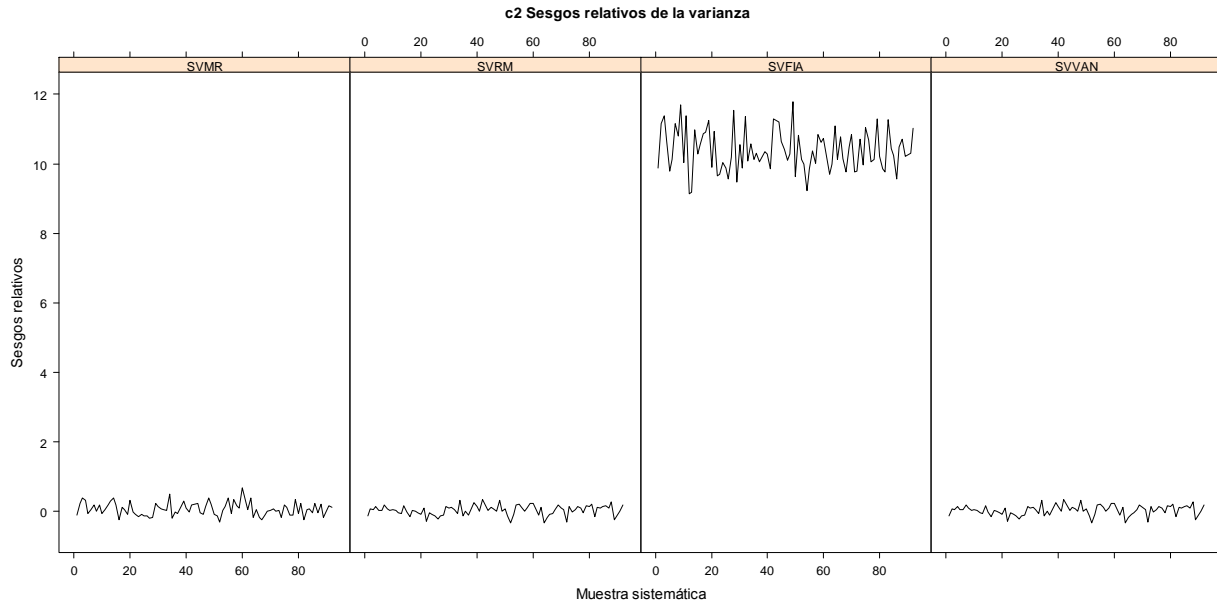


Figura 3.11. Sesgos relativos de la varianza. Variable densidad en condición c2.

3.7.2.4. Condición c0 y variable volumen ($\text{m}^3 \text{ha}^{-1}$)

En c0, los cuatro estimadores presentan los mismos valores; estiman un volumen promedio de $97.76 \text{ m}^3 \text{ha}^{-1}$. La estimación mínima fue de 94.89 y la máxima de $101.59 \text{ m}^3 \text{ha}^{-1}$. Todas las estimaciones oscilan alrededor del parámetro ($97.37 \text{ m}^3 \text{ha}^{-1}$). El sesgo relativo promedio de estimación fue 0.0040 . El sesgo relativo mínimo fue de -0.0254 y el máximo de 0.0433 (Figura 3.12). Así en c0 y para volumen, los cuatro estimadores se consideran insesgados.

Los cuatro estimadores presentan el mismo error relativo de la varianza. El error relativo promedio de la varianza fue de -0.0186 . El error relativo mínimo fue de -0.3831 y el máximo de 0.4230 (Figura 3.13). La varianza de los cuatros estimadores es insesgada.

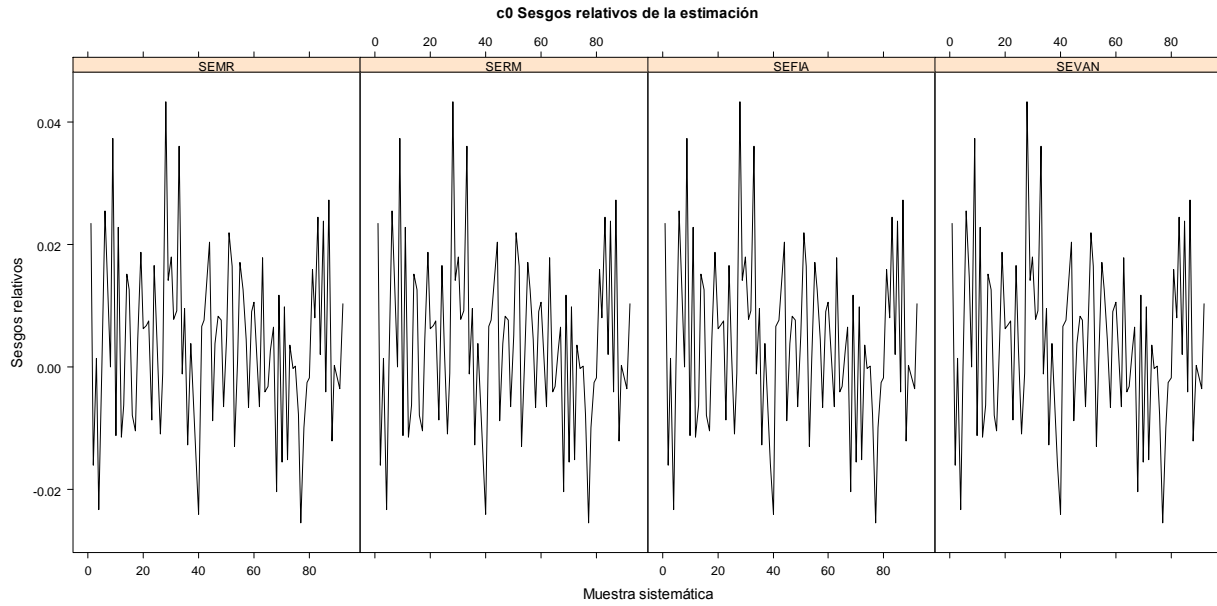


Figura 3.12. Sesgos relativos de la estimación. Variable volumen en condición c0.

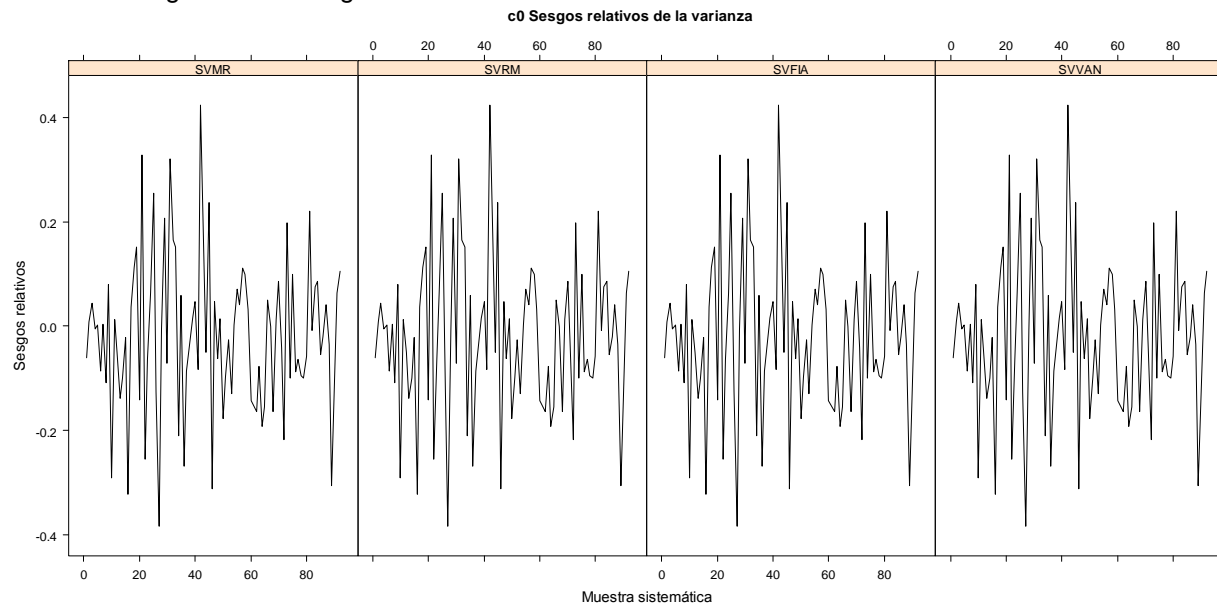


Figura 3.13. Sesgos relativos de la varianza. Variable volumen en condición c0.

3.7.2.5. Condición c1 y variable volumen ($\text{m}^3 \text{ha}^{-1}$)

En c1, los estimadores RM, FIA y VAN coinciden y en promedio estiman $97.69 \text{ m}^3 \text{ha}^{-1}$. Esta estimación varía ligeramente con MR, la cual fue de $97.60 \text{ m}^3 \text{ha}^{-1}$. La estimación mínima fue de 94.36 correspondiente a MR, y la máxima fue de $101.41 \text{ m}^3 \text{ha}^{-1}$ correspondiente a RM, FIA y VAN. Los estimadores RM, FIA y VAN presentaron un sesgo relativo promedio de estimación de 0.0033 , mientras que para MR fue de 0.0024 . El sesgo relativo de estimación mínimo fue de -0.0310 de MR y el máximo de 0.0415 correspondiente RM, FIA y VAN (Figura 3.14). Entonces, para la variable volumen en la condición c1 los cuatro estimadores se consideran incesgados.

En c1, los sesgos relativos promedio de la varianza fueron de -0.0824 , -0.0781 y -0.0775 para MR, RM y VAN, respectivamente. El valor mínimo fue de -0.4705 y el máximo de 0.4271 , ambos MR (Figura 3.15). La varianza estimada para tres estos estimadores puede considerarse incesgada. Sin embargo, el sesgo relativo promedio de la varianza para FIA fue de 1.5432 , es decir, la varianza estimada es más de dos veces la varianza verdadera, por lo tanto, la varianza estimada de FIA para volumen en c1 se considera sesgada.

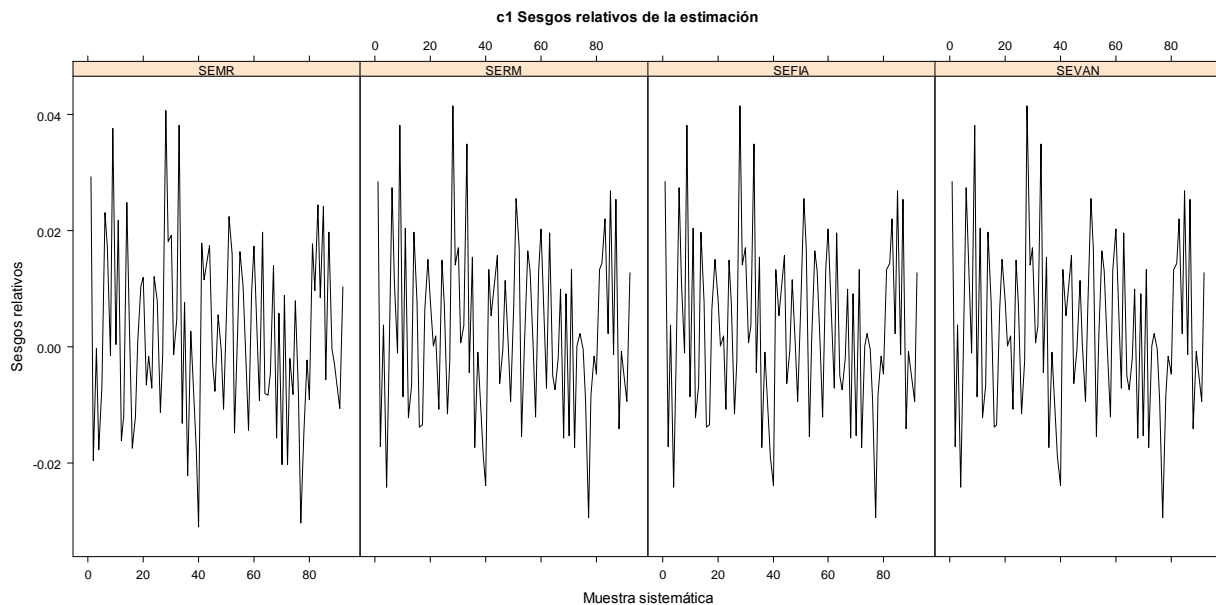


Figura 3.14. Sesgos relativos de la estimación. Variable volumen en condición c1.

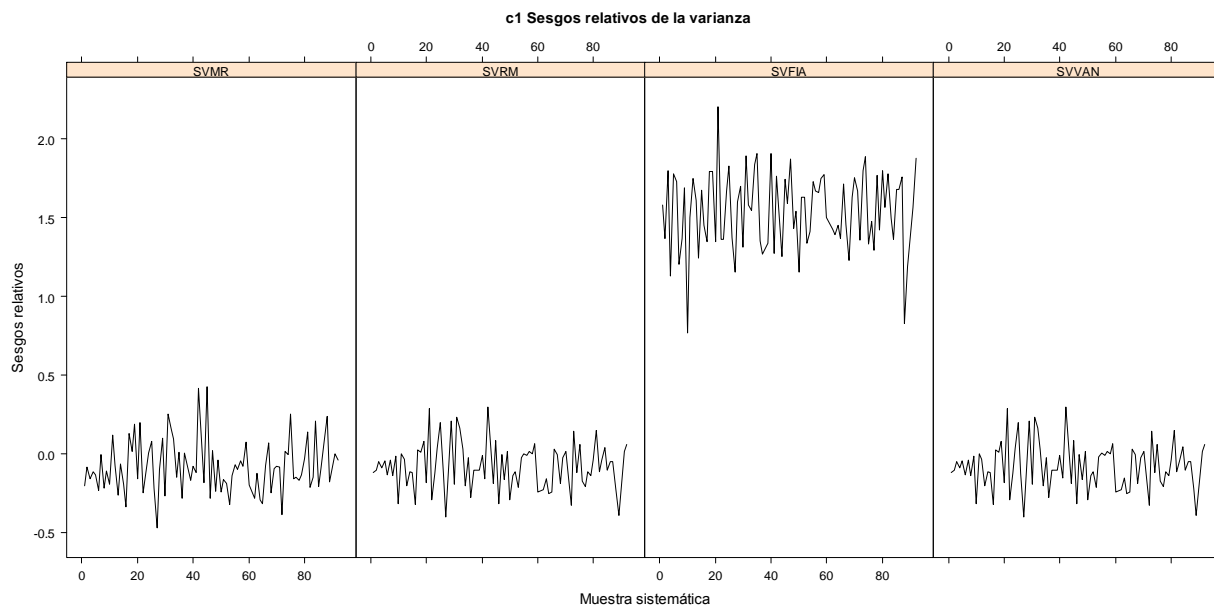


Figura 3.15. Sesgos relativos de la varianza. Variable volumen en condición c1.

3.7.2.6. Condición c2 y variable volumen ($\text{m}^3 \text{ha}^{-1}$)

En c2, los estimadores RM, FIA y VAN coinciden y en promedio estiman $97.61 \text{ m}^3 \text{ha}^{-1}$. Varían poco con MR, el cual estima $97.65 \text{ m}^3 \text{ha}^{-1}$. El valor mínimo fue de $93.34 \text{ m}^3 \text{ha}^{-1}$ correspondiente a MR y el máximo fue de $101.64 \text{ m}^3 \text{ha}^{-1}$ correspondiente a MR, FIA y VAN. El sesgo promedio para RM, FIA y VAN fue de 0.0024, mientras que para MR fue de 0.0029. El valor mínimo fue de -0.0414 correspondiente a MR y el máximo de 0.0438 correspondiente a RM, FIA y VAN (Figura 3.16). Estos estimadores se consideran insesgados en la condición C2 para la variable volumen.

En c2, el sesgo relativo promedio de la varianza para MR fue de -0.0714, para RM fue de -0.0341 y para VAN fue de -0.0319. El valor mínimo fue -0.4120 y el máximo de 0.7427, ambos de MR (Figura 3.17). Los estimadores de la varianza para MR, RM y VAN se consideran insesgados. Sin embargo, el sesgo relativo promedio de la varianza para FIA fue de 4.2576; la varianza estimada fue cinco mayor que la varianza verdadera, por lo tanto la varianza de FIA se considera altamente sesgada.

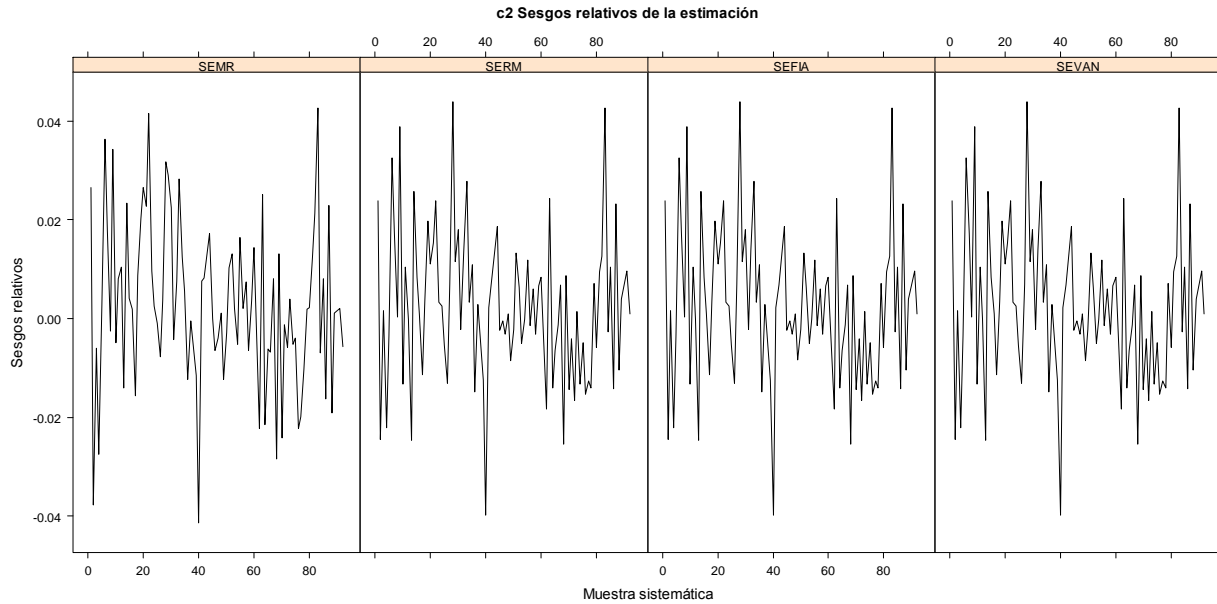


Figura 3.16. Sesgos relativos de la estimación. Variable volumen en condición c2.

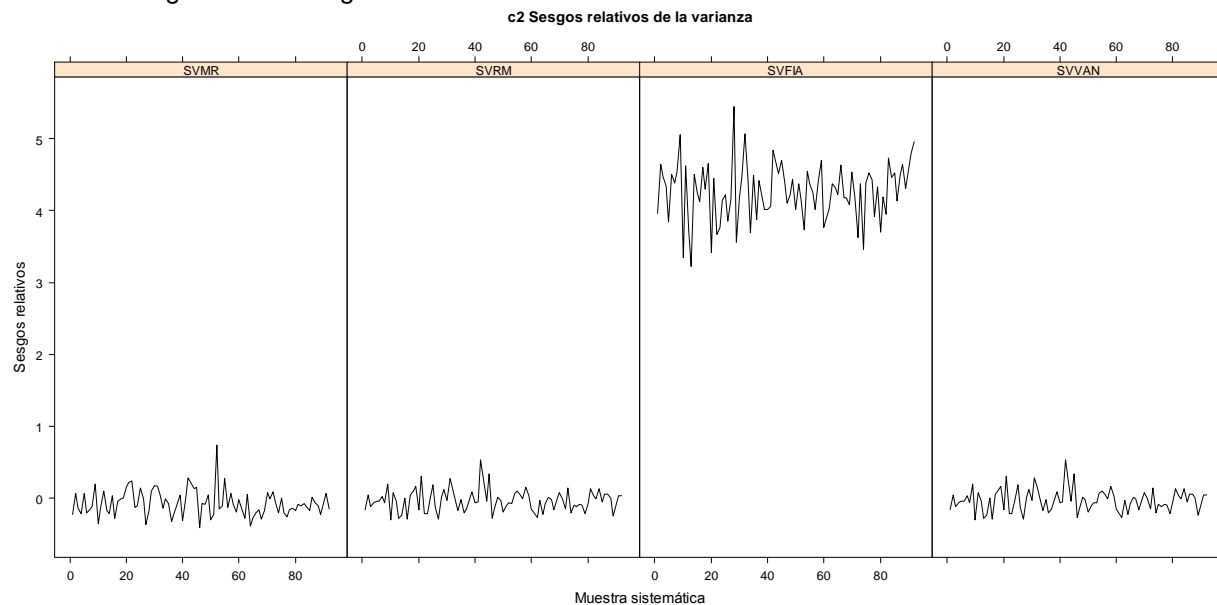


Figura 3.17. Sesgos relativos de la varianza. Variable volumen en condición c2.

Las estimaciones puntuales y por intervalo de las 88 muestras, considerando las tres condiciones estudiadas sobre la completitud del conglomerado, para densidad y volumen, se presentan respectivamente en el ANEXO 3.3 y 3.4. Como era de esperarse, en C0 los cuatro intervalos de confianza coinciden, sin embargo, a medida

que el grado de completitud disminuye, los intervalos correspondientes a FIA se hacen de mayor amplitud.

Los resultados obtenidos en la condición c_0 , estimaciones y varianzas estimadas idénticas para los cuatro estimadores en las dos variables estudiadas, eran de esperarse, tal y como lo señalan las ecuaciones (3.23) a (3.26) y (3.27) a (3.30). Así, bajo c_0 puede aplicarse cualquiera de los cuatro estimadores ensayados en este trabajo, inclusive FIA, cuya expresión matemática considera una proporción promedio de la parcela que pertenece al tipo forestal de interés. Esta situación es de particular importancia porque las fórmulas se reducen a los estimadores convencionales del muestreo aleatorio simple.

Por otra parte, el hecho de que las estimaciones puntuales de FIA, VAN y RM resultaran iguales tanto en la densidad como el volumen, incluso cuando los conglomerados son incompletos — c_1 y c_2 — no es sorprendente, según lo señalado en las expresiones (3.31) a (3.33).

Los sesgos relativos de las varianzas en c_0 y c_1 son congruentes con los reportados por Van Deusen (2004) y Van Deusen (2005). Este autor, al comparar FIA y VAN Deusen, obtuvo un sesgo relativo de la varianza de FIA de casi 2 cuando analiza una proporción promedio de parcela de 0.75, mientras que Van Deusen siempre se comporta insesgado. Estos resultados se deben a que varianza estimada de FIA está basada en la noción de que cada medición de la parcela se divide por la proporción promedio de la parcela que pertenece al tipo forestal de interés. Por lo tanto, el hecho de que esta varianza trabaje pobremente cuando las parcelas tienen proporciones que difieren ampliamente no es sorprendente.

Los resultados obtenidos por Van Deusen (2005) indican que las varianzas relativas del estimador media de razones y Van Deusen también resultaron muy similares (menores del 0.11), es decir, no hay ganancia significativa en precisión al considerar que la varianza del error es una función de la proporción al cuadrado, por esta razón el autor no recomienda el uso del estimador media de razones.

Por lo anterior resulta evidente que los estimadores MR, RM y VAN pueden tener aplicación general; se desempeñan bien —estimación puntual y por intervalo— cuando en los conglomerados se evalúan las cuatro subparcelas o sólo algunas de ellas, por el contrario, la estimación por intervalo del estimador FIA en conglomerados incompletos no es recomendable.

Las varianzas de los estimadores RM y VAN conducen a los mismos resultados porque el factor que relaciona ambas fue de 0.999 y de 0.998 en la condición uno y dos, respectivamente. Estos resultados son lógicos al considerar que cuando $\bar{a} = (1/2)a_0$ y $m = 20$, el factor es de 0.9473 (Van Deusen, 2005).

El estimador Van Deusen además de poder ser aplicado en conglomerados de subparcelas disjuntas como las del Inventario Nacional Forestal y de Suelos de México o del Forest Inventory and Analysis (FIA) de Estados Unidos, proyectos que se caracterizan por ser de gran escala, también pueden usarse en estudios forestales de nivel local, donde las parcelas de muestreo de área fija traslapan el límite de la propiedad. El requerimiento en este caso es el área de las parcelas que pertenece a la propiedad que se requiere evaluar. El supuesto más importante del estimador Van Deusen, hecho explícito en la ecuación (3.10), es que la cantidad esperada de la variable de interés en una parcela parcial es proporcional al área que pertenece al tipo forestal o propiedad objeto de estudio. Van Deusen y Roesch (2009) muestran que el error estándar del estimador Van Deusen se reduce hasta en un 15% cuando se incluyen parcelas mapeadas que traslapan el límite en comparación que cuando se excluyen.

Además de ser un estimador atractivo desde el punto de vista teórico-práctico, el estimador Van Deusen puede aplicarse fácilmente en muestreo aleatorio estratificado (Van Deusen, 2005):

$$\bar{y}_{st} = \frac{1}{A} \sum_{h=1}^L A_h \bar{y}_h \quad (3.34)$$

y

$$v(\bar{y}_{st}) = \sum_{h=1}^L \left(\frac{A_h}{A} \right)^2 v(\bar{y}_h) \quad (3.35)$$

Donde

A : Área total (ha) de la región de estudio, la cual se divide en L estratos.

A_h : Área (ha) de cada estrato.

\bar{y}_h : Media por hectárea de cada estrato obtenida mediante la expresión (3.13).

$v(\bar{y}_h)$: Varianza estimada de la media por hectárea de cada estrato obtenida mediante la expresión (3.14).

Las fórmulas convencionales de muestreo estratificado (Cochran, 1993 y Thompson, 2002) son dadas en términos de N_h y N , sin embargo, esto no es de gran utilidad en parcelas de inventarios forestales, dado que es difícil dividir la población en unidades discretas de tamaños diferentes.

3.8. CONCLUSIONES

En la condición de completitud total, los cuatro métodos estudiados condujeron a las mismas estimaciones, y a las mismas varianzas estimadas, para las variables volumen y árboles por hectárea. También en este caso, las estimaciones por hectárea y las varianzas estimadas fueron insesgadas independientemente del estimador. En esta situación es de particular importancia puesto que las fórmulas se reducen a los estimadores del muestreo aleatorio simple.

En cada una de las tres condiciones de completitud, las estimaciones de volumen y número de árboles por hectárea de MR, RM, FIA y VAN resultaron similares y prácticamente insesgadas.

Las varianzas estimadas de MR, RM y VAN pueden considerarse insesgadas aun en condiciones de completitud parcial.

A medida que el grado de completitud fue menor, tanto para volumen como para densidad, la varianza estimada del estimador FIA se volvió fuertemente sesgada.

3.9. LITERATURA CITADA

- Bechtold W. A. and C.T. Scott. 2005. The forest inventory and analysis plot design. In: The enhanced Forest Inventory and Analysis program-national sampling design and estimation procedures. W. A. Bechtold and P. L. Patterson (Eds.). General Technical Report SRS-80. Forest Service. United States Department of Agriculture. Southern Research Station, Asheville, North Carolina. USA. pp: 27-42.
- Cochran, W. G. 1993. Técnicas de Muestreo. Compañía Editorial Continental. México. 513 p.
- Condes, S. y J. Martínez-Millán. 1998. Comparación entre los índices de distribución espacial de árboles más usados en el ámbito forestal. Invest. Agr.: Sist. Recur. For. Vol. 7 (1 y 2): 173-187.
- Kangas A. and Maltano (eds.). 2009. Forest Inventory: Methodology and Applications. Springer Science+Business Media B.V. pp:295-340.
- Montgomery D.C. and Runger G.C. 2003. Applied Statistics and Probability for Engineers. John Wiley & Sons, Inc. New York, USA. 976 p.
- Scott, C.T., W. A. Bechtold, G. A. Reams, W. D. Smith, J. A. Westfall, M. H. Hansen and G.G. Moisen. 2005. Sample-based estimators used by the Forest Inventory and Analysis national information management system. In: The enhanced Forest Inventory and Analysis program-national sampling design and estimation procedures. W. A. Bechtold and P. L. Patterson (Eds.). General Technical Report SRS-80. Forest Service. United States Department of Agriculture. Southern Research Station, Asheville, North Carolina. USA. pp: 43-67.
- Stamatellos, G. y G. Panourgias (2005). Simulating spatial distributions of forest trees by using data from fixed area plots. Forestry, 78(3):305-312.
- Thompson, S.K. 2002. Sampling. John Wiley & Sons. New York, USA.367 p.
- Van Deusen, P.C. 2004. Forest inventory estimation with mapped plots. Canadian Journal of Forest Research. 34:493-497.
- Van Deusen, P.C. 2005. Stratified forest inventory estimation with mapped plots. Canadian Journal of Forest Research. 35:2392-2386.
- Van Deusen, P.C. 2006. Comparing mapped plot estimators. In Proceedings of the Sixth Annual Forest Inventory and Analysis Symposium, 21-24 September 2004. Denver, Co, U.S. For. Ser. Gen. Tech. Rep. WO-70.
- Van Deusen, P.C. 2007. Weighted estimators for forest inventory mapped plots. Canadian Journal of Forest Research.37:1142-1145.

Van Deusen, P.C. and F. A. Roesch. 2009. Application of mapped plots for single-owner forest surveys. *Journal of Forestry*. 107 (8):414-417.

Velasco B., E., H. Ramírez M., F. Moreno S. y A. de la Rosa V. 2005. Estimadores de razón para el inventario nacional forestal de México. *Revista Ciencia Forestal en México*. 28(94):23-43.

ANEXO 3.1. DERIVACIÓN DE ESTIMADORES: CASO PARTICULAR DOS OBSERVACIONES

Modelo FIA (ecuación 3.5)

$$y_s^p = \bar{a}^p \mu + \varepsilon 1_s, E(\varepsilon 1_s) = 0, E(\varepsilon 1_s, \varepsilon 1_s') = (\bar{a}^p)^2 \sigma^2 \quad \forall \quad \varepsilon 1_s = \varepsilon 1_s'$$

Sean

$$Y = \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix}; A = \begin{bmatrix} \bar{a}^p \\ \bar{a}^p \end{bmatrix}; \Sigma = \begin{bmatrix} \sigma^2 (\bar{a}^p)^2 & 0 \\ 0 & \sigma^2 (\bar{a}^p)^2 \end{bmatrix}; \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma^2 (\bar{a}^p)^2} & 0 \\ 0 & \frac{1}{\sigma^2 (\bar{a}^p)^2} \end{bmatrix}$$

Aplicando la ecuación (3.3)

$$\hat{\mu} = [A' \Sigma^{-1} A]^{-1} A' \Sigma^{-1} Y$$

$$\begin{aligned} \hat{\mu} &= \left\{ \begin{bmatrix} \bar{a}^p & \bar{a}^p \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2 (\bar{a}^p)^2} & 0 \\ 0 & \frac{1}{\sigma^2 (\bar{a}^p)^2} \end{bmatrix} \begin{bmatrix} \bar{a}^p \\ \bar{a}^p \end{bmatrix} \right\}^{-1} \left\{ \begin{bmatrix} \bar{a}^p & \bar{a}^p \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2 (\bar{a}^p)^2} & 0 \\ 0 & \frac{1}{\sigma^2 (\bar{a}^p)^2} \end{bmatrix} \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} \frac{1}{\sigma^2 (\bar{a}^p)} & \frac{1}{\sigma^2 (\bar{a}^p)} \end{bmatrix} \begin{bmatrix} \bar{a}^p \\ \bar{a}^p \end{bmatrix} \right\}^{-1} \left\{ \begin{bmatrix} \frac{1}{\sigma^2 (\bar{a}^p)} & \frac{1}{\sigma^2 (\bar{a}^p)} \end{bmatrix} \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix} \right\} \\ &= \left[\frac{1}{\sigma^2} + \frac{1}{\sigma^2} \right]^{-1} \left[\frac{y_1^p}{\sigma^2 (\bar{a}^p)} + \frac{y_2^p}{\sigma^2 (\bar{a}^p)} \right] = \left[\frac{2}{\sigma^2} \right]^{-1} \left[\frac{y_1^p + y_2^p}{\sigma^2 (\bar{a}^p)} \right] = \left[\frac{\sigma^2}{2} \right] \left[\frac{y_1^p + y_2^p}{\sigma^2 (\bar{a}^p)} \right] = \left[\frac{y_1^p + y_2^p}{2 (\bar{a}^p)} \right] \\ &= \frac{\sum_{s=1}^2 \frac{y_s^p}{\bar{a}^p}}{2} \end{aligned}$$

$$\text{En general, } \hat{\mu}^{FIA} = \frac{\sum_{s=1}^m \frac{y_s^p}{\bar{a}^p}}{m} = \frac{\sum_{s=1}^m \tilde{y}_s}{m}, \text{ donde } \tilde{y}_s = \frac{y_s^p}{\bar{a}^p}$$

Ahora, aplicando la expresión (3.4)

$$V(\hat{\mu}) = [A' \Sigma^{-1} A]^{-1}$$

$$V(\hat{\mu}) = \left\{ \begin{bmatrix} \bar{a}^p & \bar{a}^p \\ a^p & a^p \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2 (\bar{a}^p)^2} & 0 \\ 0 & \frac{1}{\sigma^2 (\bar{a}^p)^2} \end{bmatrix} \begin{bmatrix} \bar{a}^p \\ a^p \end{bmatrix} \right\}^{-1} = \left[\frac{2}{\sigma^2} \right]^{-1} = \frac{\sigma^2}{2}$$

En general, $V(\hat{\mu}^{FIA}) = \frac{\sigma^2}{m}$ y $v(\hat{\mu}^{FIA}) = \frac{\hat{\sigma}^2}{m}$, donde, $\hat{\sigma}^2 = \frac{\sum_{s=1}^m (\tilde{y}_s - \hat{\mu})^2}{m-1}$

Modelo Van Deusen (ecuación 3.10)

$$y_s^p = a_s^p \mu + \varepsilon 2_s, E(\varepsilon 2_s) = 0, E(\varepsilon 2_s, \varepsilon 2_s') = a_s^p \sigma^2 \quad \forall \quad \varepsilon 2_s = \varepsilon 2_s'$$

Sean:

$$Y = \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix}; A = \begin{bmatrix} a_1^p \\ a_2^p \end{bmatrix}; \Sigma = \begin{bmatrix} \sigma^2 a_1^p & 0 \\ 0 & \sigma^2 a_2^p \end{bmatrix}; \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma^2 a_1^p} & 0 \\ 0 & \frac{1}{\sigma^2 a_2^p} \end{bmatrix}$$

Aplicando la ecuación (3.3)

$$\hat{\mu} = [A' \Sigma^{-1} A]^{-1} A' \Sigma^{-1} Y$$

$$\begin{aligned} \hat{\mu} &= \left\{ \begin{bmatrix} a_1^p & a_2^p \\ a_1^p & a_2^p \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2 a_1^p} & 0 \\ 0 & \frac{1}{\sigma^2 a_2^p} \end{bmatrix} \begin{bmatrix} a_1^p \\ a_2^p \end{bmatrix} \right\}^{-1} \left\{ \begin{bmatrix} a_1^p & a_2^p \\ a_1^p & a_2^p \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2 a_1^p} & 0 \\ 0 & \frac{1}{\sigma^2 a_2^p} \end{bmatrix} \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} \frac{1}{\sigma^2} & \frac{1}{\sigma^2} \\ \frac{1}{\sigma^2} & \frac{1}{\sigma^2} \end{bmatrix} \begin{bmatrix} a_1^p \\ a_2^p \end{bmatrix} \right\}^{-1} \left\{ \begin{bmatrix} \frac{1}{\sigma^2} & \frac{1}{\sigma^2} \\ \frac{1}{\sigma^2} & \frac{1}{\sigma^2} \end{bmatrix} \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix} \right\} \end{aligned}$$

$$= \left[\frac{a_1^p}{\sigma^2} + \frac{a_2^p}{\sigma^2} \right]^{-1} \left[\frac{y_1^p}{\sigma^2} + \frac{y_2^p}{\sigma^2} \right] = \left[\frac{a_1^p + a_2^p}{\sigma^2} \right]^{-1} \left[\frac{y_1^p + y_2^p}{\sigma^2} \right] = \left[\frac{\sigma^2}{a_1^p + a_2^p} \right] \left[\frac{y_1^p + y_2^p}{\sigma^2} \right]$$

$$= \frac{y_1^p + y_2^p}{a_1^p + a_2^p} = \frac{\sum_{s=1}^2 y_s^p}{\sum_{s=1}^2 a_s^p}$$

En general, $\hat{\mu}^{VAN} = \frac{\sum_{s=1}^m y_s^p}{\sum_{s=1}^m a_s^p}$

Ahora, aplicando la ecuación (3.4)

$$V(\hat{\mu}) = [A' \Sigma^{-1} A]^{-1}$$

$$V(\hat{\mu}) = \left\{ \begin{bmatrix} a_1^p & a_2^p \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2 a_1^p} & 0 \\ 0 & \frac{1}{\sigma^2 a_2^p} \end{bmatrix} \begin{bmatrix} a_1^p \\ a_2^p \end{bmatrix} \right\}^{-1} = \left[\frac{a_1^p + a_2^p}{\sigma^2} \right]^{-1} = \frac{\sigma^2}{a_1^p + a_2^p} = \frac{\sigma^2}{\sum_{s=1}^2 a_s^p}$$

En general, $V(\hat{\mu}^{VAN}) = \frac{\sigma^2}{\sum_{s=1}^m a_s^p}$ y $v(\hat{\mu}^{VAN}) = \frac{\hat{\sigma}^2}{\sum_{s=1}^m a_s^p}$, donde $\hat{\sigma}^2 = \frac{\sum_{s=1}^m (y_s^p - a_s^p \hat{\mu})^2}{\sum_{s=1}^m a_s^p - 1}$

(Van Deusen, 2004).

Modelo media de razones (ecuación 3.15)

$$y_s^p = a_s^p \mu + \varepsilon 3_s, E(\varepsilon 3_s) = 0, E(\varepsilon 3_s, \varepsilon 3_s') = (a_s^p)^2 \sigma^2 \quad \forall \quad \varepsilon 3_s = \varepsilon 3_s'$$

Sean

$$Y = \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix}; A = \begin{bmatrix} a_1^p \\ a_2^p \end{bmatrix}; \Sigma = \begin{bmatrix} \sigma^2 (a_1^p)^2 & 0 \\ 0 & \sigma^2 (a_2^p)^2 \end{bmatrix}; \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma^2 (a_1^p)^2} & 0 \\ 0 & \frac{1}{\sigma^2 (a_2^p)^2} \end{bmatrix}$$

Aplicando la ecuación (3.3)

$$\hat{\mu} = [A' \Sigma^{-1} A]^{-1} A' \Sigma^{-1} Y$$

$$\hat{\mu} = \left\{ \begin{bmatrix} a_1^p & a_2^p \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2 (a_1^p)^2} & 0 \\ 0 & \frac{1}{\sigma^2 (a_2^p)^2} \end{bmatrix} \begin{bmatrix} a_1^p \\ a_2^p \end{bmatrix} \right\}^{-1} \left\{ \begin{bmatrix} a_1^p & a_2^p \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2 (a_1^p)^2} & 0 \\ 0 & \frac{1}{\sigma^2 (a_2^p)^2} \end{bmatrix} \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix} \right\}$$

$$= \left\{ \begin{bmatrix} \frac{1}{\sigma^2 a_1^p} & \frac{1}{\sigma^2 a_2^p} \end{bmatrix} \begin{bmatrix} a_1^p \\ a_2^p \end{bmatrix} \right\}^{-1} \left\{ \begin{bmatrix} \frac{1}{\sigma^2 a_1^p} & \frac{1}{\sigma^2 a_2^p} \end{bmatrix} \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix} \right\}$$

$$= \left\{ \left[\frac{1}{\sigma^2} + \frac{1}{\sigma^2} \right] \right\}^{-1} \left\{ \left[\frac{y_1^p}{\sigma^2 a_1^p} + \frac{y_2^p}{\sigma^2 a_2^p} \right] \right\} = \left[\frac{2}{\sigma^2} \right]^{-1} \frac{1}{\sigma^2} \left[\frac{y_1^p}{a_1^p} + \frac{y_2^p}{a_2^p} \right]$$

$$= \left[\frac{\sigma^2}{2} \right] \frac{1}{\sigma^2} \left[\frac{y_1^p}{a_1^p} + \frac{y_2^p}{a_2^p} \right] = \frac{1}{2} \left[\frac{y_1^p}{a_1^p} + \frac{y_2^p}{a_2^p} \right] = \frac{\frac{y_1^p}{a_1^p} + \frac{y_2^p}{a_2^p}}{2} = \frac{\sum_{s=1}^2 \frac{y_s^p}{a_s^p}}{2}$$

En general, $\hat{\mu}^{MR} = \frac{\sum_{s=1}^m \frac{y_s^p}{a_s^p}}{m} = \frac{\sum_{s=1}^m \dot{y}_s}{m}$, donde $\dot{y}_s = \frac{y_s^p}{a_s^p}$

Ahora, aplicando la ecuación (3.4)

$$V(\hat{\mu}) = [A' \Sigma^{-1} A]^{-1}$$

$$V(\hat{\mu}) = \left\{ \begin{bmatrix} a_1^p & a_2^p \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2 (a_1^p)^2} & 0 \\ 0 & \frac{1}{\sigma^2 (a_2^p)^2} \end{bmatrix} \begin{bmatrix} a_1^p \\ a_2^p \end{bmatrix} \right\}^{-1} = \left[\frac{2}{\sigma^2} \right]^{-1} = \frac{\sigma^2}{2}$$

En general, $V(\hat{\mu}^{MR}) = \frac{\sigma^2}{m}$ y $v(\hat{\mu}^{MR}) = \frac{\hat{\sigma}^2}{m}$, donde $\hat{\sigma}^2 = \frac{\sum_{s=1}^m (\dot{y}_s - \hat{\mu})^2}{m-1}$

Modelo media simple

$$y_s^p = \mu + \varepsilon 3_s, E(\varepsilon 3_s) = 0, E(\varepsilon 3_s, \varepsilon 3'_s) = \sigma^2 \quad \forall \quad \varepsilon 3_s = \varepsilon 3'_s$$

Sean

$$Y = \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix}; A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}; \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{bmatrix}$$

Aplicando la ecuación (3.3)

$$\hat{\mu} = [A' \Sigma^{-1} A]^{-1} A' \Sigma^{-1} Y$$

$$\hat{\mu} = \left\{ \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}^{-1} \left\{ \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{bmatrix} \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix} \right\}$$

$$\begin{aligned}
&= \left\{ \begin{bmatrix} \frac{1}{\sigma^2} & 1 \\ \frac{1}{\sigma^2} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}^{-1} \left\{ \begin{bmatrix} \frac{1}{\sigma^2} & 1 \\ \frac{1}{\sigma^2} & 1 \end{bmatrix} \begin{bmatrix} y_1^p \\ y_2^p \end{bmatrix} \right\} \\
&= \left\{ \begin{bmatrix} \frac{1}{\sigma^2} + \frac{1}{\sigma^2} \end{bmatrix} \right\}^{-1} \begin{bmatrix} \frac{y_1^p}{\sigma^2} + \frac{y_2^p}{\sigma^2} \end{bmatrix} = \begin{bmatrix} 2 \\ \sigma^2 \end{bmatrix}^{-1} \frac{1}{\sigma^2} [y_1^p + y_2^p] \\
&= \begin{bmatrix} \sigma^2 \\ 2 \end{bmatrix} \frac{1}{\sigma^2} [y_1^p + y_2^p] = \frac{[y_1^p + y_2^p]}{2}
\end{aligned}$$

En general,
$$\hat{\mu} = \frac{\sum_{s=1}^m y_s^p}{m} = \frac{\sum_{s=1}^m \frac{y_s}{a_0}}{m} = \frac{\sum_{s=1}^m y_s}{a_0 m}$$

Ahora, aplicando la ecuación (3.4)

$$V(\hat{\mu}) = [A' \Sigma^{-1} A]^{-1}$$

$$V(\hat{\mu}) = \left\{ \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}^{-1} = \left\{ \begin{bmatrix} \frac{1}{\sigma^2} & 1 \\ \frac{1}{\sigma^2} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}^{-1} = \left\{ \begin{bmatrix} \frac{1}{\sigma^2} + \frac{1}{\sigma^2} \end{bmatrix} \right\}^{-1} = \begin{bmatrix} 2 \\ \sigma^2 \end{bmatrix}^{-1} = \begin{bmatrix} \sigma^2 \\ 2 \end{bmatrix}$$

En general,
$$V(\hat{\mu}) = \frac{\sigma^2}{m} \text{ y } v(\hat{\mu}) = \frac{\hat{\sigma}^2}{m}, \text{ donde } \hat{\sigma}^2 = \frac{\sum_{s=1}^m \left(\frac{y_s}{a_0} - \hat{\mu} \right)^2}{m-1}$$

En consecuencia,

$$v(\hat{\mu}) = \frac{\hat{\sigma}^2}{m} = \frac{1}{m} \frac{\sum_{s=1}^m \left(\frac{y_s}{a_0} - \hat{\mu} \right)^2}{m-1} = \frac{1}{m} \frac{\sum_{s=1}^m \left(\frac{y_s}{a_0} - \frac{\sum_{s=1}^m y_s}{a_0 m} \right)^2}{m-1} = \frac{1}{m} \frac{\sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{m-1} = \frac{\sum_{s=1}^m \left(y_s - \frac{\sum_{s=1}^m y_s}{m} \right)^2}{a_0^2 m(m-1)}$$

ANEXO 3.2. DERIVACIÓN DE LA VARIANZA ESTIMADA DE RAZÓN A PARTIR DE LA VARIANZA ESTIMADA DE VAN DEUSEN

$$v(y_{ha}^{RM}) = \left(\frac{\bar{m}\bar{a} - a_0}{\bar{a}(m-1)} \right) \frac{\sum_{s=1}^m (y_s - a_s y_{ha})^2}{\sum_{s=1}^m a_s \left(\sum_{s=1}^m a_s - a_0 \right)} = \left(\frac{\frac{\sum_{s=1}^m a_s}{m} - a_0}{\frac{\sum_{s=1}^m a_s}{m} (m-1)} \right) \frac{\sum_{s=1}^m (y_s - a_s y_{ha})^2}{\sum_{s=1}^m a_s \left(\sum_{s=1}^m a_s - a_0 \right)}$$

$$= \frac{\sum_{s=1}^m a_s - a_0}{\left(\sum_{s=1}^m a_s - a_0 \right) \frac{\sum_{s=1}^m a_s}{m} \frac{\sum_{s=1}^m a_s}{m}} \frac{1}{m} \frac{\sum_{s=1}^m (y_s - a_s y_{ha})^2}{(m-1)} = \frac{\sum_{s=1}^m (y_s - a_s y_{ha})^2}{\bar{a}^2 m(m-1)}$$

Ya que, $\bar{a} = \sum_{s=1}^m a_s / m$

Si los conglomerados son completos, $\bar{a} = \sum_{s=1}^m a_0 / m = ma_0 / m = a_0$, entonces:

$$\frac{\bar{m}\bar{a} - a_0}{\bar{a}(m-1)} = \frac{\bar{m}\bar{a} - \bar{a}}{\bar{a}(m-1)} = \frac{\bar{a}(m-1)}{\bar{a}(m-1)} = 1$$

Si $\bar{a} = \frac{a_0}{2}$ y $m = 20$ entonces:

$$\frac{\bar{m}\bar{a} - a_0}{\bar{a}(m-1)} = \frac{20 \frac{a_0}{2} - a_0}{\frac{a_0}{2} (20-1)} = \frac{a_0(10-1)}{\frac{a_0}{2} (20-1)} = \frac{9}{9.5} = 0.9473$$

ANEXO 3.3. ESTIMACIONES PUNTUALES Y POR INTERVALO DE 88 MUESTRAS. VARIABLE DENSIDAD EN TRES CONDICIONES DIFERENTES.

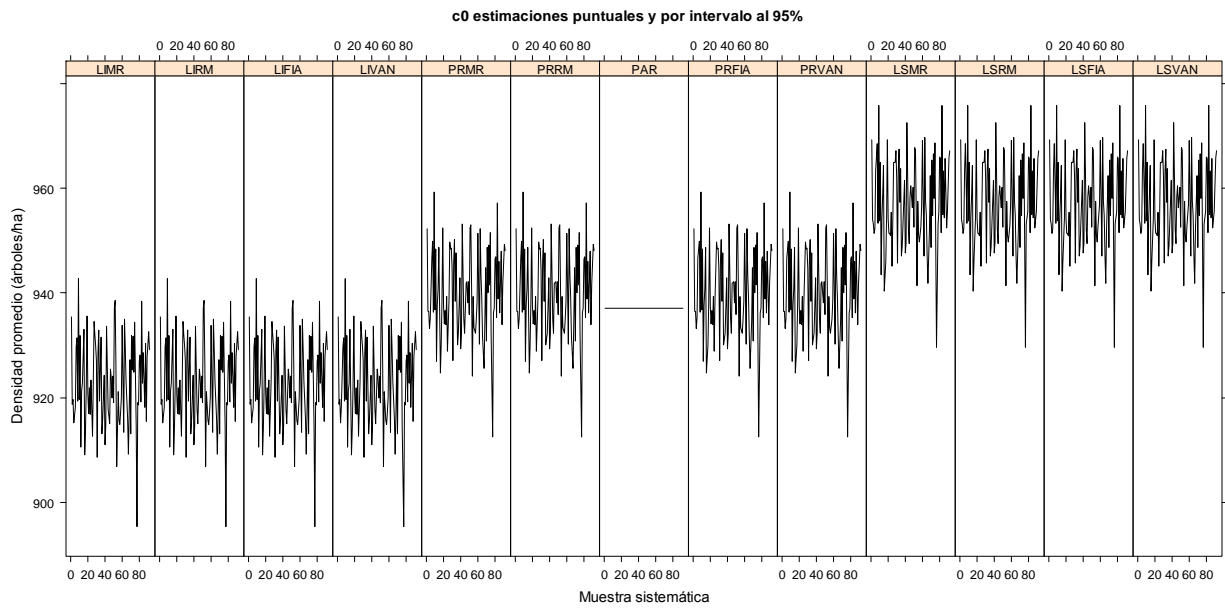


Figura 3.3.1. Estimaciones puntuales y por intervalo (95%) de 88 muestras. Variable densidad en condición c0.

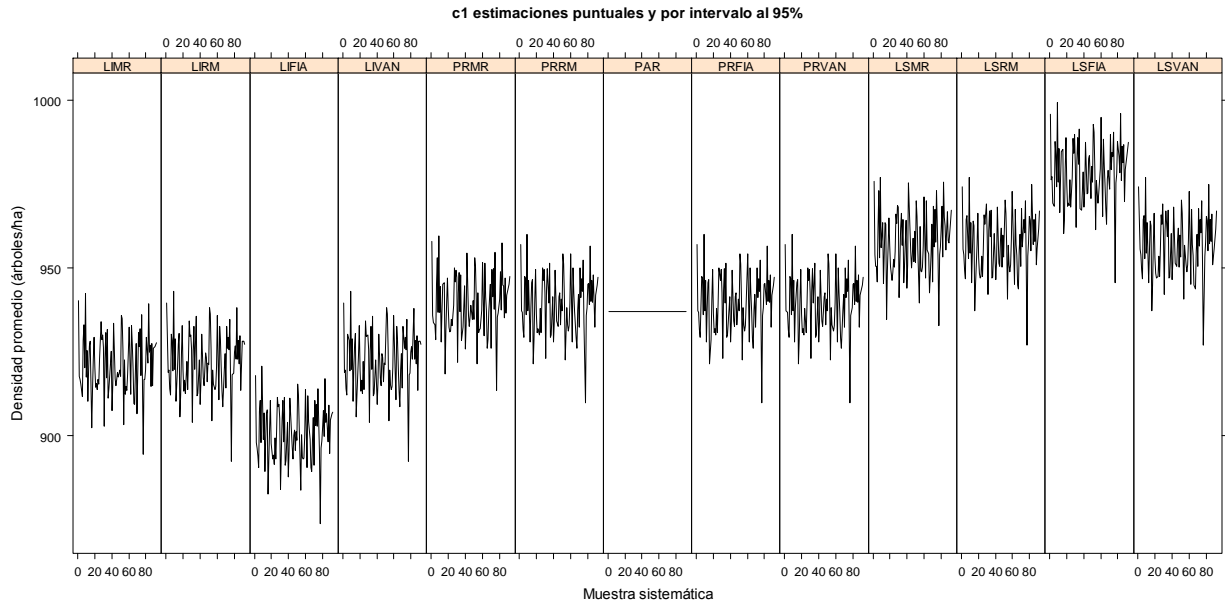


Figura 3.3.2. Estimaciones puntuales y por intervalo (95%) de 88 muestras. Variable densidad en condición c1.

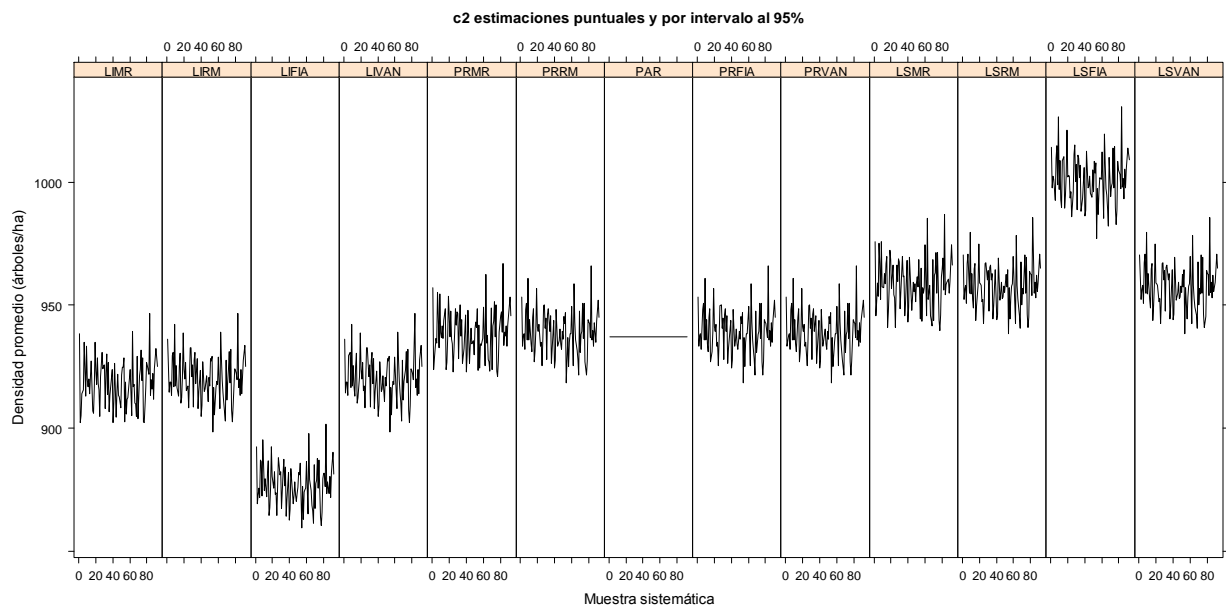


Figura 3.3.3. Estimaciones puntuales y por intervalo (95%) de 88 muestras. Variable densidad en condición c2.

ANEXO 3.4. ESTIMACIONES PUNTUALES Y POR INTERVALO DE 88 MUESTRAS. VARIABLE VOLUMEN EN TRES CONDICIONES DIFERENTES.

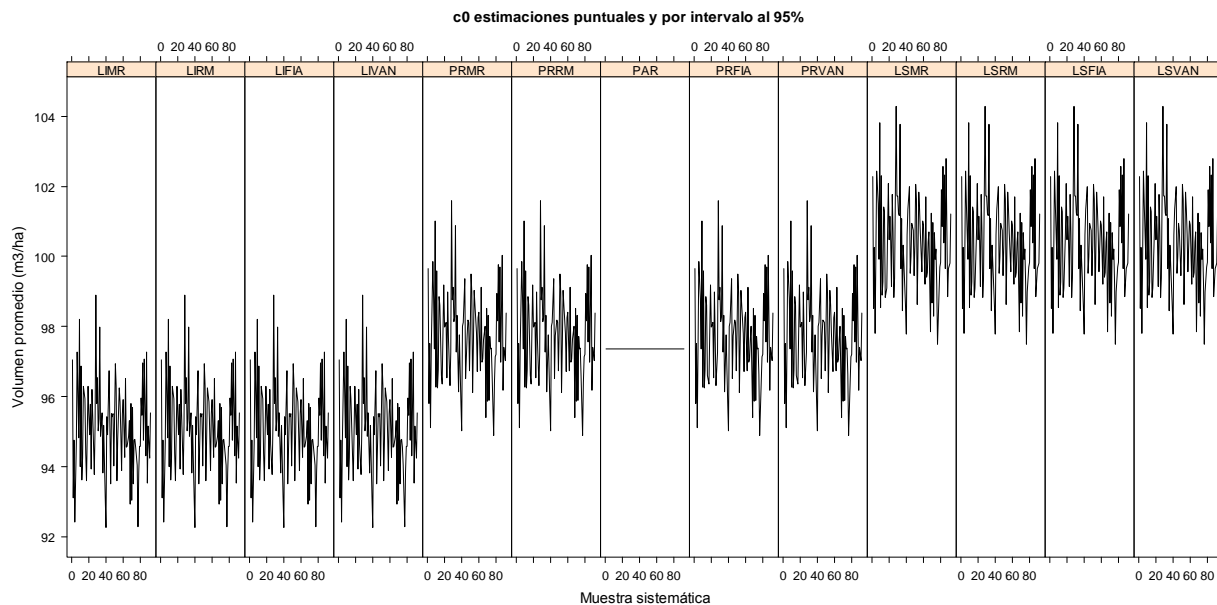


Figura 3.4.1. Estimaciones puntuales y por intervalo (95%) de 88 muestras. Variable volumen en condición c0.

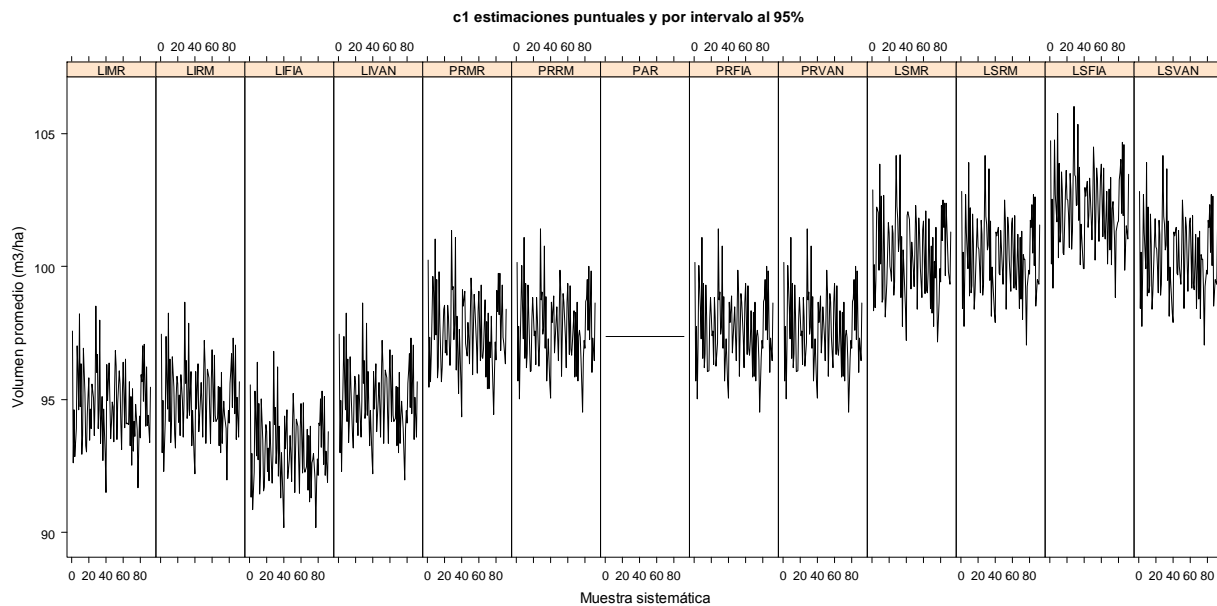


Figura 3.4.2. Estimaciones puntuales y por intervalo (95%) de 88 muestras. Variable volumen en condición c1.

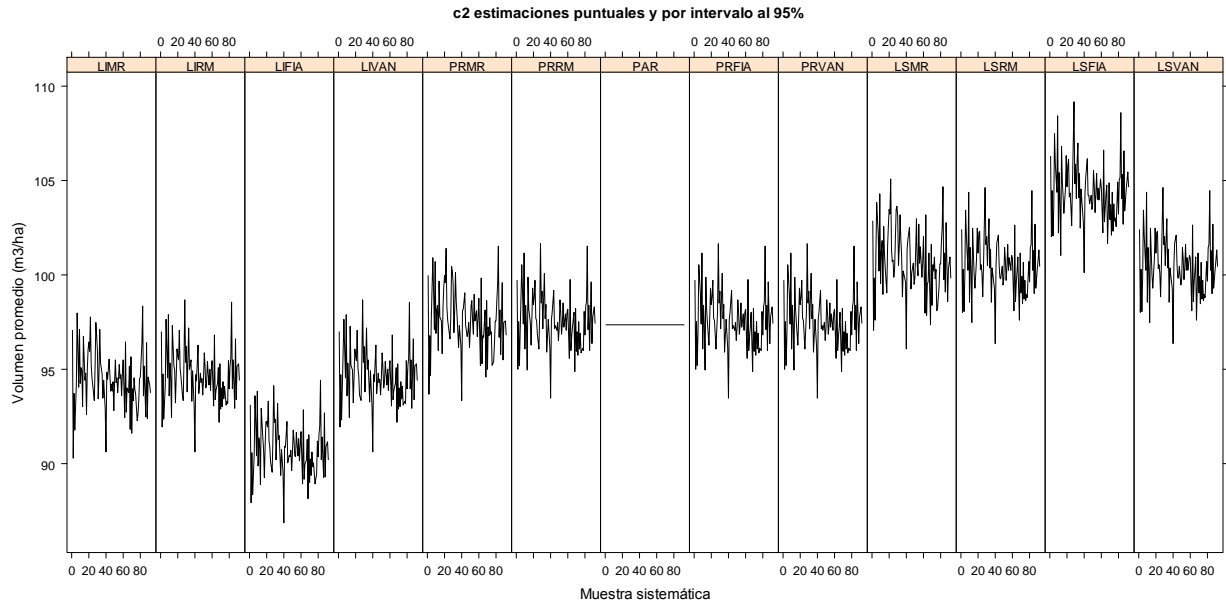


Figura 3.4.3. Estimaciones puntuales y por intervalo (95%) de 88 muestras. Variable volumen en condición c2.

CAPÍTULO IV

ESTIMACIÓN DE PARÁMETROS FORESTALES EN EL CONTEXTO DE DISEÑO DE MUESTRAS PANELIZADAS DEL INVENTARIO NACIONAL FORESTAL Y DE SUELOS DE MÉXICO (INFyS)

4.1. RESUMEN

La estrategia de remediación del INFyS corresponde a un muestreo de panel rotante o muestras panelizadas, en el cual teóricamente cada año del segundo ciclo se evalúa el 20% de la muestra total estudiada durante el primer ciclo. La medición de las mismas unidades de muestreo en muestras panelizadas produce correlación temporal que debería ser considerada en el proceso de estimación. Actualmente en México no se han investigado los posibles métodos estadísticos para analizar datos de inventario cuando se tiene un muestreo panelizado. Con el propósito de explorar la mejor técnica de análisis estadístico, bajo un diseño de muestreo de panel rotante, se estudiaron las propiedades estadísticas teóricas de los estimadores Mínimos Cuadrados Generalizados (MCG), Muestreo Replicado (MRE) y Estimador Compuesto Univariado (ECU). Se consideró como parámetro el promedio de área basal por hectárea en cada una de las siguientes ocasiones de medición: 2004, 2005, 2006, 2007, 2009, 2010, 2011 y 2012. También se consideró como parámetro la tasa de cambio de área basal por hectárea entre puntos medios de periodos; el primer periodo fue de 2004 a 2007, y el segundo, de 2009 a 2012. Para llevar a cabo el estudio se generó una pseudo-población forestal artificial de 7,597 conglomerados panelizados. Esta pseudo-población se integró al combinar las bases de datos de Durango, Quintana Roo, Hidalgo, Jalisco, Estado de México, Chihuahua y Michoacán. Los resultados indican que aunque el estimador MCG es ligeramente sesgado, comparado con MRE, presenta menor varianza en relación a MRE. El ECU generó en cada una de las ocasiones de medición un sesgo relativo intermedio entre MCG y MRE, y presentó menor varianza comparado con MRE. En estas condiciones ECU puede considerarse como un estimador insesgado (o cuasinsesgado) de mínima varianza. En la ocasión más reciente, el estimador Media Móvil (MM) con pesos iguales, caracterizado por presentar varianza pequeña, presentó un sesgo de casi el doble que MCG. En la estimación de la tasa de cambio de área basal por hectárea entre puntos medios de periodo, el ECU también tiene como propiedad reducir el sesgo de MCG y la varianza de MRE.

Palabras clave: Muestreo de panel rotante, Mínimos Cuadrados Generalizados, Muestreo Replicado y Estimador Compuesto Univariado.

ESTIMATION OF FOREST PARAMETERS IN THE CONTEXT OF THE PANELIZED SAMPLES DESIGN OF THE NATIONAL FOREST AND SOILS INVENTORY OF MEXICO (INFyS)

4.2. SUMMARY

The re-measurement INFyS strategy corresponds to a rotating panel sampling also known as panelized samples, in which, theoretically, every year of the second cycle evaluates 20% of the total sample studied during the first cycle. Measuring the same sampling units in panelized samples produces temporary correlation that should be considered in the estimation process. Nowadays in Mexico the statistical methods to analyze inventory data when there is a panelized sampling design have not yet been fully investigated. To explore the best technique for statistical analysis, under a sampling design of rotating panel, the theoretical statistical properties of the estimators Generalized Least Squares (GLS), Replicated Sampling (RES) and the Univariate Composite Estimator (UCE) were studied. The parameter of interest chosen was the average basal area per hectare in each of the following measurement occasions: 2004, 2005, 2006, 2007, 2009, 2010, 2011 and 2012. The change rate of basal area per hectare between midpoints of periods was also considered as a parameter; the first period was from 2004 to 2007 and the second from 2009 to 2012. To carry out the study an artificial forest pseudo-poblacion of 7,597 panelized clusters was generated by combining INFyS datasets of Durango, Quintana Roo, Hidalgo, Jalisco, State of Mexico, Chihuahua and Michoacan. Results indicate that even though the GLS estimator is slightly biased, compared with RES, has a smaller variance in relation to RES. UCE generated in each measurement occasions an intermediate relative bias between GLS and RES, and presented a smaller variance compared to RES. In these conditions UCE can be considered an unbiased estimator (or quasiunbiased) with minimum variance. In the most recent occasion, the Moving Average estimator (MA) with equal weights, characterized by introducing small variance, presented a bias almost double than the GLS estimate. In the estimation of the rate of change of basal area per hectare between mid-points of the period, the UCE also has a reduced GLS bias and the variance of RES.

Key words: Rotating panel sampling, Generalized Least Squares (GLS), Replicated Sampling (RES) and the Univariate Composite Estimator (UCE).

4.3. INTRODUCCIÓN

El INFyS, vigente en México, lleva dos evaluaciones en todos los ecosistemas forestales del país, el primer ciclo de medición se implementó de 2004 a 2007 (20,791 conglomerados), y el segundo, comprendió de 2009 a 2013 (meta 19,999 conglomerados, 20% anualmente). Esto quiere decir que en cada año del segundo ciclo teóricamente se evalúa el 20% de la muestra total estudiada en cada año del primer ciclo. Para cierta población de interés, por ejemplo los bosques de México o bosques del estado de Durango, la estructura espacial y temporal de las mediciones (evaluaciones diferidas en el primer ciclo y anualidades parciales en el segundo) conduce al concepto de muestras panelizadas. La medición de las mismas unidades en estudios de muestras panelizadas trae consigo correlación temporal, misma que debe ser considerada en el proceso de estimación.

El diseño de muestras panelizadas del INFyS conceptualmente se puede estudiar mediante la teoría del muestreo con reemplazo parcial en múltiples ocasiones, el cual es una extensión del muestreo con reemplazo parcial en dos ocasiones.

La estimación de los parámetros forestales en los años más recientes y la estimación del cambio de tales parámetros a través del tiempo son de interés en un sistema de evaluación y monitoreo forestal. Para tal efecto se pueden emplear por lo menos dos métodos, el primero, llamado estimación por Mínimos Cuadrados Generalizados (MCG) considera la estructura de correlación temporal producto de la remediación de los subpaneles, el segundo por su parte, se basa en ignorar dicha correlación y aplicar directamente en cada año los estimadores del Muestreo Replicado (MRE). MCG se caracteriza por combinar directamente la información de todos los años, por lo que la estimación para el año actual considera de antemano la información previa. El estimador MCG se puede usar para el muestreo con reemplazamiento parcial en dos y en múltiples ocasiones. Para la estimación del año actual, MRE puede considerar sólo la estimación de este año, o bien, combinar la información de los años anteriores y el actual mediante un estimador de Media Móvil (MM). Aun cuando MM conduce a estimaciones muy sesgadas para ocasión más reciente, resulta ser de varianza mínima, por esta razón es

el método usado en el Forest Inventory and Analysis (FIA) de los Estados Unidos de América. La combinación de MCG y MRE genera el ECU con propiedades estadísticas deseables.

Con base en lo anterior, en esta etapa de la investigación se estudiaron desde el punto de vista teórico las propiedades estadísticas de los estimadores MCG, MRE, ECU y MM. Para tal efecto se generó una pseudo-población forestal artificial con 7,597 conglomerados panelizados, la cual estuvo integrada con bases de datos del INFyS correspondientes a Durango, Quintana Roo, Hidalgo, Jalisco, Estado de México, Chihuahua y Michoacán.

4.4. OBJETIVOS

Estudiar desde el punto de vista teórico técnicas estadísticas para la estimación de los principales parámetros de interés forestal que consideren la estructura actual del sistema de muestreo panelizado del INFyS.

Evaluar el desempeño estadístico de estas técnicas en una pseudo-población forestal artificial.

4.5. REVISIÓN DE LITERATURA

De acuerdo con Schreuder *et al.* (2004), existen tres opciones principales de muestreo para considerar el muestreo en el tiempo.

1. Muestreo con reemplazo total (MRT)
2. Muestreo con remediación total (MRMT)
3. Muestreo con reemplazo parcial (MRP)

En el MRT, en cada ocasión se tiene un conjunto de parcelas de muestreo completamente nuevo. Este diseño es simple y barato de implementar, ya que la ubicación de las parcelas no es permanente. El MRT es eficiente para estimar los parámetros actuales, pero no lo es para estimar cambios, en comparación con el MRMT y el MRP.

En MRMT todas las parcelas se remiden periódicamente. Esto requiere que esas parcelas sigan siendo representativas de la población en el tiempo. Este muestreo es el más eficiente de los métodos disponibles para la estimación de los cambios.

En el muestreo con reemplazo parcial se mide un subconjunto de las parcelas medidas inicialmente y también un subconjunto nuevo de parcelas, esto es, MRP es una combinación de los dos tipos de muestreo anteriores. La estimación de regresión entre las parcelas remedidas se utiliza para actualizar las parcelas que no se remidieron. Este muestreo puede ser eficiente cuando se requiere balancear la precisión entre la estimación del año más reciente y la estimación del cambio.

El muestreo con remediación completa y el muestreo con reemplazo total son casos particulares del muestreo con reemplazo parcial desde el punto de vista de la estimación.

La aplicación más sencilla cuando se implementan estos tipos de muestreo es cuando se tienen únicamente dos mediciones en el tiempo. Aun cuando este esquema de monitoreo tienen limitaciones desde el punto de vista práctico, teóricamente es la base para estudios longitudinales de mayor complejidad.

4.5.1. Estimadores del muestreo con reemplazo total y remediación total

Sea t_1 y t_2 la primera y la segunda ocasión de muestreo, respectivamente. De manera análoga, permita que m_1 y m_2 indiquen el número de parcelas de muestreo en t_1 y t_2 , respectivamente. Además, sea $T_y(t_1)$ el total de la población o estrato de y en la primer ocasión y $T_y(t_2)$ el total en el segunda ocasión.

De acuerdo con Gregoire and Valentine (2008), el parámetro para el cambio es

$$\Delta_y = T_y(t_2) - T_y(t_1) \quad (4.1)$$

Si las estimaciones de $T_y(t_1)$ y $T_y(t_2)$ son dadas mediante Muestreo Replicado (ecuación (2.8a), Capítulo II), es decir.

$$\hat{T}_{y\pi,rep}(t_1) = \frac{1}{m_1} \sum_{s=1}^{m_1} \hat{T}_{y\pi s}(t_1) \quad (4.2)$$

y

$$\hat{T}_{y\pi,rep}(t_2) = \frac{1}{m_2} \sum_{s=1}^{m_2} \hat{T}_{y\pi s}(t_2) \quad (4.3)$$

y sus respectivas varianzas estimadas son

$$v(\hat{T}_{y\pi,rep}(t_1)) = \frac{1}{m_1(m_1-1)} \sum_{s=1}^{m_1} \left(\hat{T}_{y\pi s}(t_1) - \hat{T}_{y\pi,rep}(t_1) \right)^2 \quad (4.4)$$

y

$$v(\hat{T}_{y\pi,rep}(t_2)) = \frac{1}{m_2(m_2-1)} \sum_{s=1}^{m_2} \left(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi,rep}(t_2) \right)^2 \quad (4.5)$$

Entonces, un estimador natural de Δ_y es

$$\hat{\Delta}_y = \hat{T}_{y\pi,rep}(t_2) - \hat{T}_{y\pi,rep}(t_1) \quad (4.6)$$

Este estimador de cambio es muy general: no hay supuesto que las mismas parcelas usadas en la ocasión t_1 sean también usadas en la ocasión t_2 , y no hay supuesto que el mismo número de parcelas es usado en ambas ocasiones, ni que los mismos elementos de la población son medidos en ambas ocasiones. Sin embargo, si las parcelas muestreadas en t_1 son otra vez visitadas en t_2 , entonces $\hat{\Delta}_y$ en (4.6) puede ser expresado alternativamente como

$$\hat{\Delta}_y = \hat{T}_{y\pi,rep}(t_2) - \hat{T}_{y\pi,rep}(t_1) = \frac{1}{m} \sum_{s=1}^m \left(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi s}(t_1) \right) \quad (4.7)$$

Como ambos $\hat{T}_{y\pi,rep}(t_1)$ y $\hat{T}_{y\pi,rep}(t_2)$ son insesgados, entonces $\hat{\Delta}_y$ también es insesgado.

Su varianza es

$$V(\hat{\Delta}_y) = V(\hat{T}_{y\pi,rep}(t_2)) + V(\hat{T}_{y\pi,rep}(t_1)) - 2Cov(\hat{T}_{y\pi,rep}(t_2), \hat{T}_{y\pi,rep}(t_1)) \quad (4.8)$$

Donde $Cov(\hat{T}_{y\pi,rep}(t_2), \hat{T}_{y\pi,rep}(t_1))$ es la covarianza entre los dos estimadores. Si las parcelas muestrales son seleccionadas independientemente en t_1 y t_2 , entonces esta covarianza será cero, conduciendo al resultado bien conocido que la varianza entre dos estimadores independientes es la suma de sus respectivas varianzas. Cuando las mismas parcelas son usadas en ambas ocasiones, $\hat{T}_{y\pi,rep}(t_2)$ generalmente es correlacionado de manera positiva con $\hat{T}_{y\pi,rep}(t_1)$, aunque la fuerza de la correlación usualmente disminuye conforme el intervalo de tiempo entre t_1 y t_2 se alarga. Sin embargo, cualquier correlación positiva implica una covarianza positiva, de manera que se espera que la varianza de $\hat{\Delta}_y$ sea más pequeña si las mismas parcelas son usadas en dos ocasiones. Para ponerlo de otra manera, se espera que $\hat{\Delta}_y$ sea un estimador más preciso de Δ_y si las mismas parcelas son medidas en ambas ocasiones.

Un estimador natural de $V(\hat{\Delta}_y)$ es

$$v(\hat{\Delta}_y) = v(\hat{T}_{y\pi,rep}(t_2)) + v(\hat{T}_{y\pi,rep}(t_1)) - 2cov(\hat{T}_{y\pi,rep}(t_2), \hat{T}_{y\pi,rep}(t_1)) \quad (4.9)$$

donde $v(\hat{T}_{y\pi,rep}(t_1))$ y $v(\hat{T}_{y\pi,rep}(t_2))$ son dadas por (4.4) y (4.5), respectivamente, y

$$cov(\hat{T}_{y\pi,rep}(t_2), \hat{T}_{y\pi,rep}(t_1)) = \frac{1}{m(m-1)} \sum_{s=1}^m \left(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi,rep}(t_2) \right) \left(\hat{T}_{y\pi s}(t_1) - \hat{T}_{y\pi,rep}(t_1) \right) \quad \text{es un}$$

estimador insesgado de $Cov(\hat{T}_{y\pi,rep}(t_2), \hat{T}_{y\pi,rep}(t_1))$.

4.5.2. Muestreo con reemplazo parcial en dos ocasiones mediante estimadores de regresión

Además de la notación anterior considere la siguiente (Scott, 1984; Gregoire and Valentine, 2008):

m_{12} : Núm. de parcelas muestreadas en ambas ocasiones (parcelas permanentes).

m_1 : Núm. de parcelas muestreadas en t_1 pero no en t_2 (parcelas temporales en t_1).

m_2 : Núm. de parcelas muestreadas en t_2 pero no en t_1 (parcelas temporales t_2).

Entonces, $m_1 = m_1 + m_{12}$ y $m_2 = m_2 + m_{12}$. Si $m_{12} = 0 \Rightarrow m_1 = m_1$ y $m_2 = m_2 \Rightarrow MRT$. Además, si $m_1 = 0 \Rightarrow m_1 = m_{12}$ y si $m_2 = 0 \Rightarrow m_2 = m_{12} \Rightarrow MRMT$

En la Figura 4.1 se presenta esquemáticamente el muestreo con reemplazamiento parcial en dos ocasiones. Para fines teóricos suponga que t_2 es la segunda ocasión de medición.

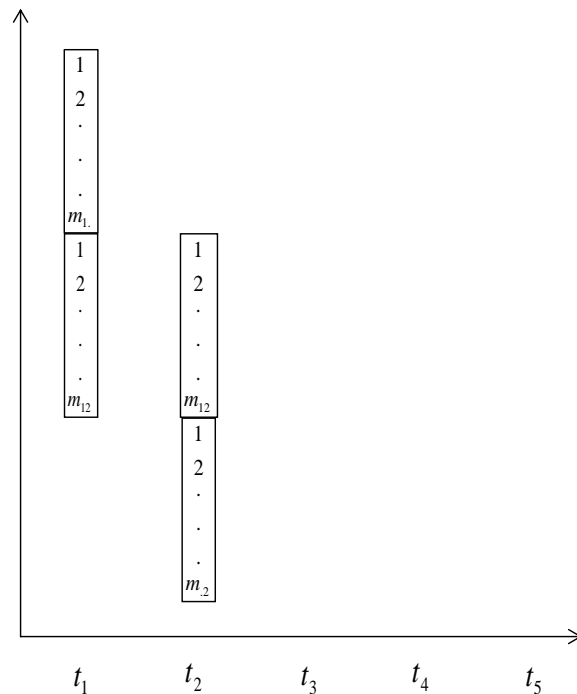


Figura 4.1. Muestreo con reemplazo parcial en dos ocasiones (t_1, t_2). m_{12} se miden en t_1 y t_2 , m_1 se miden solo en t_1 y m_2 se miden solo en t_2 .

En esta situación se distinguen cuatro estimadores. Dos de ellos basados solamente en parcelas permanentes:

Estimación de $T_y(t_1)$

$$\hat{T}_{y\pi,rep}(t_1, m_{12}) = \frac{1}{m_{12}} \sum_{s=1}^{m_{12}} \hat{T}_{y\pi s}(t_1) \quad (4.10)$$

Estimación de $T_y(t_2)$

$$\hat{T}_{y\pi,rep}(t_2, m_{12}) = \frac{1}{m_{12}} \sum_{s=1}^{m_{12}} \hat{T}_{y\pi s}(t_2) \quad (4.11)$$

Los otros dos estimadores se basan en parcelas temporales:

Estimación de $T_y(t_1)$

$$\hat{T}_{y\pi,rep}(t_1, m_{1.}) = \frac{1}{m_{1.}} \sum_{s=1}^{m_{1.}} \hat{T}_{y\pi s}(t_1) \quad (4.12)$$

Estimación de $T_y(t_2)$

$$\hat{T}_{y\pi,rep}(t_2, m_{.2}) = \frac{1}{m_{.2}} \sum_{s=1}^{m_{.2}} \hat{T}_{y\pi s}(t_2) \quad (4.13)$$

4.5.2.1. Estimador compuesto de $T_y(t_2)$

Tratando t_2 como la ocasión de interés principal se busca una combinación lineal de estos cuatro estimadores que estimen insesgadamente a $T_y(t_2)$. Esto es, se buscan valores de los coeficientes a, b, c y d para formar un estimador compuesto tal que

$$\begin{aligned} \hat{T}_{y,MRP}(t_2 : t_1) &= a\hat{T}_{y\pi,rep}(t_1, m_{1.}) + b\hat{T}_{y\pi,rep}(t_1, m_{12}) \\ &\quad + c\hat{T}_{y\pi,rep}(t_2, m_{.2}) + d\hat{T}_{y\pi,rep}(t_2, m_{12}) \end{aligned} \quad (4.14)$$

mientras se asegura que $E[\hat{T}_{y,MRP}(t_2 : t_1)] = T_y(t_2)$.

Hay redundancia en los cuatro coeficientes de (4.14) debido a que

$$E[\hat{T}_{y\pi,rep}(t_1, m_{1.})] = E[\hat{T}_{y\pi,rep}(t_1, m_{12})] = T_y(t_1)$$

Implica que $a+b=0$, y

$$E[\hat{T}_{y\pi,rep}(t_2, m_{2.})] = E[\hat{T}_{y\pi,rep}(t_2, m_{12})] = T_y(t_2)$$

implica que $c+d=1$. En consecuencia, (4.14) se simplifica a

$$\begin{aligned} \hat{T}_{y,MRP}(t_2 : t_1) &= a\hat{T}_{y\pi,rep}(t_1, m_{1.}) - a\hat{T}_{y\pi,rep}(t_1, m_{12}) \\ &\quad + (1-d)\hat{T}_{y\pi,rep}(t_2, m_{2.}) + d\hat{T}_{y\pi,rep}(t_2, m_{12}) \end{aligned} \quad (4.15)$$

Para identificar de forma exclusiva los valores de a y d en (4.15) una restricción adicional es impuesta de manera que los valores de a y d deben ser aquellos que minimicen la varianza de $\hat{T}_{y,MRP}(t_2 : t_1)$ dados $m_{1.}$, $m_{2.}$ y m_{12} . Dado que la covarianza entre los términos en (4.15) es cero para todos excepto los términos que contienen

$\hat{T}_{y\pi,rep}(t_1, m_{12})$ y $\hat{T}_{y\pi,rep}(t_2, m_{12})$, la varianza de $\hat{T}_{y,MRP}(t_2 : t_1)$ es

$$\begin{aligned} V(\hat{T}_{y,MRP}(t_2 : t_1)) &= a^2(V(\hat{T}_{y\pi,rep}(t_1, m_{1.})) + V(\hat{T}_{y\pi,rep}(t_1, m_{12}))) \\ &\quad + (1-d)^2V(\hat{T}_{y\pi,rep}(t_2, m_{2.})) + d^2V(\hat{T}_{y\pi,rep}(t_2, m_{12})) \\ &\quad - 2adCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) \end{aligned} \quad (4.16)$$

En (4.16), $Cov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12}))$ es la covarianza entre los estimadores basados en las parcelas permanentes, es decir,

$$Cov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) = \rho_{12} \sqrt{V(\hat{T}_{y\pi,rep}(t_1, m_{12}))} \sqrt{V(\hat{T}_{y\pi,rep}(t_2, m_{12}))} \quad (4.17)$$

Donde ρ_{12} es el coeficiente de correlación entre $\hat{T}_{y\pi,rep}(t_1, m_{12})$ y $\hat{T}_{y\pi,rep}(t_2, m_{12})$.

La imposición de la restricción produce el siguiente resultado (ANEXO 4.1):

$$d = \frac{V(\hat{T}_{y\pi,rep}(t_2, m_{12}))}{V(\hat{T}_{y\pi,rep}(t_2, m_{12})) + V(\hat{T}_{y\pi,rep}(t_1, m_{12})) - \gamma_{12}} \quad (4.18)$$

Donde

$$\gamma_{12} = \frac{\left(\text{Cov}(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) \right)^2}{V(\hat{T}_{y\pi,rep}(t_1, m_{12})) + V(\hat{T}_{y\pi,rep}(t_2, m_{12}))} \quad (4.19)$$

El cual se simplifica a

$$d = \frac{m_{12}}{m_2 - u_1 m_{12} \rho_{12}^2} \quad (4.20)$$

y en consecuencia,

$$a = \left(\frac{m_{12} u_1}{m_2 - u_1 m_{12} \rho_{12}^2} \right) \rho_{12} \sqrt{\frac{\phi_2}{\phi_1}} = \left(\frac{m_{12}}{m_2 - u_1 m_{12} \rho_{12}^2} \right) u_1 \rho_{12} \sqrt{\frac{\phi_2}{\phi_1}} = du_1 B_{21} \quad (4.21)$$

Donde

$$B_{21} = \rho_{12} \sqrt{\frac{\phi_2}{\phi_1}}, \quad u_1 = \frac{m_{12}}{m_1}, \quad \phi_1 \text{ y } \phi_2 \text{ son definidas en el ANEXO 4.1.}$$

Sustituyendo a y los resultados del ANEXO 4.1 (refiérase a la sección donde $\hat{T}_{y\pi,rep}(t_1)$ es expresado como un promedio ponderado) en (4.15) se obtiene

$$\hat{T}_{y,MRP}(t_2 : t_1) = d \hat{T}_{y\pi,reg}(t_2 : t_1) + (1 - d) \hat{T}_{y\pi,rep}(t_2, m_{12}) \quad (4.22)$$

Donde

$$\hat{T}_{y\pi,reg}(t_2 : t_1) = \hat{T}_{y\pi,rep}(t_2, m_{12}) + B_{21}(\hat{T}_{y\pi,rep}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12})) \quad (4.23)$$

Este resultado es algo sorprendente: el estimador óptimo de $T_y(t_2)$ para tamaños de muestras especificados m_1 , m_{12} y m_2 es una combinación del estimador de regresión basado en datos remedidos y el estimador de muestreo replicado basado en las m_2 unidades de muestreo nuevas. Éste último está basado en la información actual solamente, mientras que el primero está basado en la información previa, la cual es ajustada a través de regresión a la ocasión presente. Para el propósito de estimar $T_y(t_2)$, $\hat{T}_{y,MRP}(t_2 : t_1)$ en (4.22) es más preciso que el estimador de muestreo replicado basado en todas las $m_2 = m_2 + m_{12}$ unidades de muestreo disponibles en t_2 . Una inspección de (4.22) revela que $\hat{T}_{y,MRP}(t_2 : t_1)$ colapsa con $\hat{T}_{y\pi,rep}(t_2)$ cuando $\rho_{12} = 0$.

Ware 1962 (citado por Gregoire y Valentine, 2008) mostró que:

$$d = \frac{w(t_2 : t_1)}{w_{MRP}} \quad (4.24)$$

y

$$1 - d = \frac{w(t_2)}{w_{MRP}} \quad (4.25)$$

Donde

$$w(t_2 : t_1) = \frac{1}{V(\hat{T}_{y\pi,reg}(t_2 : t_1))} \quad (4.26)$$

$$w(t_2) = \frac{1}{V(\hat{T}_{y\pi,rep}(t_2, m_2))} \quad (4.27)$$

y

$$w_{MRP} = w(t_2 : t_1) + w(t_2) \quad (4.28)$$

Es decir,

$$w_{\text{MRP}} = \frac{1}{V(\hat{T}_{y\pi, \text{reg}}(t_2 : t_1))} + \frac{1}{V(\hat{T}_{y\pi, \text{reg}}(t_2, m.2))} = \frac{V(\hat{T}_{y\pi, \text{reg}}(t_2, m.2)) + V(\hat{T}_{y\pi, \text{reg}}(t_2 : t_1))}{V(\hat{T}_{y\pi, \text{reg}}(t_2 : t_1))V(\hat{T}_{y\pi, \text{reg}}(t_2, m.2))} \quad (4.28')$$

y

$$\frac{1}{w_{\text{MRP}}} = \frac{V(\hat{T}_{y\pi, \text{reg}}(t_2 : t_1))V(\hat{T}_{y\pi, \text{reg}}(t_2, m.2))}{V(\hat{T}_{y\pi, \text{reg}}(t_2, m.2)) + V(\hat{T}_{y\pi, \text{reg}}(t_2 : t_1))} \quad (4.29)$$

Empleando (4.24) y (4.25) se observa que (4.22) puede expresarse de la siguiente manera:

$$\hat{T}_{y, \text{MRP}}(t_2 : t_1) = \frac{w(t_2 : t_1)\hat{T}_{y\pi, \text{reg}}(t_2 : t_1) + w(t_2)\hat{T}_{y\pi, \text{reg}}(t_2, m.2)}{w_{\text{MRP}}} \quad (4.30)$$

En realidad (4.30) es un estimador ponderado, en el que se usan los recíprocos de las varianzas como los pesos (Scott, 1984; Cochran, 1993, página 424). Este método conduce a una expresión sencilla de la varianza, la cual depende de las varianzas de los estimadores que se combinan. Así la varianza de $\hat{T}_{y, \text{MRP}}(t_2 : t_1)$ se expresa simplemente de la siguiente manera (ANEXO 4.1).

$$V(\hat{T}_{y, \text{MRP}}(t_2 : t_1)) = \frac{1}{w_{\text{MRP}}} \quad (4.31)$$

En la medida en que ni B_{21} ni la varianza en los pesos $w(t_2 : t_1)$ y $w(t_2)$ son conocidos, $\hat{T}_{y, \text{MRP}}(t_2 : t_1)$ no es factible. En la práctica, estimaciones de estos parámetros son insertados en el estimador compuesto $\hat{T}_{y, \text{MRP}}(t_2 : t_1)$, produciendo así un estimador de $T_y(t_2)$ con propiedades estadísticas ligeramente diferentes. Se introduce $\hat{\hat{T}}_{y, \text{MRP}}(t_2 : t_1)$ para denotar el estimador compuesto **factible**, es decir, un estimador compuesto con estimaciones de B_{21} , $w(t_2 : t_1)$ y $w(t_2)$:

$$\hat{T}_{y,\text{MRP}}(t_2:t_1) = \frac{\hat{w}(t_2:t_1)\hat{T}_{y\pi,\text{reg}}(t_2:t_1) + \hat{w}(t_2)\hat{T}_{y\pi,\text{rep}}(t_2, m_{12})}{\hat{w}_{\text{MRP}}} \quad (4.32)$$

Donde

$$\hat{T}_{y\pi,\text{reg}}(t_2:t_1) = \hat{T}_{y\pi,\text{rep}}(t_2, m_{12}) + \hat{B}_{21}(\hat{T}_{y\pi,\text{rep}}(t_1) - \hat{T}_{y\pi,\text{rep}}(t_1, m_{12})) \quad (4.33)$$

$$\hat{w}_{\text{MRP}} = \hat{w}(t_2:t_1) + \hat{w}(t_2) \quad (4.34)$$

En general, la insesgadez de $\hat{T}_{y,\text{MRP}}(t_2:t_1)$ no se mantiene para $\hat{T}_{y,\text{MRP}}(t_2:t_1)$, sin embargo, con muestras grandes, la magnitud del sesgo debería ser despreciable.

El estimador de B_{21} que es usado en $\hat{T}_{y,\text{MRP}}(t_2:t_1)$ está basado en las m_{12} unidades emparejadas en la muestra. Suponiendo muestreo aleatorio simple, este estimador se expresa de la siguiente manera:

$$\hat{B}_{21} = \frac{s_y(t_2:t_1)}{s_y^2(t_1, m_{12})} \quad (4.35)$$

$$s_y(t_2:t_1) = \frac{1}{m_{12}-1} \sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi\mathcal{S}}(t_1) - \hat{T}_{y\pi,\text{rep}}(t_1, m_{12}) \right) \left(\hat{T}_{y\pi\mathcal{S}}(t_2) - \hat{T}_{y\pi,\text{rep}}(t_2, m_{12}) \right) \quad (4.36)$$

$$s_y^2(t_1, m_{12}) = \frac{1}{m_{12}-1} \sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi\mathcal{S}}(t_1) - \hat{T}_{y\pi,\text{rep}}(t_1, m_{12}) \right)^2 \quad (4.37)$$

Es necesario hacer hincapié que tanto el numerador como el denominador de (4.35) están basados en las parcelas remedidas solamente.

Note que $\hat{w}(t_2:t_1)$ se expresa como

$$\hat{w}(t_2:t_1) = \frac{1}{v(\hat{T}_{y\pi,\text{reg}}(t_2:t_1))} \quad (4.38)$$

Donde $v(\hat{T}_{y\pi,reg}(t_2:t_1))$ es un estimador disponible de $V(\hat{T}_{y\pi,reg}(t_2:t_1))$. El estimador propuesto por Cochran (1993) adaptado en este contexto es (Scott y Köhl, 1994):

$$v(\hat{T}_{y\pi,reg}(t_2:t_1)) = s_{\hat{T}_{y\pi}(t_2), \hat{T}_{y\pi}(t_1)}^2 \left(\frac{1}{m_{12}} + \frac{(\hat{T}_{y\pi,rep}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12}))^2}{s_y^2(t_1, m_{12})(m_{12} - 1)} \right) + \frac{s_y^2(t_2, m_{12}) - s_{\hat{T}_{y\pi}(t_2), \hat{T}_{y\pi}(t_1)}^2}{m_1} - \frac{s_y^2(t_2, m_{12})}{N} \quad (4.39)$$

Donde

$$s_y^2(t_1, m_{12})(m_{12} - 1) = \sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi s}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12}) \right)^2 \quad (4.40)$$

$$s_y^2(t_2, m_{12}) = \frac{\sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12}) \right)^2}{m_{12} - 1} \quad (4.41)$$

$$s_y^2(t_2, m_{12})(m_{12} - 1) = \sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12}) \right)^2 \quad (4.42)$$

$$s_{\hat{T}_{y\pi}(t_2), \hat{T}_{y\pi}(t_1)}^2 = \frac{\sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12}) - \hat{B}_{21}(\hat{T}_{y\pi s}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12})) \right)^2}{m_{12} - 2} \quad (4.43)$$

La expresión (4.43) corresponde a la suma de cuadrados del error de la regresión de

$\hat{T}_{y\pi s}(t_2)$ sobre $\hat{T}_{y\pi s}(t_1)$.

Por otra parte, $\hat{w}(t_2)$ se expresa como

$$\hat{w}(t_2) = \frac{1}{v(\hat{T}_{y\pi,rep}(t_2, m_2))} \quad (4.44)$$

Donde $v(\hat{T}_{y\pi,rep}(t_2, m_2))$ es un estimador disponible de $V(\hat{T}_{y\pi,rep}(t_2, m_2))$, es decir,

$$v(\hat{T}_{y\pi,rep}(t_2, m_2)) = s_y^2(t_2, m_2) = \frac{1}{m_2 - 1} \sum_{s=1}^{m_2} \left(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_2) \right)^2 \quad (4.45)$$

Finalmente, la varianza del estimador compuesto propuesta por Scott (1984) y Gregoire y Valentine (2008) es la siguiente:

$$v(\hat{T}_{y,MRP}(t_2 : t_1)) = \frac{1 + 4\Psi}{\hat{W}_{MRP}} \quad (4.46)$$

Donde

$$\Psi = \left(\frac{\hat{w}(t_2 : t_1)}{\hat{W}_{MRP}} \right) \left(\frac{\hat{w}(t_2)}{\hat{W}_{MRP}} \right) \left(\frac{1}{m_{12} - 1} + \frac{1}{m_2 - 1} \right) \quad (4.47)$$

En el marco del Sistema de Medición, Reporte y Verificación (MRV) para REDD, Köhl *et al.* (2011) al estimar biomasa aérea compararon varias estrategias de muestreo. Para un tamaño de muestra de 200 el muestreo en dos fases con estimadores de regresión (expresión 4.33) resultó ser el más preciso; presentó un error estándar relativo alrededor del 5%. El estimador del muestreo aleatorio simple arrojó un error estándar relativo del 20% aproximadamente utilizando también 200 parcelas de muestreo. Aun cuando el desempeño del muestreo doble con estimadores de regresión depende fuertemente de la correlación entre la variable de interés y la variable auxiliar, los autores señalan que cuando exista información de sensores remotos tales como LIDAR este método de estimación resulta ser el preferido.

4.5.2.2. Estimador compuesto de $T_y(t_1)$

Este estimador usa información de la muestra en la ocasión 2 para ajustar, vía regresión, la estimación de $T_y(t_1)$.

Siguiendo una línea similar de desarrollo como en la sección precedente se tiene que

$$\hat{T}_{y,MRP}(t_1 : t_2) = d'(\hat{T}_{y\pi,rep}(t_1, m_{12}) + B_{12}(\hat{T}_{y\pi,rep}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12}))) + (1 - d')\hat{T}_{y\pi,rep}(t_1, m_1) \quad (4.48)$$

Donde

$$d' = \frac{m_{12}}{m_1 - u_2 m_1 \rho_{12}^2} \quad (4.49)$$

O bien, de manera condensada

$$\hat{T}_{y,MRP}(t_1 : t_2) = d' \hat{T}_{y\pi,reg}(t_1 : t_2) + (1 - d') \hat{T}_{y\pi,rep}(t_1, m_{1.}) \quad (4.50)$$

Donde

$$\hat{T}_{y\pi,reg}(t_1 : t_2) = \hat{T}_{y\pi,rep}(t_1, m_{12}) + B_{12}(\hat{T}_{y\pi,rep}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12})) \quad (4.51)$$

La expresión (4.50) es una combinación del estimador de regresión basado en las parcelas remedidas y el estimador del muestreo replicado basado en las $m_{1.}$ parcelas temporales en la ocasión 1. El estimador de regresión considera la información actual, la cual es ajustada a través de regresión a la primera ocasión.

Al considerar las estimaciones de las varianzas y del coeficiente de regresión, el estimador combinado es el siguiente:

$$\hat{T}_{y,MRP}(t_1 : t_2) = \frac{\hat{w}(t_1 : t_2) \hat{T}_{y\pi,reg}(t_1 : t_2) + \hat{w}(t_1) \hat{T}_{y\pi,rep}(t_1, m_{1.})}{\hat{w}_{MRP}} \quad (4.52)$$

Donde

$$\hat{w}_{MRP} = \hat{w}(t_1 : t_2) + \hat{w}(t_1) \quad (4.53)$$

Y el estimador de regresión queda como

$$\hat{T}_{y\pi,reg}(t_1 : t_2) = \hat{T}_{y\pi,rep}(t_1, m_{12}) + \hat{B}_{12}(\hat{T}_{y\pi,rep}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12})) \quad (4.54)$$

Donde

$$\hat{B}_{12} = \frac{s_y(t_1 : t_2)}{s_y^2(t_2, m_{12})} \quad (4.55)$$

Aquí

$$s_y(t_1 : t_2) = \frac{1}{m_{12} - 1} \sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12}) \right) \left(\hat{T}_{y\pi s}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12}) \right) \quad (4.56)$$

$$s_y^2(t_2, m_{12}) = \frac{1}{m_{12} - 1} \sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12}) \right)^2 \quad (4.57)$$

La varianza estimada del estimador de regresión es

$$v(\hat{T}_{y\pi,reg}(t_1 : t_2)) = s_{\hat{T}_{y\pi}(t_1), \hat{T}_{y\pi}(t_2)}^2 \left(\frac{1}{m_{12}} + \frac{(\hat{T}_{y\pi,rep}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12}))^2}{s_y^2(t_2, m_{12})(m_{12} - 1)} \right) + \frac{s_y^2(t_1, m_{12}) - s_{\hat{T}_{y\pi}(t_1), \hat{T}_{y\pi}(t_2)}^2}{m_2} - \frac{s_y^2(t_1, m_{12})}{N} \quad (4.58)$$

Donde

$$s_y^2(t_2, m_{12})(m_{12} - 1) = \sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12}) \right)^2 \quad (4.59)$$

$$s_y^2(t_1, m_{12}) = \frac{\sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi s}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12}) \right)^2}{(m_{12} - 1)} \quad (4.60)$$

$$s_{\hat{T}_{y\pi}(t_1), \hat{T}_{y\pi}(t_2)}^2 = \frac{\sum_{s=1}^{m_{12}} \left(\hat{T}_{y\pi s}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12}) - \hat{B}_{12}(\hat{T}_{y\pi s}(t_2) - \hat{T}_{y\pi,rep}(t_2, m_{12})) \right)^2}{m_{12} - 2} \quad (4.61)$$

Así

$$\hat{w}(t_1 : t_2) = \frac{1}{v(\hat{T}_{y\pi,reg}(t_1 : t_2))} \quad (4.62)$$

Por otra parte, $\hat{w}(t_1)$ se expresa como

$$\hat{w}(t_1) = \frac{1}{v(\hat{T}_{y\pi,rep}(t_1, m_{1.}))} \quad (4.63)$$

Donde $v(\hat{T}_{y\pi,rep}(t_1, m_{1.}))$ está dado por

$$s_y^2(t_1, m_{1.}) = \frac{1}{m_{1.} - 1} \sum_{s=1}^{m_{1.}} \left(\hat{T}_{y\pi s}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{1.}) \right)^2 \quad (4.64)$$

Finalmente, la varianza del estimador compuesto adaptada de Scott (1984) y Gregoire and Valentine (2008) es:

$$v(\hat{T}_{y,MRP}(t_1 : t_2)) = \frac{1 + 4\Psi}{\hat{W}_{MRP}} \quad (4.65)$$

Donde

$$\Psi = \left(\frac{\hat{w}(t_1 : t_2)}{\hat{W}_{MRP}} \right) \left(\frac{\hat{w}(t_1)}{\hat{W}_{MRP}} \right) \left(\frac{1}{m_{12} - 1} + \frac{1}{m_{1.} - 1} \right) \quad (4.66)$$

En base a lo anterior, el estimador de mínima varianza de $\Delta_y(t_2, t_1) = T_y(t_2) - T_y(t_1)$ es

$$\hat{\Delta}_{y,MPR}(t_2, t_1) = \hat{T}_{y,MRP}(t_2 : t_1) - \hat{T}_{y,MRP}(t_1 : t_2) \quad (4.67)$$

Con varianza estimada dada por

$$v(\hat{\Delta}_{y,MPR}(t_2, t_1)) = v(\hat{T}_{y,MRP}(t_2 : t_1)) + v(\hat{T}_{y,MRP}(t_1 : t_2)) - 2\text{COV}(\hat{T}_{y,MRP}(t_2 : t_1), \hat{T}_{y,MRP}(t_1 : t_2)) \quad (4.68)$$

4.5.3. Muestreo con reemplazo parcial en dos ocasiones mediante Mínimos Cuadrados Generalizados

Asumiendo muestreo aleatorio simple o sistemático, Van Deusen (1989) con base en el planteamiento teórico propuesto por Dixon y Howitt (1979), formuló la teoría del muestreo con reemplazamiento parcial en un contexto de Mínimos Cuadrados Generalizados, que simultáneamente estima los valores actuales, los valores pasados y el cambio neto entre ocasiones.

En este sentido, el modelo lineal general es el siguiente:

$$\hat{T} = DT_{y,MRP} + \varepsilon \quad (4.69)$$

Donde

\hat{T} : Vector de totales estimados correspondientes a cada grupo específico. Cada elemento del vector corresponde a un grupo de parcelas permanentes o temporales medidas al mismo tiempo.

D : Matriz de ceros y unos que indica la ocasión de medición. El número de columnas de la matriz corresponde al número de ocasiones que se evalúan.

$T_{y,MRP}$: Vector de totales poblacionales. El número de filas del vector corresponde al número de ocasiones que se evalúan.

ε : Vector de errores correspondientes a los promedios de totales escalados, con $N \sim (0, \Sigma)$.

Asumiendo que Σ es conocida, el estimador MCG es (Fuller, 2009)

$$\hat{T}_{y,MRP} = [D'\Sigma^{-1}D]^{-1} D'\Sigma^{-1} \hat{T} \quad (4.70)$$

Con varianza dada por

$$V(\hat{T}_{y,MRP}) = [D'\Sigma^{-1}D]^{-1} \quad (4.71)$$

Debido a que Σ casi nunca es conocida, sus componentes deben ser estimados a partir de la muestra, obteniéndose así el estimador MCG factible. Dada la estructura del

modelo (4.69), estas estimaciones provienen de las varianzas y covarianzas muestrales de los diferentes grupos específicos involucrados en el muestreo (Scott *et al.*, 1999). Las covarianzas surgen de las parcelas de muestreo remedidas en los diferentes grupos específicos.

Por lo tanto, el estimador MCG factible es:

$$\hat{T}_{y,MRP} = \left[D' \hat{\Sigma}^{-1} D \right]^{-1} D' \hat{\Sigma}^{-1} \hat{T} \quad (4.72)$$

y

$$v(\hat{T}_{y,MRP}) = \left[D' \hat{\Sigma}^{-1} D \right]^{-1} \quad (4.73)$$

Para el muestreo con reemplazo parcial en dos ocasiones (Figura 4.1), se tiene el siguiente sistema de ecuaciones (Dixon y Howitt, 1979; Rencher y Bruce, 2008, página 297):

$$\begin{aligned} \hat{T}_{y\pi,rep}(t_1, m_{1.}) &= T_{y,MRP}(t_1) + 0T_{y,MRP}(t_2) + \varepsilon_{y\pi,rep}(t_1, m_{1.}) \\ \hat{T}_{y\pi,rep}(t_1, m_{12}) &= T_{y,MRP}(t_1) + 0T_{y,MRP}(t_2) + \varepsilon_{y\pi,rep}(t_1, m_{12}) \\ \hat{T}_{y\pi,rep}(t_2, m_{12}) &= 0T_{y,MRP}(t_1) + T_{y,MRP}(t_2) + \varepsilon_{y\pi,rep}(t_2, m_{12}) \\ \hat{T}_{y\pi,rep}(t_2, m_{2.}) &= 0T_{y,MRP}(t_1) + T_{y,MRP}(t_2) + \varepsilon_{y\pi,rep}(t_2, m_{2.}) \end{aligned} \quad (4.74a)$$

Que, expresado en forma matricial es (Van Deusen, 1989)

$$\begin{bmatrix} \hat{T}_{y\pi,rep}(t_1, m_{1.}) \\ \hat{T}_{y\pi,rep}(t_1, m_{12}) \\ \hat{T}_{y\pi,rep}(t_2, m_{12}) \\ \hat{T}_{y\pi,rep}(t_2, m_{2.}) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} T_{y,MRP}(t_1) \\ T_{y,MRP}(t_2) \end{bmatrix} + \begin{bmatrix} \varepsilon_{y\pi,rep}(t_1, m_{1.}) \\ \varepsilon_{y\pi,rep}(t_1, m_{12}) \\ \varepsilon_{y\pi,rep}(t_2, m_{12}) \\ \varepsilon_{y\pi,rep}(t_2, m_{2.}) \end{bmatrix} \quad (4.74b)$$

El vector \hat{T} presenta cuatro totales estimados porque se tienen cuatro grupos específicos.

La matriz D tiene dos columnas porque existen dos ocasiones de medición.

El vector $T_{y,MRP}$ tiene dos filas porque existen dos tiempos de medición.

El vector ε presenta cuatro elementos porque se tienen cuatro grupos específicos.

La matriz de varianzas y covarianzas de los errores, Σ , es una matriz de 4 x 4. En la diagonal principal de esta matriz están las varianzas de los errores y fuera de ella sus covarianzas, es decir,

$$\Sigma = \begin{bmatrix} Cov(\varepsilon_{y\pi,rep}(t_1, m_1), \varepsilon_{y\pi,rep}(t_1, m_1)) & Cov(\varepsilon_{y\pi,rep}(t_1, m_1), \varepsilon_{y\pi,rep}(t_1, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_1, m_1), \varepsilon_{y\pi,rep}(t_2, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_1, m_1), \varepsilon_{y\pi,rep}(t_2, m_2)) \\ Cov(\varepsilon_{y\pi,rep}(t_1, m_{12}), \varepsilon_{y\pi,rep}(t_1, m_1)) & Cov(\varepsilon_{y\pi,rep}(t_1, m_{12}), \varepsilon_{y\pi,rep}(t_1, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_1, m_{12}), \varepsilon_{y\pi,rep}(t_2, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_1, m_{12}), \varepsilon_{y\pi,rep}(t_2, m_2)) \\ Cov(\varepsilon_{y\pi,rep}(t_2, m_{12}), \varepsilon_{y\pi,rep}(t_1, m_1)) & Cov(\varepsilon_{y\pi,rep}(t_2, m_{12}), \varepsilon_{y\pi,rep}(t_1, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_2, m_{12}), \varepsilon_{y\pi,rep}(t_2, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_2, m_{12}), \varepsilon_{y\pi,rep}(t_2, m_2)) \\ Cov(\varepsilon_{y\pi,rep}(t_2, m_2), \varepsilon_{y\pi,rep}(t_1, m_1)) & Cov(\varepsilon_{y\pi,rep}(t_2, m_2), \varepsilon_{y\pi,rep}(t_1, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_2, m_2), \varepsilon_{y\pi,rep}(t_2, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_2, m_2), \varepsilon_{y\pi,rep}(t_2, m_2)) \end{bmatrix} \quad (4.75)$$

$$\Sigma = \begin{bmatrix} Var(\varepsilon_{y\pi,rep}(t_1, m_1)) & Cov(\varepsilon_{y\pi,rep}(t_1, m_1), \varepsilon_{y\pi,rep}(t_1, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_1, m_1), \varepsilon_{y\pi,rep}(t_2, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_1, m_1), \varepsilon_{y\pi,rep}(t_2, m_2)) \\ Cov(\varepsilon_{y\pi,rep}(t_1, m_{12}), \varepsilon_{y\pi,rep}(t_1, m_1)) & Var(\varepsilon_{y\pi,rep}(t_1, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_1, m_{12}), \varepsilon_{y\pi,rep}(t_2, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_1, m_{12}), \varepsilon_{y\pi,rep}(t_2, m_2)) \\ Cov(\varepsilon_{y\pi,rep}(t_2, m_{12}), \varepsilon_{y\pi,rep}(t_1, m_1)) & Cov(\varepsilon_{y\pi,rep}(t_2, m_{12}), \varepsilon_{y\pi,rep}(t_1, m_{12})) & Var(\varepsilon_{y\pi,rep}(t_2, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_2, m_{12}), \varepsilon_{y\pi,rep}(t_2, m_2)) \\ Cov(\varepsilon_{y\pi,rep}(t_2, m_2), \varepsilon_{y\pi,rep}(t_1, m_1)) & Cov(\varepsilon_{y\pi,rep}(t_2, m_2), \varepsilon_{y\pi,rep}(t_1, m_{12})) & Cov(\varepsilon_{y\pi,rep}(t_2, m_2), \varepsilon_{y\pi,rep}(t_2, m_{12})) & Var(\varepsilon_{y\pi,rep}(t_2, m_2)) \end{bmatrix} \quad (4.76)$$

Esta matriz es estimada a través de las varianzas y covarianzas muestrales de los grupos específicos involucrados (Scott *et al.*, 1999), es decir,

$$\hat{\Sigma} = \begin{bmatrix} Var(\hat{T}_{y\pi,rep}(t_1, m_1)) & Cov(\hat{T}_{y\pi,rep}(t_1, m_1), \hat{T}_{y\pi,rep}(t_1, m_{12})) & Cov(\hat{T}_{y\pi,rep}(t_1, m_1), \hat{T}_{y\pi,rep}(t_2, m_{12})) & Cov(\hat{T}_{y\pi,rep}(t_1, m_1), \hat{T}_{y\pi,rep}(t_2, m_2)) \\ Cov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_1, m_1)) & Var(\hat{T}_{y\pi,rep}(t_1, m_{12})) & Cov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) & Cov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_2)) \\ Cov(\hat{T}_{y\pi,rep}(t_2, m_{12}), \hat{T}_{y\pi,rep}(t_1, m_1)) & Cov(\hat{T}_{y\pi,rep}(t_2, m_{12}), \hat{T}_{y\pi,rep}(t_1, m_{12})) & Var(\hat{T}_{y\pi,rep}(t_2, m_{12})) & Cov(\hat{T}_{y\pi,rep}(t_2, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_2)) \\ Cov(\hat{T}_{y\pi,rep}(t_2, m_2), \hat{T}_{y\pi,rep}(t_1, m_1)) & Cov(\hat{T}_{y\pi,rep}(t_2, m_2), \hat{T}_{y\pi,rep}(t_1, m_{12})) & Cov(\hat{T}_{y\pi,rep}(t_2, m_2), \hat{T}_{y\pi,rep}(t_2, m_{12})) & Var(\hat{T}_{y\pi,rep}(t_2, m_2)) \end{bmatrix} \quad (4.77)$$

Debido a que los únicos totales correlacionados son aquellos que provienen de las parcelas remedidas (Figura 4.1), $\hat{\Sigma}$ queda finalmente de la siguiente manera:

$$\hat{\Sigma} = \begin{bmatrix} Var(\hat{T}_{y\pi,rep}(t_1, m_1)) & 0 & 0 & 0 \\ 0 & Var(\hat{T}_{y\pi,rep}(t_1, m_{12})) & Cov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) & 0 \\ 0 & Cov(\hat{T}_{y\pi,rep}(t_2, m_{12}), \hat{T}_{y\pi,rep}(t_1, m_{12})) & Var(\hat{T}_{y\pi,rep}(t_2, m_{12})) & 0 \\ 0 & 0 & 0 & Var(\hat{T}_{y\pi,rep}(t_2, m_2)) \end{bmatrix} \quad (4.78)$$

Conviene hacer hincapié que en el caso del muestreo con reemplazo parcial en dos ocasiones se tienen cuatro totales estimados correspondientes a cuatro grupos específicos (subpaneles) bien definidos y no traslapados (Figura 4.1), a saber:

Grupo específico 1 (subpanel 1): Parcelas temporales medidas en la ocasión 1.

$$\hat{T}_{y\pi,rep}(t_1, m_{1.}) = \frac{1}{m_{1.}} \sum_{s=1}^{m_{1.}} \hat{T}_{y\pi s}(t_1) \quad (4.12)$$

Grupo específico 2 (subpanel 2): Parcelas permanentes medidas en la ocasión 1.

$$\hat{T}_{y\pi,rep}(t_1, m_{12}) = \frac{1}{m_{12}} \sum_{s=1}^{m_{12}} \hat{T}_{y\pi s}(t_1) \quad (4.10)$$

Grupo específico 3 (subpanel 3): Parcelas permanentes medidas en la ocasión 2.

$$\hat{T}_{y\pi,rep}(t_2, m_{12}) = \frac{1}{m_{12}} \sum_{s=1}^{m_{12}} \hat{T}_{y\pi s}(t_2) \quad (4.11)$$

Grupo específico 4 (subpanel 4): Parcelas temporales medidas en la ocasión 2.

$$\hat{T}_{y\pi,rep}(t_2, m_{.2}) = \frac{1}{m_{.2}} \sum_{s=1}^{m_{.2}} \hat{T}_{y\pi s}(t_2) \quad (4.13)$$

Con el fin de considerar diseños de muestreo teóricos más complejos (reemplazamiento parcial en múltiples ocasiones), conviene modificar ligeramente la notación referente al número de elementos de cada grupo de la siguiente manera:

Grupo específico 1 (subpanel 1): Parcelas temporales medidas en la ocasión 1.

$$\hat{T}_{y\pi,rep}(t_1, m_{1.}) = \frac{1}{m_{1.}} \sum_{s=1}^{m_{1.}} \hat{T}_{y\pi s}(t_1) \Rightarrow \hat{T}_{y\pi,rep}(t_1, m_{11}) = \frac{1}{m_{11}} \sum_{s=1}^{m_{11}} \hat{T}_{y\pi s}(t_1) \quad (4.79)$$

Grupo específico 2 (subpanel 2): Parcelas permanentes medidas en la ocasión 1.

$$\hat{T}_{y\pi,rep}(t_1, m_{12}) = \frac{1}{m_{12}} \sum_{s=1}^{m_{12}} \hat{T}_{y\pi s}(t_1) \Rightarrow \hat{T}_{y\pi,rep}(t_1, m_{21}) = \frac{1}{m_{21}} \sum_{s=1}^{m_{21}} \hat{T}_{y\pi s}(t_1) \quad (4.80)$$

Grupo específico 3 (subpanel 3): Parcelas permanentes medidas en la ocasión 2.

$$\hat{T}_{y\pi,rep}(t_2, m_{12}) = \frac{1}{m_{12}} \sum_{s=1}^{m_{12}} \hat{T}_{y\pi s}(t_2) \Rightarrow \hat{T}_{y\pi,rep}(t_2, m_{22}) = \frac{1}{m_{22}} \sum_{s=1}^{m_{22}} \hat{T}_{y\pi s}(t_2) \quad (4.81)$$

Grupo específico 4 (subpanel 4): Parcelas temporales medidas en la ocasión 2.

$$\hat{T}_{y\pi,rep}(t_2, m_{.2}) = \frac{1}{m_{.2}} \sum_{s=1}^{m_2} \hat{T}_{y\pi s}(t_2) \Rightarrow \hat{T}_{y\pi,rep}(t_4, m_{32}) = \frac{1}{m_{32}} \sum_{s=1}^{m_{32}} \hat{T}_{y\pi s}(t_4) \quad (4.82)$$

En la Figura 4.2 se representa el muestreo con reemplazo parcial en dos ocasiones (t_1, t_4) considerando esta notación —en t_4 se lleva a cabo la segunda medición—. En el eje de las abscisas se tiene la escala temporal (usualmente en años) y en el eje de las ordenadas los grupos generales de las parcelas. El grupo general señala el tipo de parcelas que se miden a través del tiempo, es decir, temporales o permanentes. Así, el grupo 2 (g_2) identifica a las parcelas permanentes medidas en la ocasión de medición 1 (tiempo 1) y en la ocasión de medición 2 (tiempo 4), el grupo 1 (g_1) a las unidades de muestreo temporales evaluadas en la ocasión 1 (tiempo 1) y el grupo 3 (g_3) a las temporales estudiadas en la ocasión 2 (tiempo 4). Un grupo específico o subpanel de parcelas (m_{ij}) queda definido por la intersección de un grupo general (g_i) y la ocasión de medición (t_j). En este contexto, cada subpanel se identifica por dos subíndices, el primero se refiere al grupo general y el segundo a la ocasión de medición. De esta manera el subpanel m_{32} corresponde a las parcelas del grupo 3 medidas en la ocasión 2 (tiempo 4).

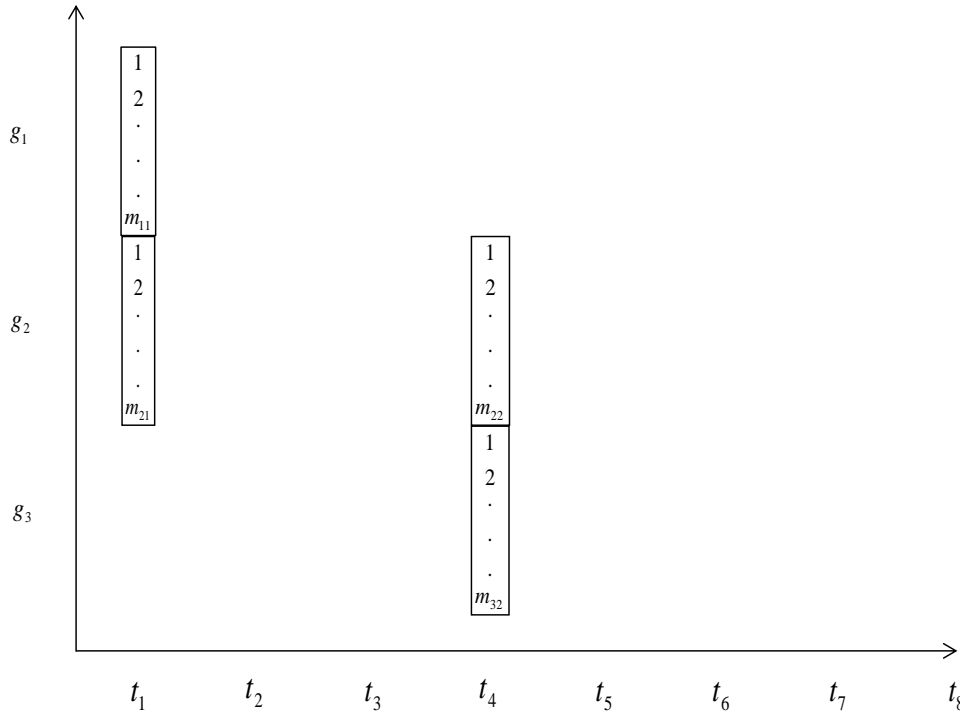


Figura 4.2. Muestreo con reemplazo parcial en dos ocasiones (t_1, t_4). Un grupo específico o subpanel de parcelas (m_{ij}) queda definido por la intersección de un grupo general (g_i) y la ocasión de medición (t_j). Todos los subpaneles en t_j integran un panel.

Una vez que los subpaneles han sido identificados bajo la notación esbozada, la matriz diseño y la matriz de varianzas y covarianzas se obtienen fácilmente, y en consecuencia, la estimación mediante MCG es casi inmediata.

Fuller (2009, pág. 236) presenta la teoría del estimador Mínimos Cuadrados Generalizados para el muestreo en dos ocasiones bajo el título *estimadores compuestos*, los cuales define como estimadores que son una función de muestras observadas en el tiempo.

4.5.4. Muestreo con reemplazo parcial en múltiples ocasiones vía Mínimos Cuadrados Generalizados

En un sistema de inventario y monitoreo forestal a largo plazo y a gran escala, un muestreo con reemplazamiento parcial en dos ocasiones no es aplicable, por lo que es imprescindible diseñar esquemas de muestreo con reemplazamiento parcial en ocasiones múltiples. Scott *et al.* (1999) muestran varios diseños teóricos panelizados

complejos. Con el objetivo de mostrar el modelo en forma matricial para situaciones como éstas, se presentan los siguientes ejemplos.

En la Figura 4.3 se representa una alternativa del muestreo con reemplazo parcial en tres ocasiones (Scott *et al.*, 1999). Ahora se tienen nuevos grupos específicos o subpaneles. El subpanel m_{42} corresponde al grupo general 4 medido en la ocasión 2 (t_4). El subpanel m_{43} corresponde al grupo general 4 medido en la ocasión 3 (t_7). Se observa que el ejemplo mostrado en la Figura 4.2 queda incluido en este nuevo diseño. El hecho de incluir otro tiempo de medición, trae consigo cinco subpaneles adicionales, luego entonces la representación matricial del diseño y la matriz de varianzas y covarianzas también crecen en complejidad.

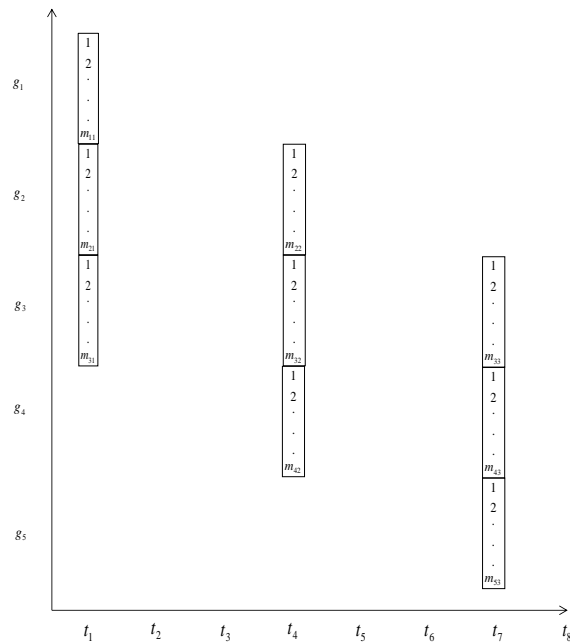


Figura 4.3. Una alternativa del muestreo con reemplazo parcial en tres ocasiones.

Para el muestreo con reemplazo parcial en tres ocasiones (Figura 4.3), se tiene el siguiente sistema de ecuaciones:

$$\hat{T}_{y\pi,rep}(t_1, m_{11}) = T_{y,MRP}(t_1) + 0T_{y,MRP}(t_4) + 0T_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_1, m_{11})$$

$$\hat{T}_{y\pi,rep}(t_1, m_{21}) = T_{y,MRP}(t_1) + 0T_{y,MRP}(t_4) + 0T_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_1, m_{21})$$

$$\hat{T}_{y\pi,rep}(t_1, m_{31}) = T_{y,MRP}(t_1) + 0T_{y,MRP}(t_4) + 0T_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_1, m_{31})$$

$$\begin{aligned}
\hat{T}_{y\pi,rep}(t_4, m_{22}) &= 0T_{y,MRP}(t_1) + T_{y,MRP}(t_4) + 0T_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_4, m_{22}) \\
\hat{T}_{y\pi,rep}(t_4, m_{32}) &= 0T_{y,MRP}(t_1) + T_{y,MRP}(t_4) + 0T_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_4, m_{32}) \\
\hat{T}_{y\pi,rep}(t_4, m_{42}) &= 0T_{y,MRP}(t_1) + T_{y,MRP}(t_4) + 0T_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_4, m_{42}) \\
\hat{T}_{y\pi,rep}(t_7, m_{33}) &= 0T_{y,MRP}(t_1) + 0T_{y,MRP}(t_4) + T_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_7, m_{33}) \\
\hat{T}_{y\pi,rep}(t_7, m_{43}) &= 0T_{y,MRP}(t_1) + 0T_{y,MRP}(t_4) + T_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_7, m_{43}) \\
\hat{T}_{y\pi,rep}(t_7, m_{53}) &= 0T_{y,MRP}(t_1) + 0T_{y,MRP}(t_4) + T_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_7, m_{53})
\end{aligned} \tag{4.83a}$$

Al considerar estas nueve ecuaciones, el modelo en forma matricial y la matriz de varianzas y covarianzas del diseño mostrado en la Figura (4.3) se presentan en las expresiones (4.83b) y (4.84), respectivamente.

$$\begin{bmatrix} \hat{T}_{y\pi,rep}(t_1, m_{11}) \\ \hat{T}_{y\pi,rep}(t_1, m_{21}) \\ \hat{T}_{y\pi,rep}(t_1, m_{31}) \\ \hat{T}_{y\pi,rep}(t_4, m_{22}) \\ \hat{T}_{y\pi,rep}(t_4, m_{32}) \\ \hat{T}_{y\pi,rep}(t_4, m_{42}) \\ \hat{T}_{y\pi,rep}(t_7, m_{33}) \\ \hat{T}_{y\pi,rep}(t_7, m_{43}) \\ \hat{T}_{y\pi,rep}(t_7, m_{53}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} T_{y,MRP}(t_1) \\ T_{y,MRP}(t_4) \\ T_{y,MRP}(t_7) \end{bmatrix} + \begin{bmatrix} \varepsilon_{y\pi,rep}(t_1, m_{11}) \\ \varepsilon_{y\pi,rep}(t_1, m_{21}) \\ \varepsilon_{y\pi,rep}(t_1, m_{31}) \\ \varepsilon_{y\pi,rep}(t_4, m_{22}) \\ \varepsilon_{y\pi,rep}(t_4, m_{32}) \\ \varepsilon_{y\pi,rep}(t_4, m_{42}) \\ \varepsilon_{y\pi,rep}(t_7, m_{33}) \\ \varepsilon_{y\pi,rep}(t_7, m_{43}) \\ \varepsilon_{y\pi,rep}(t_7, m_{53}) \end{bmatrix} \tag{4.83b}$$

$$\hat{\Sigma} = \begin{bmatrix} v(\hat{T}_{y\pi,rep}(t_1, m_{11})) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & v(\hat{T}_{y\pi,rep}(t_1, m_{21})) & 0 & c(m_{21}, m_{22}) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & v(\hat{T}_{y\pi,rep}(t_1, m_{31})) & 0 & c(m_{31}, m_{32}) & 0 & c(m_{31}, m_{33}) & 0 & 0 \\ 0 & c(m_{21}, m_{22}) & 0 & v(\hat{T}_{y\pi,rep}(t_4, m_{22})) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & c(m_{31}, m_{32}) & 0 & v(\hat{T}_{y\pi,rep}(t_4, m_{32})) & 0 & c(m_{32}, m_{33}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & v(\hat{T}_{y\pi,rep}(t_4, m_{42})) & 0 & c(m_{42}, m_{43}) & 0 \\ 0 & 0 & c(m_{31}, m_{33}) & 0 & c(m_{32}, m_{33}) & 0 & v(\hat{T}_{y\pi,rep}(t_7, m_{33})) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & c(m_{42}, m_{43}) & 0 & v(\hat{T}_{y\pi,rep}(t_7, m_{43})) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & v(\hat{T}_{y\pi,rep}(t_7, m_{53})) \end{bmatrix}$$

(4.84)

Donde

$$c(m_{21}, m_{22}) = \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{21}), \hat{T}_{y\pi,rep}(t_4, m_{22})), \quad c(m_{32}, m_{33}) = \text{cov}(\hat{T}_{y\pi,rep}(t_4, m_{32}), \hat{T}_{y\pi,rep}(t_7, m_{33})),$$

$$c(m_{31}, m_{32}) = \text{cov}(\hat{T}_{y\pi, rep}(t_1, m_{31}), \hat{T}_{y\pi, rep}(t_4, m_{32})), c(m_{42}, m_{43}) = \text{cov}(\hat{T}_{y\pi, rep}(t_4, m_{42}), \hat{T}_{y\pi, rep}(t_7, m_{43})),$$

$$c(m_{31}, m_{33}) = \text{cov}(\hat{T}_{y\pi, rep}(t_1, m_{31}), \hat{T}_{y\pi, rep}(t_7, m_{33}))$$

La teoría del muestreo con reemplazo parcial en dos y tres ocasiones (forma matricial) puede ser adaptada para diseños de muestreo multitemporales panelizados. En la Figura 4.4 se muestra un diseño que considera tres mallas sistemáticas sobre un región completa. En cada ocasión, una de las tres redes debe ser medida (Scott *et al.*, 1999), así el ciclo de remediación es de tres años. En este diseño también se generan nueve subpaneles o grupos específicos.

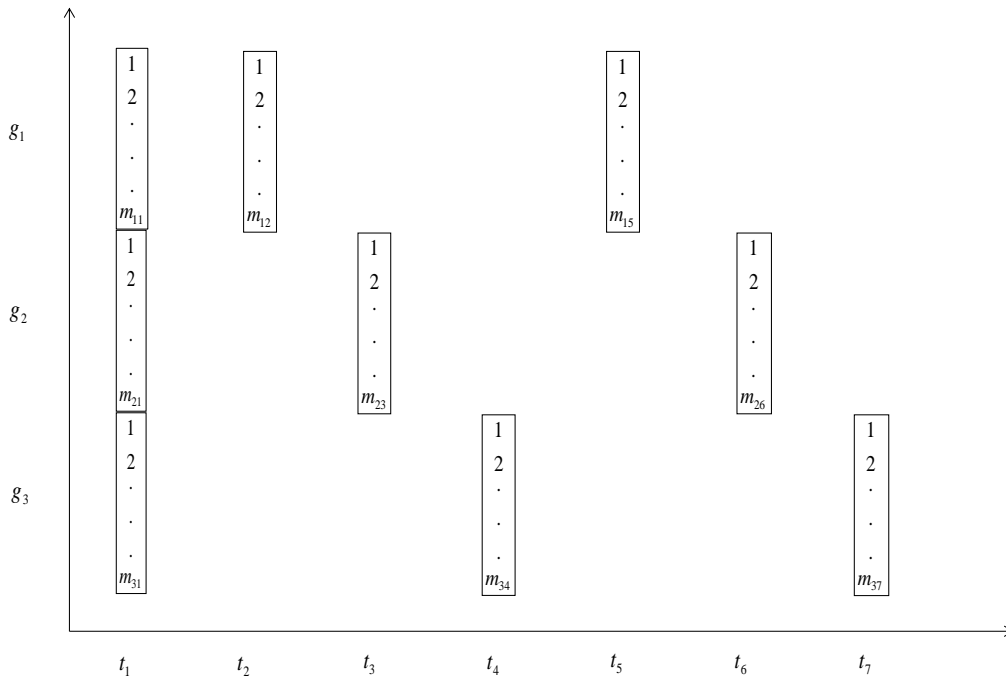


Figura 4.4. Muestreo multitemporal panelizado con ciclo de remediación de tres años.

Para el muestreo multitemporal panelizado con ciclo de remediación de tres años (Figura 4.4) se genera el siguiente sistema de ecuaciones:

$$\hat{T}_{y\pi, rep}(t_1, m_{11}) = T_{y, MRP}(t_1) + 0T_{y, MRP}(t_2) + 0T_{y, MRP}(t_3) + 0T_{y, MRP}(t_4) + 0T_{y, MRP}(t_5) + 0T_{y, MRP}(t_6) + 0T_{y, MRP}(t_7) + \varepsilon_{y\pi, rep}(t_1, m_{11})$$

$$\hat{T}_{y\pi, rep}(t_1, m_{21}) = T_{y, MRP}(t_1) + 0T_{y, MRP}(t_2) + 0T_{y, MRP}(t_3) + 0T_{y, MRP}(t_4) + 0T_{y, MRP}(t_5) + 0T_{y, MRP}(t_6) + 0T_{y, MRP}(t_7) + \varepsilon_{y\pi, rep}(t_1, m_{21})$$

$$\hat{T}_{y\pi, rep}(t_1, m_{31}) = T_{y, MRP}(t_1) + 0T_{y, MRP}(t_2) + 0T_{y, MRP}(t_3) + 0T_{y, MRP}(t_4) + 0T_{y, MRP}(t_5) + 0T_{y, MRP}(t_6) + 0T_{y, MRP}(t_7) + \varepsilon_{y\pi, rep}(t_1, m_{31})$$

$$\hat{T}_{y\pi, rep}(t_2, m_{12}) = 0T_{y, MRP}(t_1) + T_{y, MRP}(t_2) + 0T_{y, MRP}(t_3) + 0T_{y, MRP}(t_4) + 0T_{y, MRP}(t_5) + 0T_{y, MRP}(t_6) + 0T_{y, MRP}(t_7) + \varepsilon_{y\pi, rep}(t_2, m_{12})$$

$$\hat{T}_{y\pi, rep}(t_3, m_{23}) = 0T_{y, MRP}(t_1) + 0T_{y, MRP}(t_2) + T_{y, MRP}(t_3) + 0T_{y, MRP}(t_4) + 0T_{y, MRP}(t_5) + 0T_{y, MRP}(t_6) + 0T_{y, MRP}(t_7) + \varepsilon_{y\pi, rep}(t_3, m_{23})$$

$$\begin{aligned}
\hat{T}_{y\pi,rep}(t_4, m_{34}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + T_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_4, m_{34}) \\
\hat{T}_{y\pi,rep}(t_5, m_{15}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + T_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_5, m_{15}) \\
\hat{T}_{y\pi,rep}(t_6, m_{26}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + T_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_6, m_{26}) \\
\hat{T}_{y\pi,rep}(t_7, m_{37}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + T_{y,MRP}(t_7) + \varepsilon_{y\pi,rep}(t_7, m_{37})
\end{aligned}$$

(4.85a)

El modelo en forma matricial para el diseño de la Figura 4.4 es el siguiente:

$$\begin{bmatrix} \hat{T}_{y\pi,rep}(t_1, m_{11}) \\ \hat{T}_{y\pi,rep}(t_1, m_{21}) \\ \hat{T}_{y\pi,rep}(t_1, m_{31}) \\ \hat{T}_{y\pi,rep}(t_2, m_{12}) \\ \hat{T}_{y\pi,rep}(t_3, m_{23}) \\ \hat{T}_{y\pi,rep}(t_4, m_{34}) \\ \hat{T}_{y\pi,rep}(t_5, m_{15}) \\ \hat{T}_{y\pi,rep}(t_6, m_{26}) \\ \hat{T}_{y\pi,rep}(t_7, m_{37}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} T_{y,MRP}(t_1) \\ T_{y,MRP}(t_2) \\ T_{y,MRP}(t_3) \\ T_{y,MRP}(t_4) \\ T_{y,MRP}(t_5) \\ T_{y,MRP}(t_6) \\ T_{y,MRP}(t_7) \end{bmatrix} + \begin{bmatrix} \varepsilon_{y\pi,rep}(t_1, m_{11}) \\ \varepsilon_{y\pi,rep}(t_1, m_{21}) \\ \varepsilon_{y\pi,rep}(t_1, m_{31}) \\ \varepsilon_{y\pi,rep}(t_2, m_{12}) \\ \varepsilon_{y\pi,rep}(t_3, m_{23}) \\ \varepsilon_{y\pi,rep}(t_4, m_{34}) \\ \varepsilon_{y\pi,rep}(t_5, m_{15}) \\ \varepsilon_{y\pi,rep}(t_6, m_{26}) \\ \varepsilon_{y\pi,rep}(t_7, m_{37}) \end{bmatrix} \quad (4.85b)$$

Mientras que la matriz de varianzas y covarianzas estimadas es

$$\hat{\Sigma} = \begin{bmatrix} v(\hat{T}_{y\pi,rep}(t_1, m_{11})) & 0 & 0 & c(m_{11}, m_{12}) & 0 & 0 & c(m_{11}, m_{15}) & 0 & 0 \\ 0 & v(\hat{T}_{y\pi,rep}(t_1, m_{21})) & 0 & 0 & c(m_{21}, m_{23}) & 0 & 0 & c(m_{21}, m_{26}) & 0 \\ 0 & 0 & v(\hat{T}_{y\pi,rep}(t_1, m_{31})) & 0 & 0 & c(m_{31}, m_{34}) & 0 & 0 & c(m_{31}, m_{37}) \\ c(m_{11}, m_{12}) & 0 & 0 & v(\hat{T}_{y\pi,rep}(t_2, m_{12})) & 0 & 0 & c(m_{12}, m_{15}) & 0 & 0 \\ 0 & c(m_{21}, m_{23}) & 0 & 0 & v(\hat{T}_{y\pi,rep}(t_3, m_{23})) & 0 & 0 & c(m_{23}, m_{26}) & 0 \\ 0 & 0 & c(m_{31}, m_{34}) & 0 & 0 & v(\hat{T}_{y\pi,rep}(t_4, m_{34})) & 0 & 0 & c(m_{34}, m_{37}) \\ c(m_{11}, m_{15}) & 0 & 0 & c(m_{12}, m_{15}) & 0 & 0 & v(\hat{T}_{y\pi,rep}(t_5, m_{15})) & 0 & 0 \\ 0 & c(m_{21}, m_{26}) & 0 & 0 & c(m_{23}, m_{26}) & 0 & 0 & v(\hat{T}_{y\pi,rep}(t_6, m_{26})) & 0 \\ 0 & 0 & c(m_{31}, m_{37}) & 0 & 0 & c(m_{34}, m_{37}) & 0 & 0 & v(\hat{T}_{y\pi,rep}(t_7, m_{37})) \end{bmatrix}$$

(4.86)

Donde

$$\begin{aligned}
c(m_{11}, m_{12}) &= \text{COV}(\hat{T}_{y\pi,rep}(t_1, m_{11}), \hat{T}_{y\pi,rep}(t_2, m_{12})), \quad c(m_{11}, m_{15}) = \text{COV}(\hat{T}_{y\pi,rep}(t_1, m_{11}), \hat{T}_{y\pi,rep}(t_5, m_{15})) \\
c(m_{12}, m_{15}) &= \text{COV}(\hat{T}_{y\pi,rep}(t_2, m_{12}), \hat{T}_{y\pi,rep}(t_5, m_{15})), \quad c(m_{21}, m_{23}) = \text{COV}(\hat{T}_{y\pi,rep}(t_1, m_{21}), \hat{T}_{y\pi,rep}(t_3, m_{23})) \\
c(m_{21}, m_{26}) &= \text{COV}(\hat{T}_{y\pi,rep}(t_1, m_{21}), \hat{T}_{y\pi,rep}(t_6, m_{26})), \quad c(m_{23}, m_{26}) = \text{COV}(\hat{T}_{y\pi,rep}(t_3, m_{23}), \hat{T}_{y\pi,rep}(t_6, m_{26})) \\
c(m_{31}, m_{34}) &= \text{COV}(\hat{T}_{y\pi,rep}(t_1, m_{31}), \hat{T}_{y\pi,rep}(t_4, m_{34})), \quad c(m_{31}, m_{37}) = \text{COV}(\hat{T}_{y\pi,rep}(t_1, m_{31}), \hat{T}_{y\pi,rep}(t_7, m_{37})) \\
c(m_{34}, m_{37}) &= \text{COV}(\hat{T}_{y\pi,rep}(t_4, m_{34}), \hat{T}_{y\pi,rep}(t_7, m_{37}))
\end{aligned}$$

En la Figura 4.5 se presenta un diseño multitemporal similar al mostrado la Figura 4.4, sólo que en este caso la muestra original fue dividida en cinco grupos, por lo que el ciclo del estudio es de cinco años. En este diseño se generan 15 subpaneles, lo cual conduce a un sistema de 15 ecuaciones.

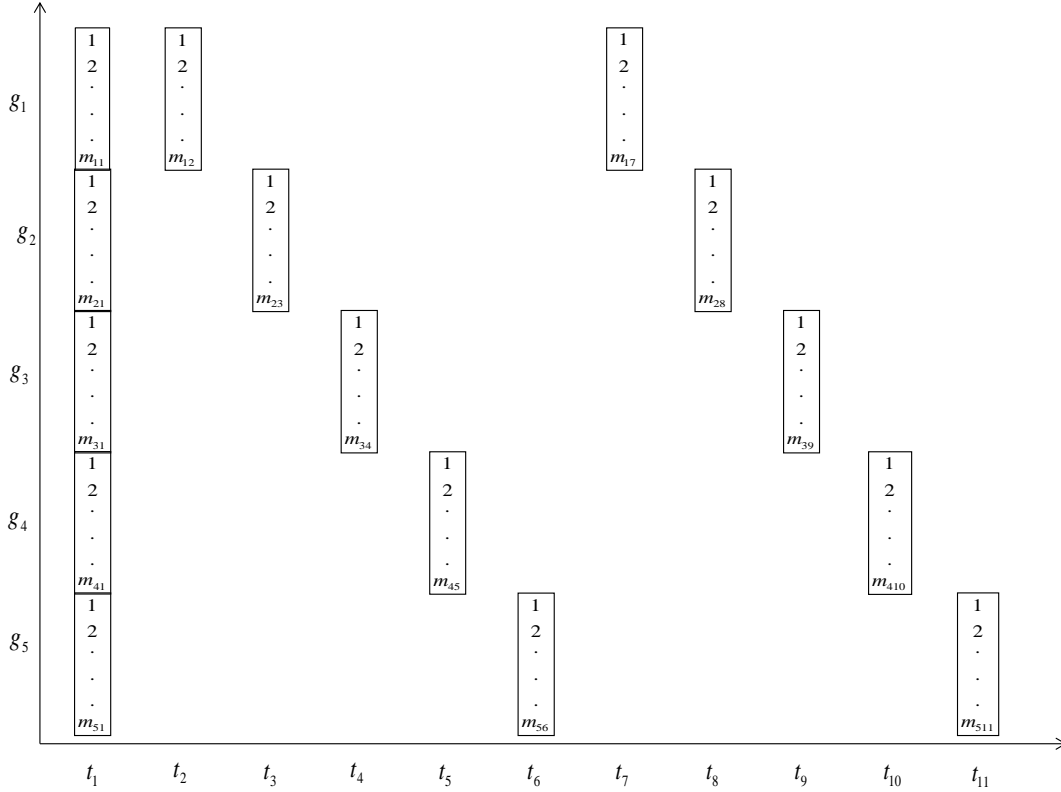


Figura 4.5. Muestreo multitemporal panelizado con ciclo de remediación de cinco años.

Para el muestreo multitemporal panelizado con ciclo de remediación de cinco años (Figura 4.5), se genera un sistema con las siguientes 15 ecuaciones:

$$\begin{aligned} \hat{T}_{y\pi,rep}(t_1, m_{11}) &= T_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_1, m_{11}) \\ \hat{T}_{y\pi,rep}(t_1, m_{21}) &= T_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_1, m_{21}) \\ \hat{T}_{y\pi,rep}(t_1, m_{31}) &= T_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_1, m_{31}) \\ \hat{T}_{y\pi,rep}(t_1, m_{41}) &= T_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_1, m_{41}) \\ \hat{T}_{y\pi,rep}(t_1, m_{51}) &= T_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_1, m_{51}) \\ \hat{T}_{y\pi,rep}(t_2, m_{12}) &= OT_{y,MRP}(t_1) + T_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_2, m_{12}) \\ \hat{T}_{y\pi,rep}(t_3, m_{23}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + T_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_3, m_{23}) \\ \hat{T}_{y\pi,rep}(t_4, m_{34}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + T_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_4, m_{34}) \\ \hat{T}_{y\pi,rep}(t_5, m_{45}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + T_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_5, m_{45}) \\ \hat{T}_{y\pi,rep}(t_6, m_{56}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + T_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_6, m_{56}) \\ \hat{T}_{y\pi,rep}(t_7, m_{17}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + T_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_7, m_{17}) \\ \hat{T}_{y\pi,rep}(t_8, m_{28}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + T_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_8, m_{28}) \end{aligned}$$

$$\begin{aligned}\hat{T}_{y\pi,rep}(t_9, m_{39}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + T_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_9, m_{39}) \\ \hat{T}_{y\pi,rep}(t_{10}, m_{410}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + T_{y,MRP}(t_{10}) + OT_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_{10}, m_{410}) \\ \hat{T}_{y\pi,rep}(t_{11}, m_{511}) &= OT_{y,MRP}(t_1) + OT_{y,MRP}(t_2) + OT_{y,MRP}(t_3) + OT_{y,MRP}(t_4) + OT_{y,MRP}(t_5) + OT_{y,MRP}(t_6) + OT_{y,MRP}(t_7) + OT_{y,MRP}(t_8) + OT_{y,MRP}(t_9) + OT_{y,MRP}(t_{10}) + T_{y,MRP}(t_{11}) + \varepsilon_{y\pi,rep}(t_{11}, m_{511})\end{aligned}$$

(4.87a)

La representación matricial y la matriz de varianzas y covarianzas del diseño correspondiente a la Figura 4.5 son, respectivamente,

$$\begin{bmatrix} \hat{T}_{y\pi,rep}(t_1, m_{11}) \\ \hat{T}_{y\pi,rep}(t_1, m_{21}) \\ \hat{T}_{y\pi,rep}(t_1, m_{31}) \\ \hat{T}_{y\pi,rep}(t_1, m_{41}) \\ \hat{T}_{y\pi,rep}(t_1, m_{51}) \\ \hat{T}_{y\pi,rep}(t_2, m_{12}) \\ \hat{T}_{y\pi,rep}(t_3, m_{23}) \\ \hat{T}_{y\pi,rep}(t_4, m_{34}) \\ \hat{T}_{y\pi,rep}(t_5, m_{45}) \\ \hat{T}_{y\pi,rep}(t_6, m_{56}) \\ \hat{T}_{y\pi,rep}(t_7, m_{17}) \\ \hat{T}_{y\pi,rep}(t_8, m_{28}) \\ \hat{T}_{y\pi,rep}(t_9, m_{39}) \\ \hat{T}_{y\pi,rep}(t_{10}, m_{410}) \\ \hat{T}_{y\pi,rep}(t_{11}, m_{511}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} T_{y,MRP}(t_1) \\ T_{y,MRP}(t_2) \\ T_{y,MRP}(t_3) \\ T_{y,MRP}(t_4) \\ T_{y,MRP}(t_5) \\ T_{y,MRP}(t_6) \\ T_{y,MRP}(t_7) \\ T_{y,MRP}(t_8) \\ T_{y,MRP}(t_9) \\ T_{y,MRP}(t_{10}) \\ T_{y,MRP}(t_{11}) \end{bmatrix} + \begin{bmatrix} \varepsilon_{y\pi,rep}(t_1, m_{11}) \\ \varepsilon_{y\pi,rep}(t_1, m_{21}) \\ \varepsilon_{y\pi,rep}(t_1, m_{31}) \\ \varepsilon_{y\pi,rep}(t_1, m_{41}) \\ \varepsilon_{y\pi,rep}(t_1, m_{51}) \\ \varepsilon_{y\pi,rep}(t_2, m_{12}) \\ \varepsilon_{y\pi,rep}(t_3, m_{23}) \\ \varepsilon_{y\pi,rep}(t_4, m_{34}) \\ \varepsilon_{y\pi,rep}(t_5, m_{45}) \\ \varepsilon_{y\pi,rep}(t_6, m_{56}) \\ \varepsilon_{y\pi,rep}(t_7, m_{17}) \\ \varepsilon_{y\pi,rep}(t_8, m_{28}) \\ \varepsilon_{y\pi,rep}(t_9, m_{39}) \\ \varepsilon_{y\pi,rep}(t_{10}, m_{410}) \\ \varepsilon_{y\pi,rep}(t_{11}, m_{511}) \end{bmatrix} \quad (4.87b)$$

$$\hat{\Sigma} = \begin{pmatrix}
\hat{v}(\hat{T}_{y\pi,rep}(t_1, m_{11})) & 0 & 0 & 0 & 0 & c(m_{11}, m_{12}) & 0 & 0 & 0 & 0 \\
0 & \hat{v}(\hat{T}_{y\pi,rep}(t_1, m_{21})) & 0 & 0 & 0 & 0 & c(m_{21}, m_{23}) & 0 & 0 & 0 \\
0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_1, m_{31})) & 0 & 0 & 0 & 0 & c(m_{31}, m_{34}) & 0 & 0 \\
0 & 0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_1, m_{41})) & 0 & 0 & 0 & 0 & c(m_{41}, m_{45}) & 0 \\
0 & 0 & 0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_1, m_{51})) & 0 & 0 & 0 & 0 & c(m_{51}, m_{56}) \\
c(m_{11}, m_{12}) & 0 & 0 & 0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_2, m_{12})) & 0 & 0 & 0 & 0 \\
0 & c(m_{21}, m_{23}) & 0 & 0 & 0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_3, m_{23})) & 0 & 0 & 0 \\
0 & 0 & c(m_{31}, m_{34}) & 0 & 0 & 0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_4, m_{34})) & 0 & 0 \\
0 & 0 & 0 & c(m_{41}, m_{45}) & 0 & 0 & 0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_5, m_{45})) & 0 \\
0 & 0 & 0 & 0 & c(m_{51}, m_{56}) & 0 & 0 & 0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_6, m_{56})) \\
c(m_{11}, m_{17}) & 0 & 0 & 0 & 0 & c(m_{12}, m_{17}) & 0 & 0 & 0 & 0 \\
0 & c(m_{21}, m_{28}) & 0 & 0 & 0 & 0 & c(m_{23}, m_{28}) & 0 & 0 & 0 \\
0 & 0 & c(m_{31}, m_{39}) & 0 & 0 & 0 & 0 & c(m_{34}, m_{39}) & 0 & 0 \\
0 & 0 & 0 & c(m_{41}, m_{410}) & 0 & 0 & 0 & 0 & c(m_{45}, m_{410}) & 0 \\
0 & 0 & 0 & 0 & c(m_{51}, m_{511}) & 0 & 0 & 0 & 0 & c(m_{56}, m_{511}) \\
c(m_{11}, m_{17}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & c(m_{21}, m_{28}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & c(m_{31}, m_{39}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & c(m_{41}, m_{410}) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & c(m_{51}, m_{511}) & 0 & 0 & 0 & 0 & 0 \\
c(m_{12}, m_{17}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & c(m_{23}, m_{28}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & c(m_{34}, m_{39}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & c(m_{45}, m_{410}) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & c(m_{56}, m_{511}) & 0 & 0 & 0 & 0 & 0 \\
\hat{v}(\hat{T}_{y\pi,rep}(t_7, m_{17})) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \hat{v}(\hat{T}_{y\pi,rep}(t_8, m_{28})) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_9, m_{39})) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_{10}, m_{410})) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \hat{v}(\hat{T}_{y\pi,rep}(t_{11}, m_{511})) & 0 & 0 & 0 & 0 & 0
\end{pmatrix} \quad (4.88)$$

Donde

$$\begin{aligned}
c(m_{11}, m_{12}) &= \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{11}), \hat{T}_{y\pi,rep}(t_2, m_{12})), c(m_{11}, m_{17}) = \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{11}), \hat{T}_{y\pi,rep}(t_7, m_{17})) \\
c(m_{12}, m_{17}) &= \text{cov}(\hat{T}_{y\pi,rep}(t_2, m_{12}), \hat{T}_{y\pi,rep}(t_7, m_{17})), c(m_{21}, m_{23}) = \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{21}), \hat{T}_{y\pi,rep}(t_3, m_{23})) \\
c(m_{21}, m_{28}) &= \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{21}), \hat{T}_{y\pi,rep}(t_8, m_{28})), c(m_{23}, m_{28}) = \text{cov}(\hat{T}_{y\pi,rep}(t_3, m_{23}), \hat{T}_{y\pi,rep}(t_8, m_{28})) \\
c(m_{31}, m_{34}) &= \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{31}), \hat{T}_{y\pi,rep}(t_4, m_{34})), c(m_{31}, m_{39}) = \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{31}), \hat{T}_{y\pi,rep}(t_9, m_{39})) \\
c(m_{34}, m_{39}) &= \text{cov}(\hat{T}_{y\pi,rep}(t_4, m_{34}), \hat{T}_{y\pi,rep}(t_9, m_{39})), c(m_{41}, m_{45}) = \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{41}), \hat{T}_{y\pi,rep}(t_5, m_{45})) \\
c(m_{41}, m_{410}) &= \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{41}), \hat{T}_{y\pi,rep}(t_{10}, m_{410})), c(m_{45}, m_{410}) = \text{cov}(\hat{T}_{y\pi,rep}(t_5, m_{45}), \hat{T}_{y\pi,rep}(t_{10}, m_{410})) \\
c(m_{51}, m_{56}) &= \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{51}), \hat{T}_{y\pi,rep}(t_6, m_{56})), c(m_{51}, m_{511}) = \text{cov}(\hat{T}_{y\pi,rep}(t_1, m_{51}), \hat{T}_{y\pi,rep}(t_{11}, m_{511})) \\
c(m_{56}, m_{511}) &= \text{cov}(\hat{T}_{y\pi,rep}(t_6, m_{56}), \hat{T}_{y\pi,rep}(t_{11}, m_{511}))
\end{aligned}$$

La estimación simultánea de medias poblacionales bajo el enfoque de Mínimos Cuadrados Generalizados, en estudios repetidos a través del tiempo, fue también abordada por Binder y Hidiroglou (1988). Estos autores refieren a las estimaciones en los subpaneles como “estimaciones elementales”.

Bajo la teoría de Mínimos Cuadrados Generalizados las estimaciones del cambio neto entre ocasiones son determinadas construyendo matrices de contrastes, C , las cuales toman las diferencias entre los componentes de $\hat{T}_{y,MRP}$. Por lo tanto, la expresión para estimar el cambio entre las ocasiones t_j y $t_{j'}$ es (Scott *et al.*, 1999; Arner *et al.* 2004)

$$\hat{\Delta}_{yjj',MRP} = C \hat{T}_{y,MRP} = C \left[D' \hat{\Sigma}^{-1} D \right]^{-1} D' \hat{\Sigma}^{-1} \hat{T} \quad (4.89)$$

Con varianza estimada por (Scott *et al.*, 1999; Arner *et al.* 2004; Rencher and Bruce, 2008)

$$v \left(\hat{\Delta}_{yjj',MRP} \right) = C v \left(\hat{T}_{y,MRP} \right) C' \quad (4.90)$$

4.5.4.1 Propiedades del estimador Mínimos Cuadrados Generalizados

Las propiedades del estimador MCG en donde Σ se supone conocida se resumen en el Teorema de Aitken (Greene, 2012, página 265):

“El estimador MCG es lineal, insesgado, consistente y distribuido normalmente desde el punto de vista asintótico. El estimador MCG es el estimador lineal insesgado de mínima varianza en el modelo de regresión general”.

Por otra parte, si $\hat{\Sigma}$ es un estimador consistente de Σ , el estimador MCG factible

es consistente, $p \lim \hat{T}_{y,MRP} = T_{y,MRP}$. Además si

$$p \lim \left[\left(\frac{1}{m} D' \hat{\Sigma}^{-1} D \right) - \left(\frac{1}{m} D' \Sigma^{-1} D \right) \right] = 0 \quad (4.91)$$

y

$$p \lim \left[\left(\frac{1}{\sqrt{m}} D' \hat{\Sigma}^{-1} \varepsilon \right) - \left(\frac{1}{\sqrt{m}} D' \Sigma^{-1} \varepsilon \right) \right] = 0 \quad (4.92)$$

el estimador de MCG factible es asintóticamente equivalente a MCG.

En el ANEXO 4.2 se presentan de manera detallada las propiedades del estimador MCG.

4.5.5. Estimadores del Muestreo Replicado

Un segundo método para estimar el parámetro de interés en cada ocasión (incluyendo el año más reciente) consiste en aplicar el estimador MRE en cada panel, es decir,

$$\hat{T}_{y\pi,rep}(t_j) = \frac{1}{m_j} \sum_{s=1}^{m_j} \hat{T}_{y\pi s}(t_j) \quad (4.93)$$

Donde, $j = 1, 2, \dots, P$.

En este contexto un panel está integrado por todos los subpaneles en t_j . La estimación de cada panel es tratada como una estimación independiente de los otros paneles.

La varianza estimada de (4.93) se expresa como

$$v(\hat{T}_{y\pi,rep}(t_j)) = \frac{1}{m_j(m_j - 1)} \sum_{s=1}^{m_j} \left(\hat{T}_{y\pi s}(t_j) - \hat{T}_{y\pi,rep}(t_j) \right)^2 \quad (4.94)$$

Dado que las estimaciones derivadas de los paneles se consideran independientes, el estimador del cambio entre dos ocasiones sucesivas y su varianza son, respectivamente,

$$\hat{\Delta}_y = \hat{T}_{y\pi,rep}(t_{j'}) - \hat{T}_{y\pi,rep}(t_j), \quad \forall t_{j'} > t_j \quad (4.95)$$

$$v(\hat{\Delta}_y) = v(\hat{T}_{y\pi,rep}(t_j)) + v(\hat{T}_{y\pi,rep}(t_{j'})) \quad (4.96)$$

Como ambos $\hat{T}_{y\pi,rep}(t_j)$ y $\hat{T}_{y\pi,rep}(t_{j'})$ son insesgados, entonces $\hat{\Delta}_y$ también es insesgado.

4.5.6. Estimador Media Móvil

En inventarios periódicos (tiempo entre estudios de 6 a 18 años) las estadísticas poblacionales pueden estimarse independientemente de una ocasión a otra. Las estimaciones de una ocasión dada poseen un alto nivel de precisión. En virtud de que la longitud del ciclo de medición es de varios años, las estimaciones previas contienen poca información acerca de las condiciones actuales. En inventarios anualizados sin embargo, los datos previos son de años relativamente recientes, de manera que ellos contienen una cantidad de información relevante acerca de las condiciones en el tiempo presente. No obstante, las estimaciones anuales tendrán menor precisión que los estudios periódicos debido simplemente a la muestra más pequeña medida cada año. Con esto en mente, se ha sugerido que las estimaciones para las condiciones actuales deben tomar ventaja de los datos previos (Johnson *et al.*, 2003).

Un estimador de la estadística poblacional actual que combina las estimaciones previas independientes es el Media Móvil (MM) (Moving Average=MA), dado por (Johnson *et al.*, 2003; Patterson and Reams, 2005)

$$\hat{T}_{MM,rep} = \sum_{j=1}^P w_j \hat{T}_{y\pi,rep}(t_j) \quad (4.97)$$

Cuya varianza es

$$v(\hat{T}_{MM,rep}) = \sum_{j=1}^P w_j^2 v(\hat{T}_{y\pi,rep}(t_j)) \quad (4.98)$$

Donde $\hat{T}_{y\pi,rep}(t_j)$ es un estimador insesgado de $T(t_j)$, la característica poblacional en la ocasión t_j , y w_j es un grupo de constantes las cuales suman 1 (Patterson and Reams, 2005). El FIA utiliza el estimador MM con pesos iguales (McRoberts *et al.*, 2005).

El estimador de media móvil con pesos iguales tiene muchas ventajas. Tanto el estimador como su varianza son fáciles para calcular a partir de datos de inventario. El estimador media móvil tendrá una varianza menor que cualquiera de los estimadores anuales. Estos hechos son atractivos tanto desde el punto de vista práctico como teórico. El mayor problema con el estimador media móvil es que resulta sesgado como un estimador del parámetro poblacional del año actual (McRoberts *et al.*, 2005). La justificación para el estimador de media móvil es que usando los datos del pasado se fortalece el tamaño de muestra. Mientras se genera un sesgo cuando se estima el parámetro de la condición actual, esto es más que compensado por la reducción en la varianza. Johnson *et al.* (2003) han mostrado, con simulaciones basadas en datos del FIA, que en muchas situaciones el promedio móvil con pesos iguales tiene el cuadrado medio del error más pequeño. Sin embargo, el estimador de pesos iguales puede ser pensado como un estimador insesgado para el parámetro de la población en algún tiempo aproximadamente en la mitad del ciclo de rotación.

Con el fin de evitar el sesgo del estimador promedio móvil con pesos iguales, algunos investigadores han sugerido asignar mayores pesos a los años más recientes. Por ejemplo, para un ciclo de cinco años se pueden usar pesos de 0.1, 0.1, 0.2, 0.3, y 0.3 y para una rotación de cuatro años los pesos pueden ser 0.2, 0.2, 0.3 y 0.3 (Eskelson *et al.*, 2009).

4.5.7. Estimador Compuesto Univariado

Los estimadores compuestos univariados son bien conocidos en aplicaciones de encuestas por muestreo, tal como muestreo sobre el tiempo o estimación de área-pequeña (Schreuder *et al.*, 1993, página 321; Gregoire y Valentine, 2008, página 388). Czaplewski y Thompson (2009) proveen una descripción tutorial del filtro de Kalman univariado en el contexto de datos de series de tiempo que ayuda a un entendimiento intuitivo de este estimador. Se puede combinar un estimador de varianza pequeña y un estimador de bajo sesgo.

Dada la importancia teórica y práctica del estimador compuesto presentado en la expresión (4.30) —y en (4.52)—, vale la pena mostrar otras equivalencias algebraicas encontradas comúnmente en la literatura, para tal efecto considérese la siguiente notación:

$\hat{\theta}_1$: Primer estimador.

$V(\hat{\theta}_1)$: Varianza del primer estimador.

$\hat{\theta}_2$: Segundo estimador.

$V(\hat{\theta}_2)$: Varianza del segundo estimador.

$\hat{\theta}$: Estimador compuesto.

$V(\hat{\theta})$: Varianza del estimador compuesto.

$\hat{\theta}_1$ y $\hat{\theta}_2$ en principio se consideran estimadores independientes.

De forma tal que el estimador compuesto de (4.30), en general se puede expresar de la siguiente manera (Gregoire y Valentine, 2008):

$$\hat{\theta} = \frac{\frac{1}{V(\hat{\theta}_1)}\hat{\theta}_1 + \frac{1}{V(\hat{\theta}_2)}\hat{\theta}_2}{\frac{1}{V(\hat{\theta}_1)} + \frac{1}{V(\hat{\theta}_2)}} \quad (4.99)$$

La varianza de (4.99) es

$$\begin{aligned} V(\hat{\theta}) &= \frac{1}{\left(\frac{1}{V(\hat{\theta}_1)} + \frac{1}{V(\hat{\theta}_2)}\right)^2} \left[\frac{1}{(V(\hat{\theta}_1))^2} V(\hat{\theta}_1) + \frac{1}{(V(\hat{\theta}_2))^2} V(\hat{\theta}_2) \right] \\ &= \frac{1}{\left(\frac{1}{V(\hat{\theta}_1)} + \frac{1}{V(\hat{\theta}_2)}\right)^2} \left[\frac{1}{V(\hat{\theta}_1)} + \frac{1}{V(\hat{\theta}_2)} \right] = \frac{1}{\frac{1}{V(\hat{\theta}_1)} + \frac{1}{V(\hat{\theta}_2)}} \quad (*) \end{aligned}$$

6

(*) esta expresión tiene la misma forma que la expresión (4.31)

$$V(\hat{\theta}) = \frac{1}{\frac{V(\hat{\theta}_2) + V(\hat{\theta}_1)}{V(\hat{\theta}_1)V(\hat{\theta}_2)}} = \frac{V(\hat{\theta}_1)V(\hat{\theta}_2)}{V(\hat{\theta}_2) + V(\hat{\theta}_1)} = \frac{V(\hat{\theta}_1)V(\hat{\theta}_2)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \quad (4.100)$$

Otra forma alternativa de (4.99) es (Czaplewski, 2010)

$$\hat{\theta} = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \hat{\theta}_1 + \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \hat{\theta}_2 \quad (4.101)$$

ó

$$\hat{\theta} = k \hat{\theta}_1 + (1-k) \hat{\theta}_2 \quad (4.101')$$

$$\text{donde } k = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \text{ y } 1-k = \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)}$$

Para obtener (4.101) basta desarrollar (4.99):

$$\begin{aligned} \hat{\theta} &= \frac{\frac{1}{V(\hat{\theta}_1)} \hat{\theta}_1 + \frac{1}{V(\hat{\theta}_2)} \hat{\theta}_2}{\frac{1}{V(\hat{\theta}_1)} + \frac{1}{V(\hat{\theta}_2)}} = \frac{\frac{1}{V(\hat{\theta}_1)}}{\frac{1}{V(\hat{\theta}_1)} + \frac{1}{V(\hat{\theta}_2)}} \hat{\theta}_1 + \frac{\frac{1}{V(\hat{\theta}_2)}}{\frac{1}{V(\hat{\theta}_1)} + \frac{1}{V(\hat{\theta}_2)}} \hat{\theta}_2 \\ &= \frac{\frac{1}{V(\hat{\theta}_1)}}{\frac{V(\hat{\theta}_2) + V(\hat{\theta}_1)}{V(\hat{\theta}_1)V(\hat{\theta}_2)}} \hat{\theta}_1 + \frac{\frac{1}{V(\hat{\theta}_2)}}{\frac{V(\hat{\theta}_2) + V(\hat{\theta}_1)}{V(\hat{\theta}_1)V(\hat{\theta}_2)}} \hat{\theta}_2 = \frac{V(\hat{\theta}_1)V(\hat{\theta}_2)}{V(\hat{\theta}_1)[V(\hat{\theta}_2) + V(\hat{\theta}_1)]} \hat{\theta}_1 + \frac{V(\hat{\theta}_1)V(\hat{\theta}_2)}{V(\hat{\theta}_2)[V(\hat{\theta}_2) + V(\hat{\theta}_1)]} \hat{\theta}_2 \\ &= \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_2) + V(\hat{\theta}_1)} \hat{\theta}_1 + \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_2) + V(\hat{\theta}_1)} \hat{\theta}_2 = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \hat{\theta}_1 + \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \hat{\theta}_2 \end{aligned}$$

Además,

$$k + (1-k) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} + \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} = \frac{V(\hat{\theta}_1) + V(\hat{\theta}_2)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} = 1$$

En el siguiente desarrollo se observa que la varianza de (4.101) es la misma expresión dada en (4.100):

$$\begin{aligned}
V(\hat{\theta}) &= \left(\frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \right)^2 V(\hat{\theta}_1) + \left(\frac{V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \right)^2 V(\hat{\theta}_2) \\
&= \frac{(V(\hat{\theta}_2))^2}{(V(\hat{\theta}_1) + V(\hat{\theta}_2))^2} V(\hat{\theta}_1) + \frac{(V(\hat{\theta}_1))^2}{(V(\hat{\theta}_1) + V(\hat{\theta}_2))^2} V(\hat{\theta}_2) \\
&= \frac{1}{(V(\hat{\theta}_1) + V(\hat{\theta}_2))^2} [(V(\hat{\theta}_2))^2 V(\hat{\theta}_1) + (V(\hat{\theta}_1))^2 V(\hat{\theta}_2)] = \frac{V(\hat{\theta}_2)V(\hat{\theta}_1)}{(V(\hat{\theta}_1) + V(\hat{\theta}_2))^2} [V(\hat{\theta}_2) + V(\hat{\theta}_1)] \\
&= \frac{V(\hat{\theta}_1)V(\hat{\theta}_2)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \tag{4.102}
\end{aligned}$$

El estimador (4.101) puede manipularse algebraicamente para representarlo en el estimador univariado del filtro de Kalman:

$$\begin{aligned}
\hat{\theta} &= \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \hat{\theta}_1 + \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \hat{\theta}_2 \\
&= \frac{V(\hat{\theta}_2)\hat{\theta}_1}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} + \frac{V(\hat{\theta}_1)\hat{\theta}_2}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} = \frac{\hat{\theta}_1 V(\hat{\theta}_2) + \hat{\theta}_2 V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} = \frac{\hat{\theta}_1 V(\hat{\theta}_1) + \hat{\theta}_1 V(\hat{\theta}_2) + \hat{\theta}_2 V(\hat{\theta}_1) - \hat{\theta}_1 V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \\
&= \frac{\hat{\theta}_1(V(\hat{\theta}_1) + V(\hat{\theta}_2)) + \hat{\theta}_2 V(\hat{\theta}_1) - \hat{\theta}_1 V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} = \frac{\hat{\theta}_1(V(\hat{\theta}_1) + V(\hat{\theta}_2))}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} + \frac{\hat{\theta}_2 V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} - \frac{\hat{\theta}_1 V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \\
&= \hat{\theta}_1 + (\hat{\theta}_2 - \hat{\theta}_1) \left(\frac{V(\hat{\theta}_1)}{V(\hat{\theta}_1) + V(\hat{\theta}_2)} \right) = \hat{\theta}_1 + (\hat{\theta}_2 - \hat{\theta}_1)(1 - k) \tag{4.103}
\end{aligned}$$

Donde, $1 - k$ como se definió en (4.101').

La expresión (4.103) corresponde al estimador de filtro de Kalman univariado (Czaplewski y Thompson, 2009; página 9, ecuación (11)).

La expresión (4.102) también puede manipularse algebraicamente para representar la varianza del estimador univariado del filtro de Kalman (Czaplewski y Thompson, 2009; página 10, ecuación (14)).

$$\begin{aligned}
V(\hat{\theta}) &= \frac{V(\hat{\theta}_1)V(\hat{\theta}_2)}{V(\hat{\theta}_1)+V(\hat{\theta}_2)} \\
&= \frac{V(\hat{\theta}_1)V(\hat{\theta}_1)+V(\hat{\theta}_1)V(\hat{\theta}_2)-V(\hat{\theta}_1)V(\hat{\theta}_1)}{V(\hat{\theta}_1)+V(\hat{\theta}_2)} = \frac{V(\hat{\theta}_1)[V(\hat{\theta}_1)+V(\hat{\theta}_2)]-V(\hat{\theta}_1)V(\hat{\theta}_1)}{V(\hat{\theta}_1)+V(\hat{\theta}_2)} \\
&= \frac{V(\hat{\theta}_1)[V(\hat{\theta}_1)+V(\hat{\theta}_2)]}{V(\hat{\theta}_1)+V(\hat{\theta}_2)} - \frac{V(\hat{\theta}_1)V(\hat{\theta}_1)}{V(\hat{\theta}_1)+V(\hat{\theta}_2)} = V(\hat{\theta}_1) - V(\hat{\theta}_1) \left(\frac{V(\hat{\theta}_1)}{V(\hat{\theta}_1)+V(\hat{\theta}_2)} \right) \\
&= V(\hat{\theta}_1) - V(\hat{\theta}_1)(1-k) \tag{4.104}
\end{aligned}$$

Cuando $\hat{\theta}_1$ y $\hat{\theta}_2$ se consideran estimadores no independientes, el valor de k se obtiene de la siguiente manera (Czaplewski, 2010):

$$k = \frac{V(\hat{\theta}_2) + Cov(\hat{\theta}_1, \hat{\theta}_2)}{V(\hat{\theta}_1) + V(\hat{\theta}_2) + 2Cov(\hat{\theta}_1, \hat{\theta}_2)} \tag{4.105}$$

Donde $Cov(\hat{\theta}_1, \hat{\theta}_2)$ es la covarianza entre $\hat{\theta}_1$ y $\hat{\theta}_2$.

En este caso la $V(\hat{\theta})$ es

$$V(\hat{\theta}) = k^2V(\hat{\theta}_1) + (1-k)^2V(\hat{\theta}_2) + 2k(1-k)Cov(\hat{\theta}_1, \hat{\theta}_2) \tag{4.106}$$

El estimador compuesto presenta las siguientes características (Czaplewski, 2010):

- Pondera cada estimador independiente inversamente proporcional a su varianza.
- El estimador más preciso (menor varianza) recibe el mayor peso, aunque el estimador menos preciso recibe un peso positivo. El estimador $\hat{\theta}_1$ puede considerarse como el más preciso.
- La varianza del estimador compuesto resultante será menor que la varianza de cualquiera de los estimadores cuando éstos son independientes.
- Si ambos estimadores son independientes e igualmente precisos cada uno de ellos recibe un peso igual de 0.5. El estimador compuesto tendrá una varianza

que es la mitad de cualquiera de las varianzas de los estimadores independientes.

Desde el punto de vista teórico, el producto de dos funciones densidad de probabilidad (FDP) normal es otra normal escalada (ANEXO 4.3) (Figura 4.6). De tal manera que si se conoce la media y la varianza verdadera de dos variables distribuidas normalmente es posible determinar los parámetros de la tercera. Bajo el esquema del estimador compuesto se tienen los siguientes resultados:

Sean $f(x)$ y $g(x)$ FDP Gaussianas con medias $\bar{\theta}_1$ y $\bar{\theta}_2$ y desviaciones estándar $\sqrt{V(\hat{\theta}_1)}$ y $\sqrt{V(\hat{\theta}_2)}$, respectivamente. Es decir,

$$f(x) = \frac{1}{\sqrt{2\pi}\sqrt{V(\hat{\theta}_1)}} e^{-\frac{(x-\bar{\theta}_1)^2}{2V(\hat{\theta}_1)}} \quad (4.107)$$

$$g(x) = \frac{1}{\sqrt{2\pi}\sqrt{V(\hat{\theta}_2)}} e^{-\frac{(x-\bar{\theta}_2)^2}{2V(\hat{\theta}_2)}} \quad (4.108)$$

El producto de estas dos FDP Gaussianas $f(x)$ y $g(x)$ es una PDF Gaussiana escalada:

$$f(x)g(x) \propto \frac{1}{\sqrt{2\pi}\sqrt{V(\hat{\theta})}} e^{-\frac{(x-\bar{\theta})^2}{2V(\hat{\theta})}} \quad (4.109)$$

En (4.109), $\bar{\theta}$ y $V(\hat{\theta})$ se expresan de la siguiente manera:

$$\bar{\theta} = \frac{\bar{\theta}_1 V(\hat{\theta}_2) + \bar{\theta}_2 V(\hat{\theta}_1)}{V(\hat{\theta}_2) + V(\hat{\theta}_1)} \quad (4.110)$$

$$V(\hat{\theta}) = \frac{V(\hat{\theta}_1)V(\hat{\theta}_2)}{V(\hat{\theta}_2) + V(\hat{\theta}_1)} \quad (4.111)$$

Se observa que la media de la función de densidad resultante, $\hat{\theta}$, es una media ponderada de las medias de las funciones de densidad originales (estimador compuesto de medias y varianzas verdaderas). El factor de ponderación es inversamente proporcional a las varianzas verdaderas. Resultados similares a los anteriores se obtienen al considerar el producto de una distribución normal *a priori* de un parámetro θ y la distribución normal de una variable Y (Hoff, 2009, página 71). Bajo el enfoque de densidades normales, Maybeck (1979, página 11) y Czaplewski y Thompson (2009; páginas 7-8) explican el estimador del filtro de Kalman univariado.

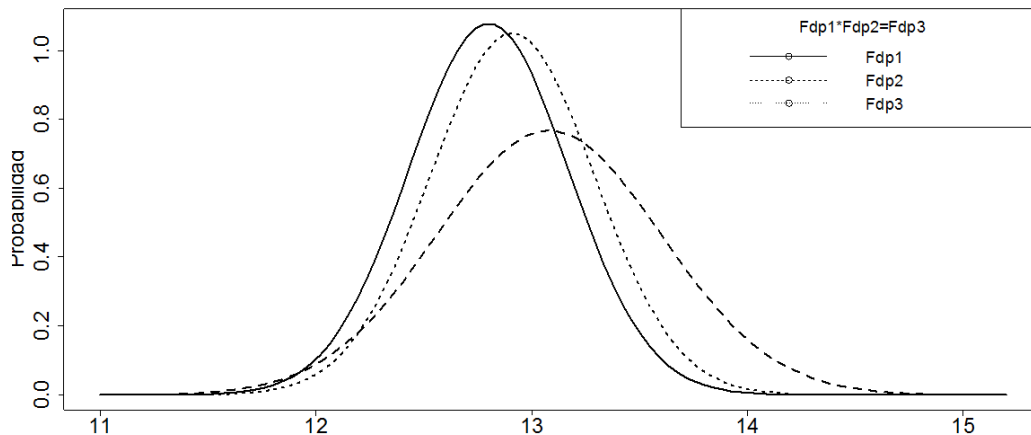


Figura 4.6. Representación del producto de dos funciones de densidad de probabilidad normal. El producto de $Fdp1 \times Fdp2$ es igual a $Fdp3$. La media de $Fdp3$ es una media ponderada de la media de $Fdp1$ y $Fdp2$.

En el ANEXO 4.3 se presenta el desarrollo teórico del producto de dos funciones de densidad normal.

El estimador mostrado en (4.30) —y en (4.52)— es un ejemplo de un estimador compuesto integrado por dos estimadores independientes. No obstante, el estimador compuesto que combina MCG y MRE, para una cierta ocasión de medición, es un ejemplo de estimador compuesto que combina dos estimadores no independientes, dado que un conjunto de unidades de muestreo es común en ambos.

4.5.8. Resumen de estimadores

Las estimaciones por ocasión se obtienen aplicando los siguientes estimadores:

$\hat{\theta}_j^1$: Estimador de Mínimos Cuadrados Generalizados en la ocasión j .
 $j = 1, 2, 3, 4, 5, 6, 7, 8$. Corresponde al j -ésimo elemento del siguiente vector:

$$\hat{T}_y = \begin{bmatrix} D' \hat{\Sigma}^{-1} & D \end{bmatrix}^{-1} D' \hat{\Sigma}^{-1} \hat{T} \quad (4.112)$$

Su varianza estimada, $v(\hat{\theta}_j^1)$, corresponde al elemento jj de la diagonal principal de la siguiente matriz:

$$v(\hat{T}_y) = \begin{bmatrix} D' \hat{\Sigma}^{-1} & D \end{bmatrix}^{-1} \quad (4.113)$$

$\hat{\theta}_j^2$: Estimador de Muestreo Replicado en la ocasión j . $j = 1, 2, 3, 4, 5, 6, 7, 8$. Se obtiene mediante la expresión:

$$\hat{T}_{y\pi,rep}(t_j) = \frac{1}{m_j} \sum_{s=1}^{m_j} \hat{T}_{y\pi s}(t) \quad (4.114)$$

Con varianza estimada

$$v(\hat{T}_{y\pi,rep}(t_j)) = \frac{1}{m_j(m_j - 1)} \sum_{s=1}^{m_j} \left(\hat{T}_{y\pi s}(t_j) - \hat{T}_{y\pi,rep}(t_j) \right)^2 \quad (4.115)$$

$\hat{\theta}_j^3$: Estimador Combinado Univariado en la ocasión j . $j = 1, 2, 3, 4, 5, 6, 7, 8$.

$$\hat{\theta}_j^3 = k \hat{\theta}_j^1 + (1 - k) \hat{\theta}_j^2 \quad (4.116)$$

Con varianza estimada

$$v(\hat{\theta}_j^3) = k^2 v(\hat{\theta}_j^1) + (1 - k)^2 v(\hat{\theta}_j^2) + 2k(1 - k) \text{cov}(\hat{\theta}_j^1, \hat{\theta}_j^2) \quad (4.117)$$

Donde

$$k = \frac{v(\hat{\theta}_j^2) + \text{cov}(\hat{\theta}_j^1, \hat{\theta}_j^2)}{v(\hat{\theta}_j^1) + v(\hat{\theta}_j^2) + 2\text{cov}(\hat{\theta}_j^1, \hat{\theta}_j^2)} \quad (4.118)$$

Donde el superíndice indica el tipo de estimador:

1: MCG, 2: MRE y 3: ECU

Bajo este enfoque, el estimador ECU combina el estimador MCG y MRE para estimar el parámetro de interés en cada ocasión de medición.

Si se supone una población teórica y el procedimiento de muestreo se repite un número grande de veces (R), entonces se puede conocer el valor esperado de cada estimador y su varianza verdadera:

$$E(\hat{\theta}_j^l) = \bar{\theta}_j^l = \frac{\sum_{r=1}^R \hat{\theta}_{rj}^l}{R} \quad (4.119)$$

$$V(\hat{\theta}_j^l) = \frac{\sum_{r=1}^R \left(\hat{\theta}_{rj}^l - \bar{\theta}_j^l \right)^2}{R} \quad (4.120)$$

4.5.9. Estimación del cambio

Con el fin de obtener estimaciones de cambio notables se propone que éstas sean entre los puntos medios de periodo, para tal efecto se considera lo siguiente:

Periodo 1: ocasiones de medición 2004, 2005, 2006 y 2007.

Periodo 2: ocasiones de medición 2009, 2010, 2011 y 2012.

$\hat{\theta}_{1j}^1$: Estimador de Mínimos Cuadrados Generalizados en la ocasión j del periodo uno. $j = 1, 2, 3, 4$. Obtenido como en la sección anterior.

$\hat{\theta}_{1j}^2$: Estimador de Muestreo Replicado en la ocasión j del periodo uno. $j = 1, 2, 3, 4$. Obtenido como en la sección anterior.

$\hat{\theta}_{1j}^3$: Estimador Combinado Univariado en la ocasión j del periodo uno. $j=1,2,3,4$.
Obtenido como en la sección anterior.

$$\hat{\theta}_{1j}^3 = k \hat{\theta}_{1j}^1 + (1-k) \hat{\theta}_{1j}^2 \quad (4.121)$$

El superíndice indica el tipo de estimador y el primer subíndice se refiere al periodo uno.

$\hat{\theta}_{2j}^1$: Estimador de Mínimos Cuadrados Generalizados en la ocasión j del periodo dos. Obtenido como en la sección anterior.

$\hat{\theta}_{2j}^2$: Estimador de Muestreo Replicado en la ocasión j del periodo dos. Obtenido como en la sección anterior.

$\hat{\theta}_{2j}^3$: Estimador Combinado Univariado en la ocasión j del periodo dos. Obtenido como en la sección anterior.

$$\hat{\theta}_{2j}^3 = k \hat{\theta}_{2j}^1 + (1-k) \hat{\theta}_{2j}^2 \quad (4.122)$$

El superíndice indica el tipo de estimador y el primer subíndice se refiere al periodo dos.

$\hat{\theta}^{-c1}$: Estimador de Mínimos Cuadrados Generalizados para el cambio entre puntos medios de periodo.

$$\hat{\theta}^{-c1} = \hat{\theta}_2^{-1} - \hat{\theta}_1^{-1} = \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{2j}^1 - \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{1j}^1 \quad (4.123)$$

y

$$v(\hat{\theta}^{-c1}) = v(\hat{\theta}_2^{-1}) + v(\hat{\theta}_1^{-1}) - 2 \text{cov}(\hat{\theta}_2^{-1}, \hat{\theta}_1^{-1}) \quad (4.124)$$

$\hat{\theta}^{-c2}$: Estimador de Muestreo Replicado para el cambio entre puntos medios de periodo.

$$\hat{\theta}^{-c2} = \hat{\theta}_2^{-2} - \hat{\theta}_1^{-2} = \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{2j}^2 - \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{1j}^2 \quad (4.125)$$

y

$$v(\hat{\theta}^{-c2}) = v(\hat{\theta}_2^{-2}) + v(\hat{\theta}_1^{-2}) - 2 \text{cov}(\hat{\theta}_2^{-2}, \hat{\theta}_1^{-2}) \quad (4.126)$$

$\hat{\theta}^{-c3}$: Estimador Combinado Univariado para el cambio entre puntos medios de periodo.

$$\hat{\theta}^{-c3} = \hat{\theta}_2^{-3} - \hat{\theta}_1^{-3} = \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{2j}^{-3} - \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{1j}^{-3} \quad (4.127)$$

y

$$v(\hat{\theta}^{-c3}) = v(\hat{\theta}_2^{-3}) + v(\hat{\theta}_1^{-3}) - 2 \text{cov}(\hat{\theta}_2^{-3}, \hat{\theta}_1^{-3}) \quad (4.128)$$

Si se supone una población teórica y el procedimiento de muestreo se repite un número grande de veces (R), entonces se puede conocer el valor esperado de cada estimador de cambio y su varianza verdadera:

$$E(\hat{\theta}^{-cl}) = \hat{\theta}^{-cl} = \frac{\sum_{r=1}^R \hat{\theta}_r^{-cl}}{R} \quad (4.129)$$

$$V(\hat{\theta}^{-cl}) = \frac{\sum_{r=1}^R \left(\hat{\theta}_r^{-cl} - \hat{\theta}^{-cl} \right)^2}{R} \quad (4.130)$$

La última expresión puede obtenerse alternativamente de la siguiente manera:

$$V(\hat{\theta}^{-cl}) = V(\hat{\theta}_2^{-l}) + V(\hat{\theta}_1^{-l}) - 2 \text{Cov}(\hat{\theta}_2^{-l}, \hat{\theta}_1^{-l}) \quad (4.131)$$

Los componentes de la parte derecha de (4.131) son dados por (Ross, 2006, pág. 356-357):

$$V(\hat{\theta}_2^{-l}) = V\left(\frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{2j}^{-l}\right) = \frac{1}{16} V\left(\sum_{j=1}^4 \hat{\theta}_{2j}^{-l}\right) = \frac{1}{16} \left(\sum_{j=1}^4 v(\hat{\theta}_{2j}^{-l}) + 2 \sum_{j < j'} \text{Cov}(\hat{\theta}_{2j}^{-l}, \hat{\theta}_{2j'}^{-l}) \right) \quad (4.132)$$

$$V(\hat{\theta}_1^{-l}) = V\left(\frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{1j}^{-l}\right) = \frac{1}{16} V\left(\sum_{j=1}^4 \hat{\theta}_{1j}^{-l}\right) = \frac{1}{16} \left(\sum_{j=1}^4 v(\hat{\theta}_{1j}^{-l}) + 2 \sum_{j < j'} \text{Cov}(\hat{\theta}_{1j}^{-l}, \hat{\theta}_{1j'}^{-l}) \right) \quad (4.133)$$

$$\text{Cov}(\hat{\theta}_2^{-l}, \hat{\theta}_1^{-l}) = \text{Cov}\left(\frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{2j}^{-l}, \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{1j}^{-l}\right) = \frac{1}{16} \text{Cov}\left(\sum_{j=1}^4 \hat{\theta}_{2j}^{-l}, \sum_{j=1}^4 \hat{\theta}_{1j}^{-l}\right) = \frac{1}{16} \sum_{j=1}^4 \sum_{j=1}^4 \text{Cov}(\hat{\theta}_{2j}^{-l}, \hat{\theta}_{1j}^{-l}) \quad (4.134)$$

Las expresiones (4.129) y (4.130) pueden ser aplicadas también cuando el interés es la tasa de cambio entre los dos periodos (punto medio del periodo dos respecto al punto medio del periodo uno). Entonces, para la repetición r del estimador l se tiene la siguiente expresión.

$$\hat{\theta}_r^{-l} = \frac{\hat{\theta}_{2r}^{-l}}{\hat{\theta}_{1r}^{-l}} = \frac{\frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{2rj}^{-l}}{\frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{1rj}^{-l}} \quad (4.135)$$

En consecuencia, el valor esperado y la varianza de (4.135) son, respectivamente,

$$E(\hat{\theta}^{-l}) = \hat{\theta}^{-l} = \frac{\sum_{r=1}^R \hat{\theta}_r^{-l}}{R} \quad (4.136)$$

$$V(\hat{\theta}^{-l}) = \frac{\sum_{r=1}^R \left(\hat{\theta}_r^{-l} - \hat{\theta}^{-l} \right)^2}{R} \quad (4.137)$$

Donde, $l = 1, 2, 3$; 1: MCG, 2: MRE, 3: ECU

Cuando solamente se tiene una muestra panelizada, es decir, no se tiene muestreo repetido, las estimaciones en cada ocasión, así como las estimaciones del cambio y la razón de cambio, y sus respectivas varianzas se pueden obtener empleando bootstrap.

4.6. MATERIALES Y MÉTODOS

Al considerar muestreo repetido en una pseudo-población forestal artificial se estudiaron desde el punto de vista teórico las propiedades estadísticas de los siguientes estimadores:

- Estimador MCG factible $\hat{\theta}_j^1$ (Estimador del parámetro poblacional en cada ocasión de medición j , incluye la actual). $j=1,2,3,4,5,6,7,8$. Corresponde al elemento j del siguiente vector:

$$\hat{T}_y = \left[\begin{array}{c|c} D'\hat{\Sigma}^{-1} & D \end{array} \right]^{-1} D'\hat{\Sigma}^{-1} \hat{T} \quad (4.112)$$

- Estimador MRE $\hat{\theta}_j^2$ (Estimador del parámetro poblacional en cada ocasión j , incluye la actual). $j=1,2,3,4,5,6,7,8$

$$\hat{T}_{y\pi,rep}(t_j) = \frac{1}{m_j} \sum_{s=1}^{m_j} \hat{T}_{y\pi s}(t) \quad (4.114)$$

- Estimador ECU $\hat{\theta}_j^3$ (Estimador del parámetro poblacional en cada ocasión j , incluye la actual). $j=1,2,3,4,5,6,7,8$

$$\hat{\theta}_j^3 = k \hat{\theta}_j^1 + (1-k) \hat{\theta}_j^2 \quad (4.116)$$

- Estimador de MCG para el cambio entre puntos medios de periodo.

$$\bar{\theta}^{c1} = \bar{\theta}_2^{-1} - \bar{\theta}_1^{-1} = \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{2j}^1 - \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{1j}^1 \quad (4.123)$$

- Estimador de MRE para el cambio entre puntos medios de periodo.

$$\bar{\theta}^{c2} = \bar{\theta}_2^{-2} - \bar{\theta}_1^{-2} = \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{2j}^2 - \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{1j}^2 \quad (4.125)$$

- ECU para el cambio entre puntos medios de periodo.

$$\hat{\theta}^{-c3} = \hat{\theta}_2^{-3} - \hat{\theta}_1^{-3} = \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{2j}^3 - \frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{1j}^3 \quad (4.127)$$

- Razón de cambio entre puntos medios de periodo.

$$\hat{\theta}_r^{-l} = \frac{\hat{\theta}_{2r}^{-l}}{\hat{\theta}_{1r}^{-l}} = \frac{\frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{2rj}^l}{\frac{1}{4} \sum_{j=1}^4 \hat{\theta}_{1rj}^l} \quad (4.135)$$

Donde, $l = 1, 2, 3$

- Estimador de MM (Estimador del parámetro poblacional actual)

$$\hat{T}_{MM,rep} = \sum_{j=1}^P w_j \hat{T}_{y\pi,rep}(t_j) \quad (4.97)$$

Debido a que en el estudio se consideró como parámetro el área basal por hectárea, entonces en las expresiones anteriores $\hat{T}_{y\pi,rep}$ se sustituye por:

$$\hat{\lambda}_{y\pi,rep} = \frac{\hat{T}_{y\pi,rep}}{A} \quad ((2.14), \text{Capítulo II})$$

La pseudo-población forestal artificial de conglomerados se generó combinando las bases de datos de Durango, Hidalgo, Jalisco, Michoacán, Estado de México, Chihuahua y Quintana Roo. De esta forma, la pseudo-población se constituyó por conglomerados de bosques y de selvas. Esta información fue proporcionada por la Comisión Nacional Forestal y proviene de muestras panelizadas correspondientes a las ocasiones de medición 2004, 2005, 2006, 2007, 2009, 2010, 2011 y 2012. La pseudo-población estuvo integrada por 7,597 conglomerados con estructura panelizada, es decir, grupos específicos distribuidos en cada una de las ocho ocasiones de medición.

Johnson *et al.* (2003) señalan que esta forma de crear la población garantiza que se obtengan tendencias y varianzas realísticas de las variables de interés seleccionadas, área basal en este caso.

De la pseudo-población generada se tomaron 100 muestras aleatorias (repeticiones) en cada uno de dos tipos de condiciones o escenarios diferentes. Cada condición se definió en base al número de conglomerados forestales panelizados de los estados que integraron la pseudo-población. De esta manera, se consideraron tamaños de muestras panelizadas de 1000 conglomerados (condición 1), y de 1500 (condición 2).

El parámetro de interés fue el área basal promedio por hectárea (ABHA) en cada una de las ocho ocasiones de medición, así como el cambio y la razón de cambio de área basal por hectárea entre puntos medios de periodo. Los valores verdaderos de ABHA fueron 12.8074, 12.4569, 11.4169, 11.1347, 12.5070, 13.1689, 13.2585 y 13.0764 ($\text{m}^2 \text{ha}^{-1}$), para 2004, 2005, 2006, 2007, 2009, 2010, 2011 y 2012, respectivamente. El cambio y la tasa del cambio verdadero entre los puntos medios de periodo fueron respectivamente, $1.05 \text{ m}^2 \text{ha}^{-1}$ y 1.09.

El análisis conjunto de las 100 muestras aleatorias, en cada condición, permitió observar el comportamiento de los estimadores estudiados bajo el concepto de muestreo repetido. Para determinar el mejor estimador (l) en cada condición estudiada (c) se usó el Sesgo Relativo (SER), la Varianza y la Raíz del Cuadrado Medio del Error Relativo (RaizCMEREL) (Gregoire y Valentine, 2008):

$$E(\hat{\theta}_j^l) = \bar{\hat{\theta}}_j^l = \frac{\sum_{r=1}^R \hat{\theta}_{rj}^l}{R} : \text{valor esperado} \quad (4.119)$$

$$B(\hat{\theta}_j^l) = (\bar{\hat{\theta}}_j^l - \theta_j) : \text{sesgo}$$

$$SRE(\hat{\theta}_j^l) = \left(\frac{\bar{\hat{\theta}}_j^l - \theta_j}{\theta_j} \right) * 100 : \text{sesgo relativo}$$

$$V(\hat{\theta}_j^l) = \frac{\sum_{r=1}^R \left(\hat{\theta}_{rj}^l - \bar{\hat{\theta}}_j^l \right)^2}{R} : \text{varianza} \quad (4.120)$$

$$RaizCMEREL(\hat{\theta}_j^l) = \left(\frac{\sqrt{V(\hat{\theta}_j^l) + (B(\hat{\theta}_j^l))^2}}{\theta_j} \right) * 100 : \text{Raíz del Cuadrado Medio del Error}$$

Relativo

θ_j : Área basal por ha ($m^2 ha^{-1}$) en cada ocasión (parámetro).

Las muestras panelizadas que constituyeron la pseudo-población forestal artificial se representan en la Figura 4.7. Se generaron 40 subpaneles distribuidos en las ocho ocasiones de medición. Dado que se tienen ocho ocasiones de medición y 40 subpaneles se utilizó la matriz diseño mostrada en la Figura 4.8.

Con el propósito de obtener un valor de magnitud apreciable, la estimación del cambio de ABHA se obtuvo entre periodos. Para tal efecto, en el periodo uno se consideraron las ocasiones de medición 2004 a 2007, y en el periodo dos de 2009 a 2012. La media del periodo uno se obtuvo como el promedio de la media de la ocasión 2004, 2005, 2006 y 2007. La media del periodo dos se calculó como el promedio de la media de la ocasión 2009, 2010, 2011 y 2012. El cambio entre periodos se estimó como la diferencia entre las medias de los dos periodos. Este procedimiento se aplicó para cada tipo de estimador y en cada condición. Para generar los valores verdaderos del cambio entre periodos se promediaron los parámetros de las ocasiones correspondientes a cada periodo. En todos los casos la media del periodo uno fue sustraída de la media del periodo dos. De tal manera que un valor positivo indicó incremento de ABHA entre el punto medio de los dos periodos. Para determinar el mejor estimador (l) en cada condición estudiada (c) también se usó el Sesgo Relativo (SER), la Varianza y la Raíz del Cuadrado Medio del Error Relativo (RaizCMEREL) (Gregoire y Valentine, 2008).

Grupo general	2004	2005	2006	2007	2009	2010	2011	2012
	1	2	3	4	5	6	7	8
1	11							
2	21				25			
3	31					36		
4	41						47	
5	51							58
6		62						
7		72			75			
8		82				86		
9		92					97	
10		102						108
11			113					
12			123		125			
13			133			136		
14			143				147	
15			153					158
16				164				
17				174	175			
18					185			
19				194		196		
20						206		
21				214			217	
22							227	
23				234				238
24								248

Figura 4.7. Muestras panelizadas que constituyeron la pseudo-población forestal artificial. Los números en el interior de la Figura indican el grupo general y la ocasión de medición. Así por ejemplo, el 62 indica que la muestra panelizada corresponde al grupo general 6 y a la ocasión de medición 2, mientras que el 248, indica que la muestra panelizada pertenece al grupo general 24 y a la ocasión de medición 8.

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figura 4.8. Matriz diseño utilizada en el proceso de estimación de Mínimos Cuadrados Generalizados. El número 1 indica la ocasión en que el subpanel fue medido.

$$E(\hat{\theta}^{\bar{cl}}) = \hat{\theta}^{\bar{cl}} = \frac{\sum_{r=1}^R \hat{\theta}_r^{\bar{cl}}}{R} : \text{valor esperado} \quad (4.129)$$

$$B(\hat{\theta}^{\bar{cl}}) = (\hat{\theta}^{\bar{cl}} - \bar{\theta}^{\bar{c}}) : \text{sesgo}$$

$$SRE(\hat{\theta}^{\bar{cl}}) = \left(\frac{\hat{\theta}^{\bar{cl}} - \bar{\theta}^{\bar{c}}}{\bar{\theta}^{\bar{c}}} \right) * 100 : \text{sesgo relativo}$$

$$V(\hat{\theta}^{\bar{cl}}) = \frac{\sum_{r=1}^R \left(\hat{\theta}_r^{\bar{cl}} - \hat{\theta}^{\bar{cl}} \right)^2}{R} : \text{varianza} \quad (4.130)$$

$$RaizCMEREL(\hat{\theta}^{\bar{cl}}) = \left(\frac{\sqrt{V(\hat{\theta}^{\bar{cl}}) + (B(\hat{\theta}^{\bar{cl}}))^2}}{\bar{\theta}^{\bar{c}}} \right) * 100 : \text{Raíz del Cuadrado Medio del Error}$$

Relativo

$\bar{\theta}^{\bar{c}}$: Cambio de área basal por ha (m^2ha^{-1}) entre dos puntos medios de periodos (parámetro).

Para seleccionar el mejor estimador de la tasa de cambio entre puntos medios de periodo se emplearon expresiones similares a las anteriores, sólo que en este caso se consideraron las ecuaciones (4.136) y (4.137) para el valor esperado y la varianza, respectivamente.

Con el fin de mejorar las propiedades estadísticas de MCG (4.112) y MRE (4.114), en cada ocasión de medición y en cada condición de estudio se aplicó el estimador compuesto univariado ECU (4.116), en el cual se consideró MCG como el estimador más preciso por tener la menor varianza. Dada la posible correlación entre MCG y MRE, en el factor de ponderación se consideró la covarianza entre ambos estimadores (4.118).

Para el caso de la media móvil se aplicaron los pesos presentados en el Cuadro 4.1. En el estimador MM-FIA0 se considera el mismo peso para cada ocasión de

medición. El promedio móvil con pesos iguales se emplea en el programa FIA-USA (McRoberts *et al.*, 2005). En el resto, las ponderaciones se establecieron de acuerdo al criterio sugerido por Eskelson *et al.* (2009), de tal manera que la suma de los pesos correspondientes a las ocasiones de medición del segundo ciclo de inventario fuera al menos del 0.5. El término BIA obedece a que el primer autor de Eskelson *et al.* (2009) es Bianca N.I. Eskelson.

Cuadro 4.1. Pesos considerados en la aplicación de media móvil.

ESTIMADOR	2004	2005	2006	2007	2009	2010	2011	2012
MM-FIA0	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
MM-BIA1	0.10	0.10	0.10	0.10	0.15	0.15	0.15	0.15
MM-BIA2	0.00	0.10	0.10	0.10	0.15	0.15	0.20	0.20
MM-BIA3	0.00	0.00	0.10	0.10	0.15	0.15	0.25	0.25
MM-BIA4	0.00	0.00	0.00	0.10	0.15	0.15	0.30	0.30

El análisis estadístico, incluyendo la selección de las muestras aleatorias, se realizó en SAS (v9) (SAS, 2009). Para la implementación de Mínimos Cuadrados Generalizados se utilizó el PROC IML (SAS, 2010). La modelación gráfica de la normalidad se llevó en R.

4.7. RESULTADOS Y DISCUSIÓN

4.7.1. Estimación de ABHA en cada ocasión de medición

Las estimaciones de área basal por hectárea —obtenidas mediante el método de MCG, estimadores de MRE y el ECU— se presentan en los Cuadros 4.2 y 4.3. Estos resultados corresponden a las ocasiones de medición 2004, 2005, 2006, 2007, 2009, 2010, 2011 y 2012 considerando los tamaños de muestra panelizada de 1000 y 1500 conglomerados, condición uno y dos, respectivamente.

En las dos condiciones de estudio y para cada una de las ocho ocasiones de medición se observa que el ABHA de MCG es ligeramente sesgada en comparación con el ABHA de MRE. En la condición uno el sesgo relativo mayor fue de -5.05% correspondiente a 2004, mientras que en la condición dos fue de 4.68% para 2007; por lo que el desvío de MCG respecto al parámetro puede considerarse insignificante. Salvo en 2006 y 2007 de ambas condiciones, el sesgo de MCG se manifiesta es una subestimación (Figura 4.9 y 4.10).

En contraste, el sesgo relativo en MRE en ningún caso fue mayor al 1%; inclusive en 2007 y 2012 de la condición uno, y 2011 y 2012 de la condición dos, la estimación y el valor verdadero de la población son prácticamente los mismos. Así, las estimaciones obtenidas mediante MRE se consideran insesgadas (equitativamente por arriba y por abajo del parámetro (Figura 4.9 y 4.10)). El ECU por su parte generó en cada una de las ocasiones un sesgo relativo intermedio entre MCG y MRE. En este caso el sesgo relativo mayor en la condición uno fue cercano al -3%, es decir, la distancia entre la estimación y el parámetro se redujo en dos unidades porcentuales comparada con MCG. En la condición dos el sesgo relativo mayor del ECU fue del 3.33%; 1.35% menor que el sesgo relativo de MCG.

Cuadro 4.2. Estimación de área basal por hectárea (ABHA) ($m^2 ha^{-1}$) mediante Mínimos Cuadrados Generalizados, Muestreo Replicado y Estimador Combinado Univariado bajo la condición uno (tamaño de muestras panelizadas de 1000).

	Ocasión de medición							
	2004	2005	2006	2007	2009	2010	2011	2012
Condición 1/Estimador de Mínimos Cuadrados Generalizados								
Estimación	12.16	11.85	11.51	11.49	11.98	12.63	12.81	12.70
Parámetro	12.81	12.46	11.42	11.13	12.51	13.17	13.26	13.08
Sesgo	-0.65	-0.60	0.10	0.35	-0.52	-0.54	-0.45	-0.37
Sesgo relativo (%)	-5.05	-4.83	0.86	3.18	-4.18	-4.06	-3.41	-2.85
Varianza	1.33	0.12	0.18	0.17	0.15	0.27	0.25	0.26
CME	1.75	0.49	0.19	0.30	0.43	0.56	0.46	0.40
RaizCME	1.32	0.70	0.43	0.54	0.65	0.75	0.68	0.63
RaizCMEREL (%)	10.32	5.60	3.81	4.89	5.22	5.68	5.09	4.85
Condición 1/Estimador de Muestreo Replicado								
Estimación	12.79	12.40	11.39	11.14	12.46	13.12	13.17	13.07
Parámetro	12.81	12.46	11.42	11.13	12.51	13.17	13.26	13.08
Sesgo	-0.02	-0.05	-0.03	0.003	-0.05	-0.05	-0.09	-0.01
Sesgo relativo (%)	-0.15	-0.44	-0.26	0.03	-0.41	-0.37	-0.68	-0.04
Varianza	2.11	0.11	0.24	0.40	0.32	0.25	0.36	0.39
CME	2.11	0.11	0.24	0.40	0.32	0.26	0.37	0.39
RaizCME	1.45	0.34	0.49	0.64	0.56	0.50	0.61	0.63
RaizCMEREL (%)	11.34	2.72	4.32	5.71	4.51	3.83	4.57	4.80
Condición 1/Estimador Combinado Univariado								
Estimación	12.42	12.14	11.46	11.37	12.16	12.88	12.97	12.86
Parámetro	12.81	12.46	11.42	11.13	12.51	13.17	13.26	13.08
Sesgo	-0.38	-0.32	0.04	0.23	-0.35	-0.29	-0.29	-0.21
Sesgo relativo (%)	-2.99	-2.57	0.35	2.11	-2.78	-2.17	-2.20	-1.64
Varianza	1.17	0.09	0.17	0.16	0.14	0.20	0.23	0.22
CME	1.32	0.20	0.17	0.21	0.26	0.28	0.32	0.27
RaizCME	1.15	0.44	0.42	0.46	0.51	0.53	0.56	0.52
RaizCMEREL (%)	8.97	3.57	3.65	4.13	4.11	4.00	4.23	3.98

Cuadro 4.3. Estimación de área basal por hectárea (ABHA) ($m^2 ha^{-1}$) mediante Mínimos Cuadrados Generalizados, Muestreo Replicado y Estimador Combinado Univariado bajo la condición dos (tamaño de muestras panelizadas de 1500).

	Ocasión de medición							
	2004	2005	2006	2007	2009	2010	2011	2012
Condición 2/Mínimos Cuadrados Generalizados								
Estimación	12.28	12.06	11.68	11.66	12.20	12.91	13.07	12.80
Parámetro	12.81	12.46	11.42	11.13	12.51	13.17	13.26	13.08
Sesgo	-0.53	-0.40	0.26	0.52	-0.31	-0.26	-0.19	-0.28
Sesgo relativo (%)	-4.11	-3.21	2.30	4.68	-2.46	-2.00	-1.40	-2.12
Varianza	0.43	0.05	0.10	0.12	0.09	0.09	0.08	0.13
CME	0.71	0.21	0.17	0.39	0.18	0.16	0.12	0.21
RaizCME	0.84	0.45	0.41	0.62	0.42	0.40	0.34	0.46
RaizCMEREL (%)	6.58	3.65	3.56	5.60	3.39	3.07	2.58	3.51
Condición 2/Muestreo Replicado								
Estimación	12.87	12.52	11.46	11.15	12.57	13.25	13.26	13.08
Parámetro	12.81	12.46	11.42	11.13	12.51	13.17	13.26	13.08
Sesgo	0.07	0.06	0.04	0.02	0.06	0.09	0.002	0.01
Sesgo relativo (%)	0.53	0.52	0.39	0.18	0.48	0.65	0.01	0.05
Varianza	1.44	0.07	0.12	0.37	0.22	0.24	0.20	0.27
CME	1.45	0.07	0.12	0.37	0.22	0.25	0.20	0.27
RaizCME	1.20	0.26	0.35	0.61	0.47	0.50	0.45	0.52
RaizCMEREL (%)	9.40	2.12	3.04	5.44	3.79	3.79	3.37	3.96
Condición 2/Estimador Combinado Univariado								
Estimación	12.46	12.27	11.58	11.51	12.33	13.03	13.14	12.91
Parámetro	12.81	12.46	11.42	11.13	12.51	13.17	13.26	13.08
Sesgo	-0.35	-0.19	0.16	0.37	-0.18	-0.14	-0.12	-0.17
Sesgo relativo (%)	-2.74	-1.53	1.40	3.33	-1.42	-1.07	-0.90	-1.27
Varianza	0.47	0.05	0.09	0.12	0.10	0.10	0.09	0.14
CME	0.59	0.08	0.12	0.26	0.13	0.12	0.10	0.17
RaizCME	0.77	0.29	0.35	0.51	0.36	0.35	0.32	0.41
RaizCMEREL (%)	6.00	2.31	3.04	4.54	2.87	2.68	2.39	3.15

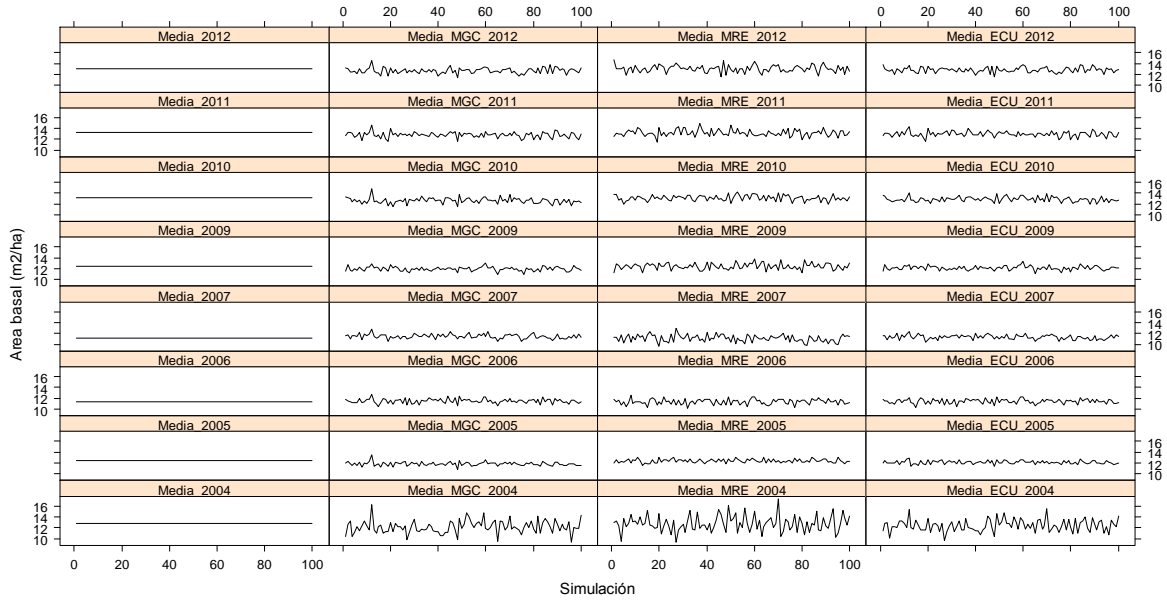


Figura 4.9. Estimaciones de ABHA obtenidas mediante MCG, MRE y ECU en la condición uno. El sesgo de mayor magnitud para MCG ocurre en 2004 (-5.05%). Las estimaciones de la ocasión 2004 presentan mayor variabilidad comparadas con los valores del 2005. Las estimaciones amplificadas para cada ocasión en la condición uno se presentan el ANEXO 4.4.

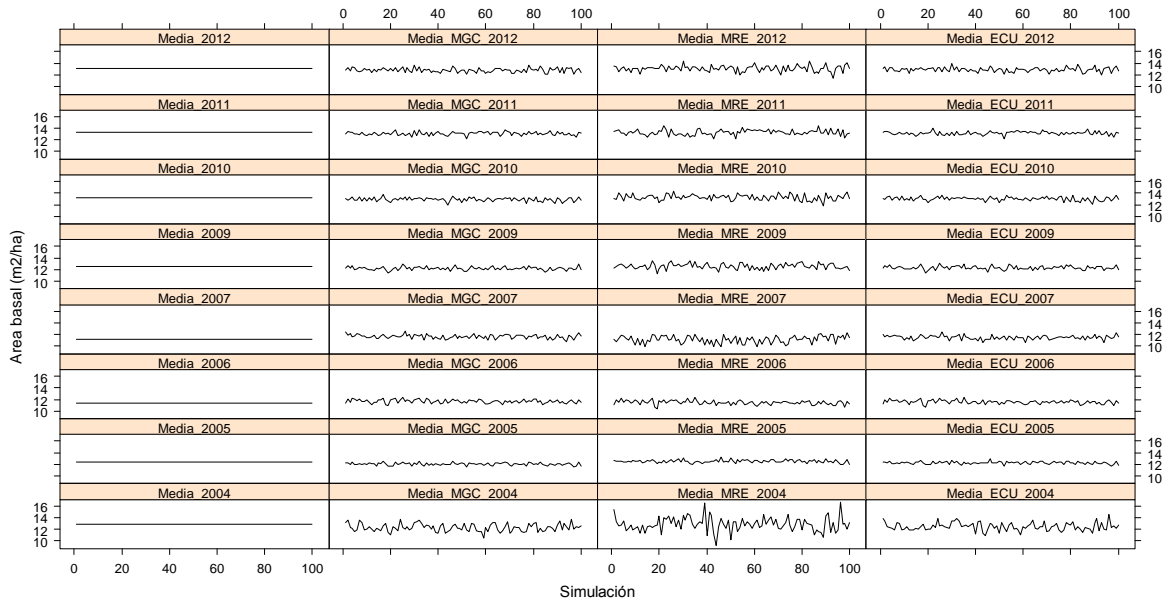


Figura 4.10. Estimaciones de ABHA obtenidas mediante MCG, MRE y ECU en la condición dos. El sesgo de mayor magnitud ocurre en 2007 para MCG (4.68%). Las estimaciones de la ocasión 2004 presentan mayor variabilidad comparadas con los valores del 2005. Las estimaciones amplificadas para cada ocasión en la condición dos se presentan el ANEXO 4.5.

A pesar de que MCG subestima ligeramente el parámetro, también presenta una varianza menor comparada con MRE, a excepción del 2005 y 2010 de la condición uno. La proporción de la varianza de MCG respecto a MRE fue alrededor del 70% en la condición uno y cerca del 50% en la condición dos. De esta manera, las estimaciones de ABHA de MCG se encuentran concentradas alrededor de su promedio, mientras que los estimados de MRE presentan desviaciones notables con respecto a su valor esperado (Figura 4.9 y 4.10).

El estimador combinado univariado además de reducir el sesgo respecto al MCG, también presenta menor varianza comparado con MRE. En la condición uno la proporción de la varianza ECU sobre MRE fue en promedio alrededor del 60% y en la condición dos del 50%. Por otro lado, en la condición dos la varianza de ECU fue en promedio casi igual que MCG, mientras que en la condición uno la varianza de ECU fue del 86% respecto a MCG. En la Figura 4.11 se muestran las funciones de densidad normal modeladas para ABHA de todas las ocasiones de medición correspondientes a la condición uno y en la Figura 4.12 para la condición dos. En ambas se observa que la media de ECU está ubicada entre la media de MCG y MRE, además, la campana de Gauss del ECU es notablemente más estrecha comparada con la MRE. En la Figura 4.11 se nota además como la varianza de ECU es ligeramente menor que MCG.

Lo anterior indica que, por una parte ECU reduce el sesgo de MCG, y por la otra reduce la varianza de MRE. De esta manera, ECU puede considerarse como un estimador insesgado (o cuasinsesgado) de mínima varianza (Czaplewski, 2010). De hecho, excepto en 2005 y 2010 de la condición uno, así como en 2005 de la condición dos, la Raíz del Cuadrado Medio del Error Relativo de ECU fue menor que MCG y MRE.

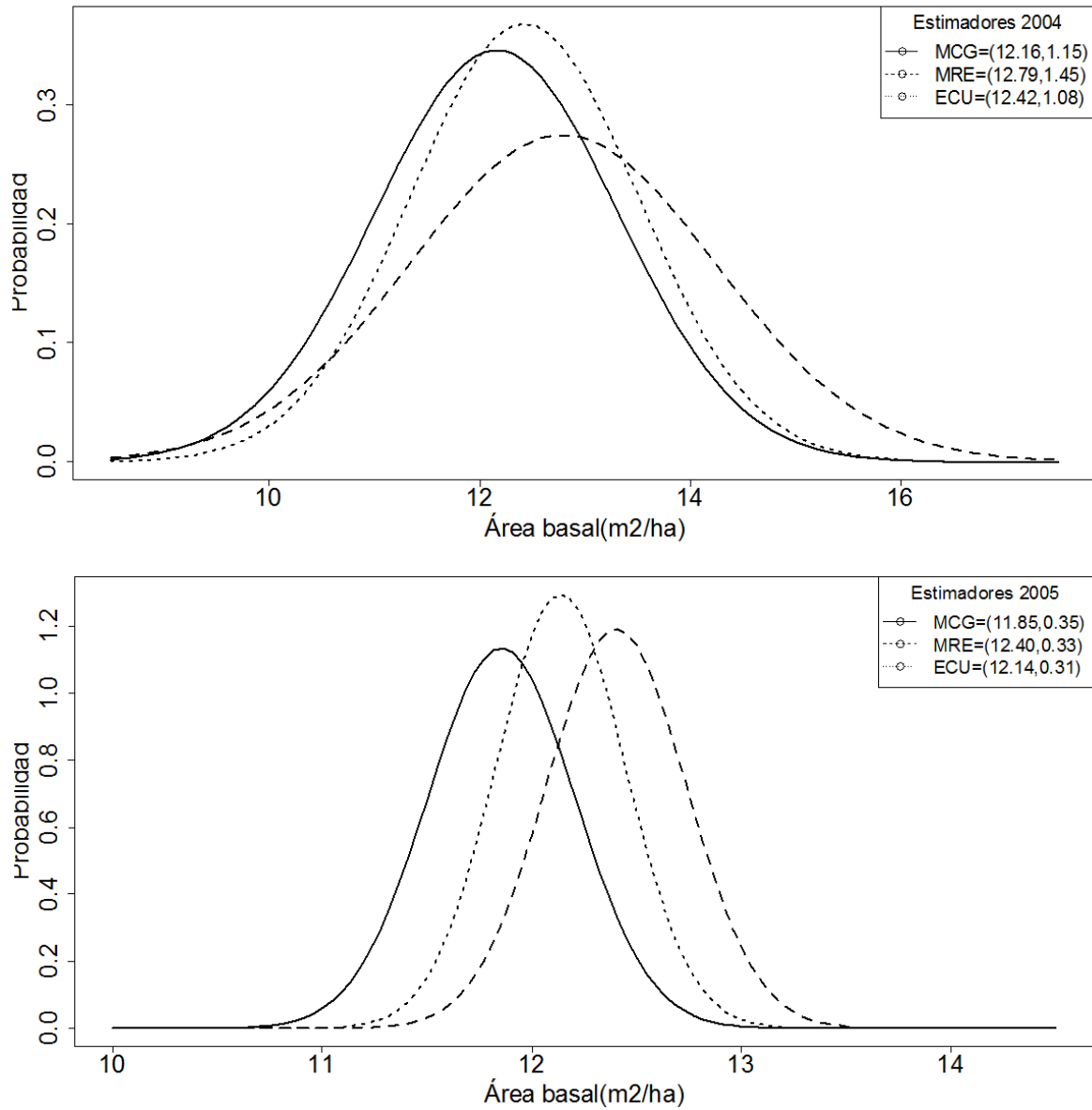


Figura 4.11a. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2004 y 2005 en la condición uno (MCG: línea sólida, MRE: línea punteada grande, MCU: línea punteada pequeña). El primer número entre paréntesis indica la media del estimador y el segundo su desviación estándar.

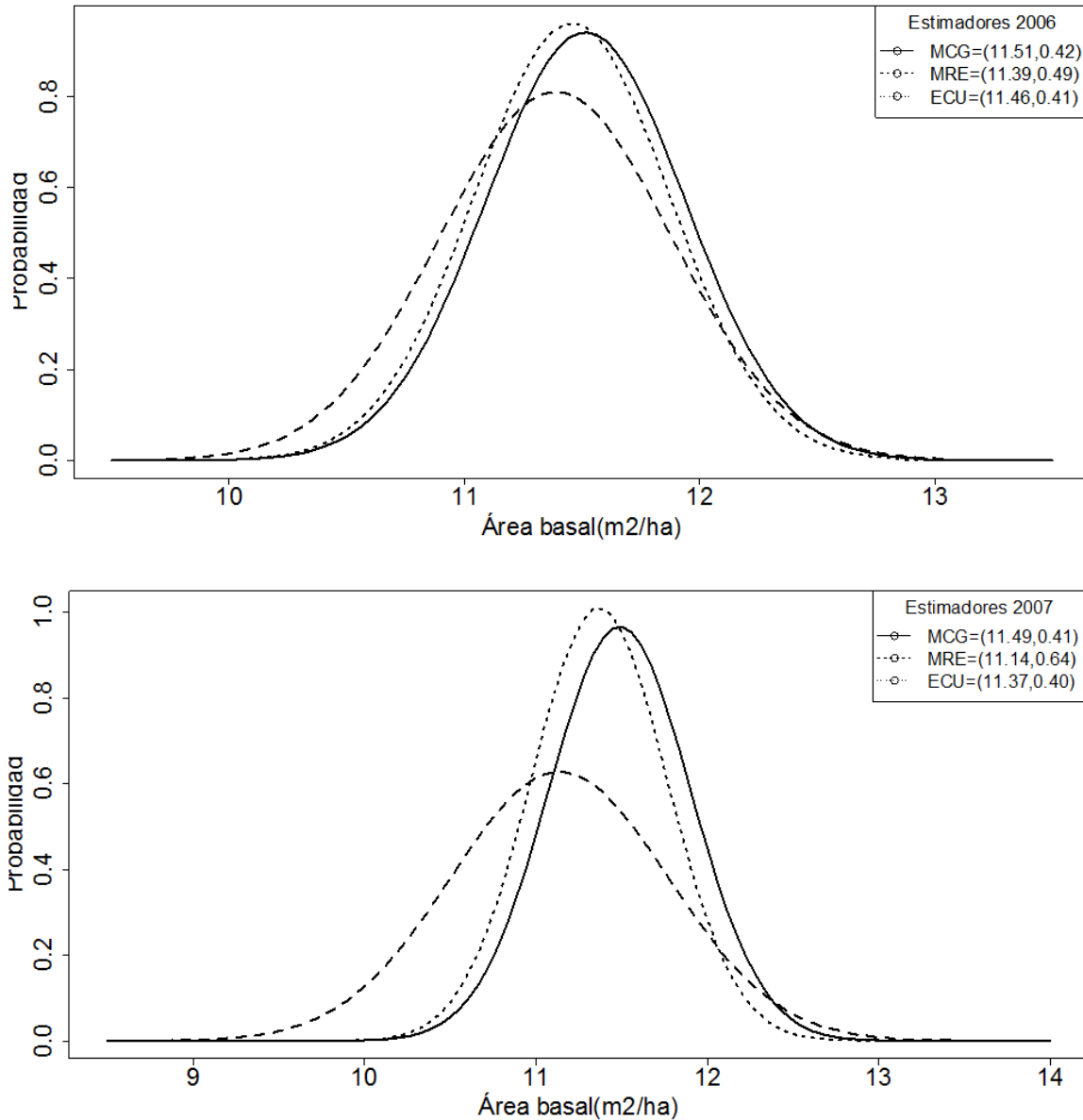


Figura 4.11b. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2006 y 2007 en la condición uno (MCG: línea sólida, MRE: línea punteada grande, MRE: línea punteada pequeña). El primer número entre paréntesis indica la media del estimador y el segundo su desviación estándar.

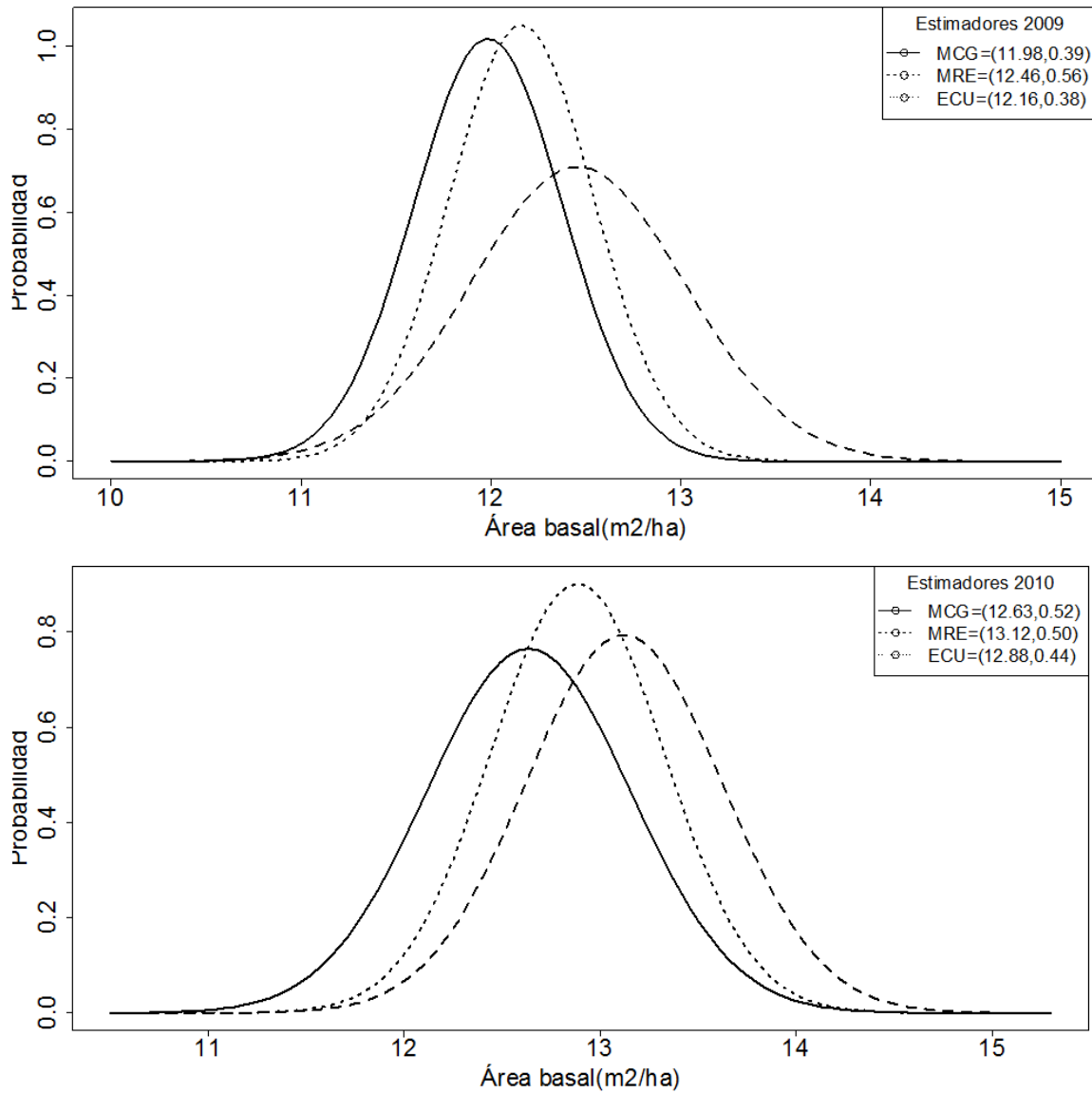


Figura 4.11c. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2009 y 2010 en la condición uno (MCG: línea sólida, MRE: línea punteada grande, ECU: línea punteada pequeña). El primer número entre paréntesis indica la media del estimador y el segundo su desviación estándar.

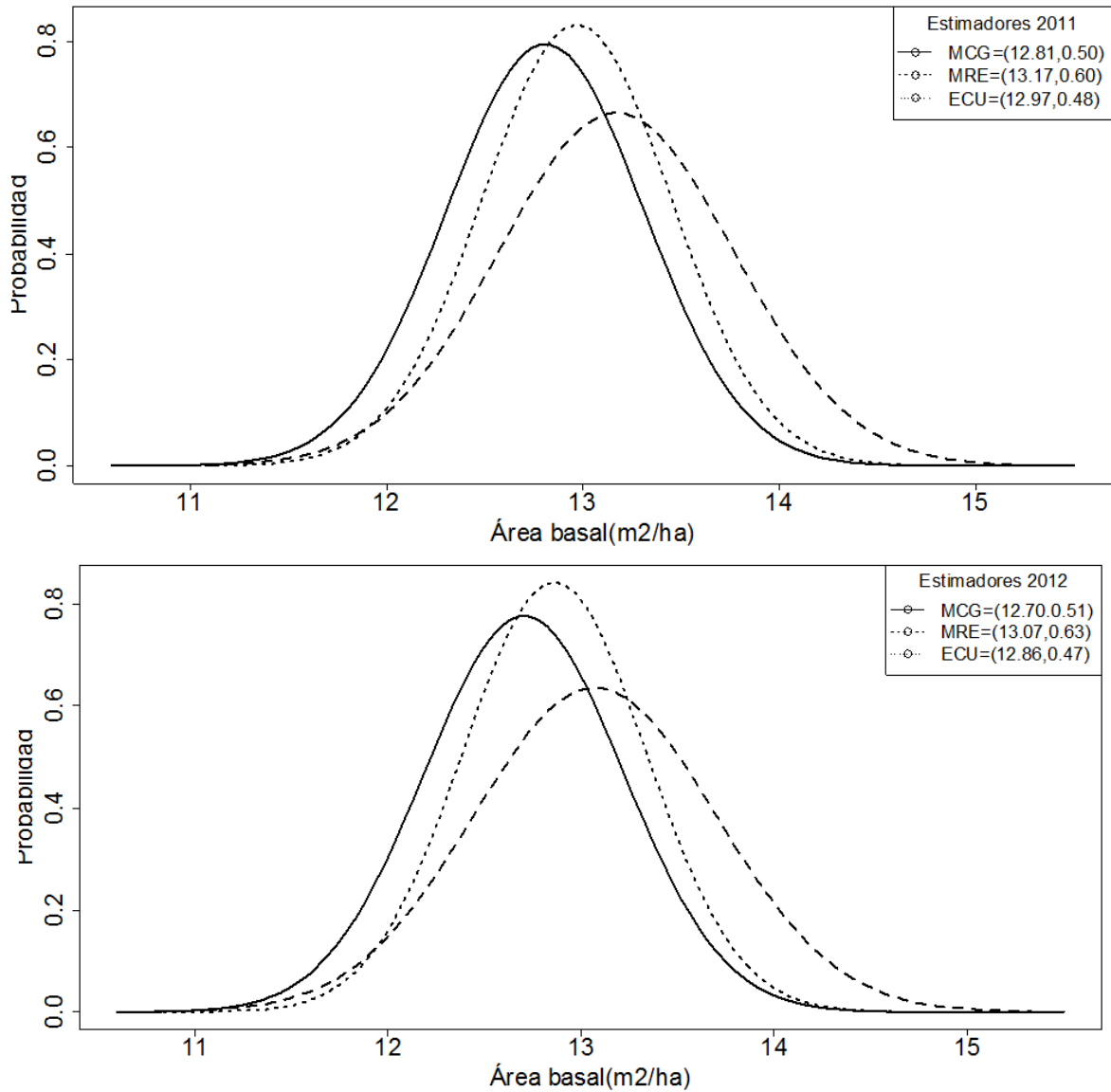


Figura 4.11d. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2011 y 2012 en la condición uno (MCG: línea sólida, MRE: línea punteada grande, MCU: línea punteada pequeña). El primer número entre paréntesis indica la media del estimador y el segundo su desviación estándar.

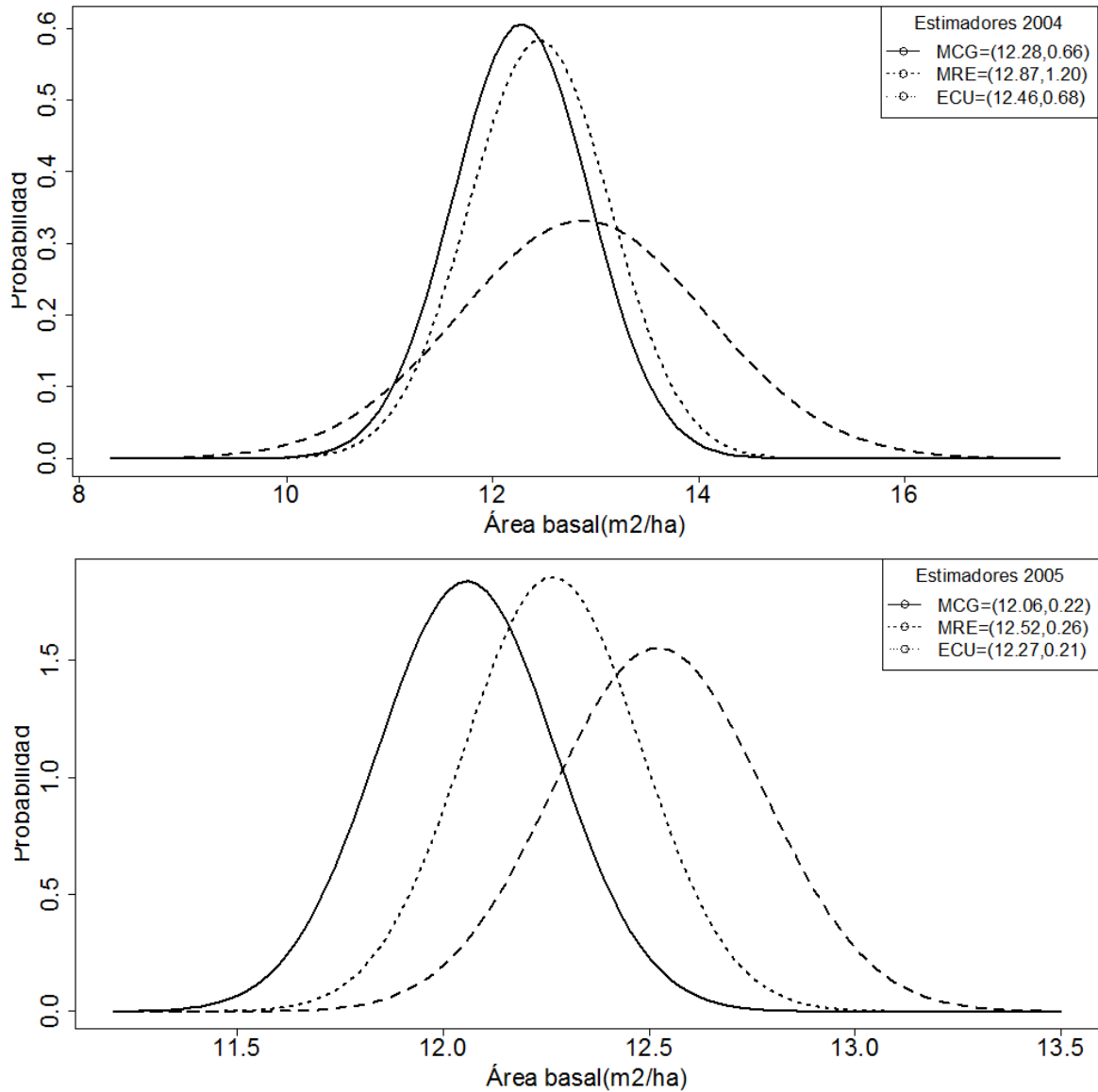


Figura 4.12a. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2004 a 2005 en la condición dos (MCG: línea sólida, MRE: línea punteada grande, MCU: línea punteada pequeña). El primer número entre paréntesis indica la media del estimador y el segundo su desviación estándar.

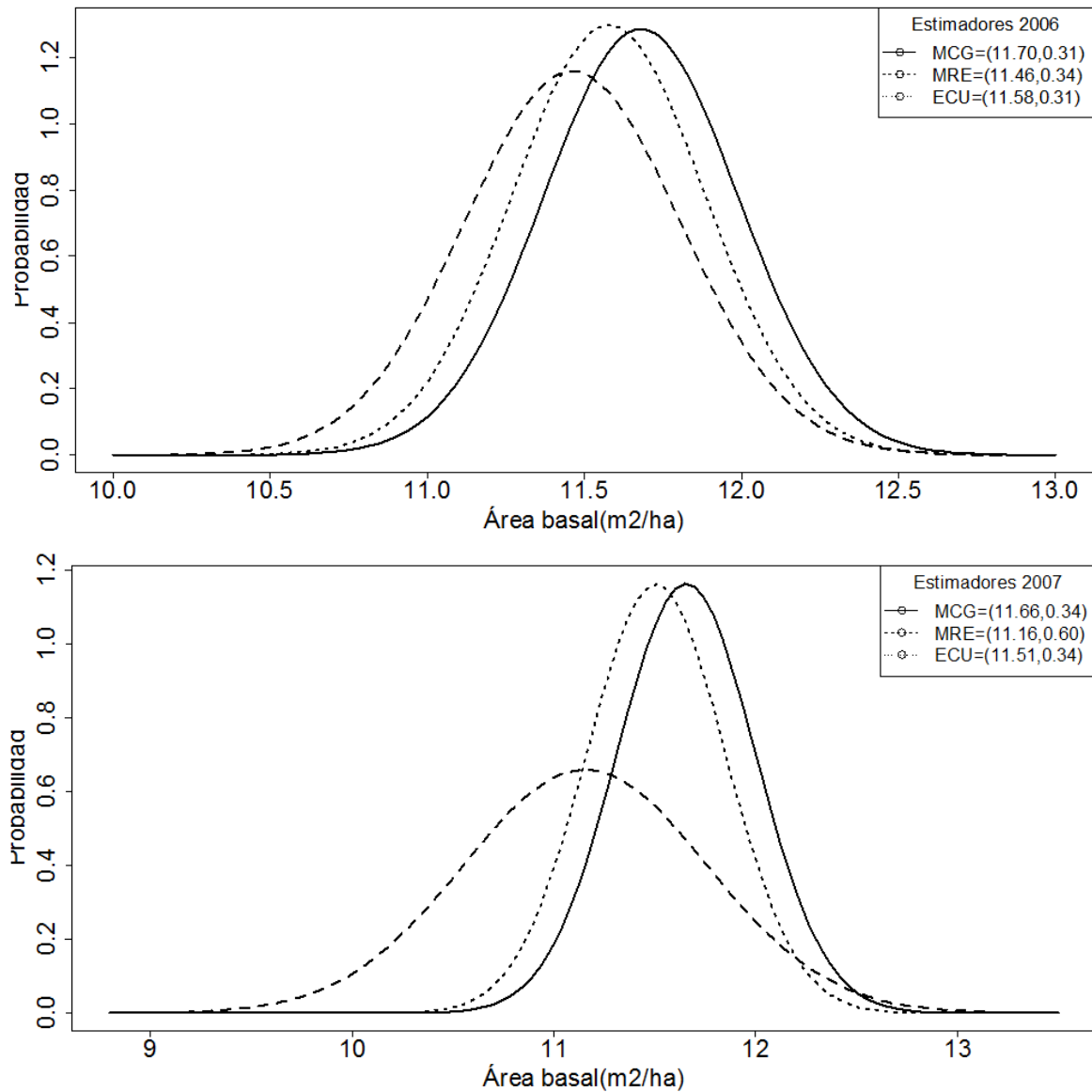


Figura 4.12b. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2006 a 2007 en la condición dos (MCG: línea sólida, MRE: línea punteada grande, MCU: línea punteada pequeña). El primer número entre paréntesis indica la media del estimador y el segundo su desviación estándar.

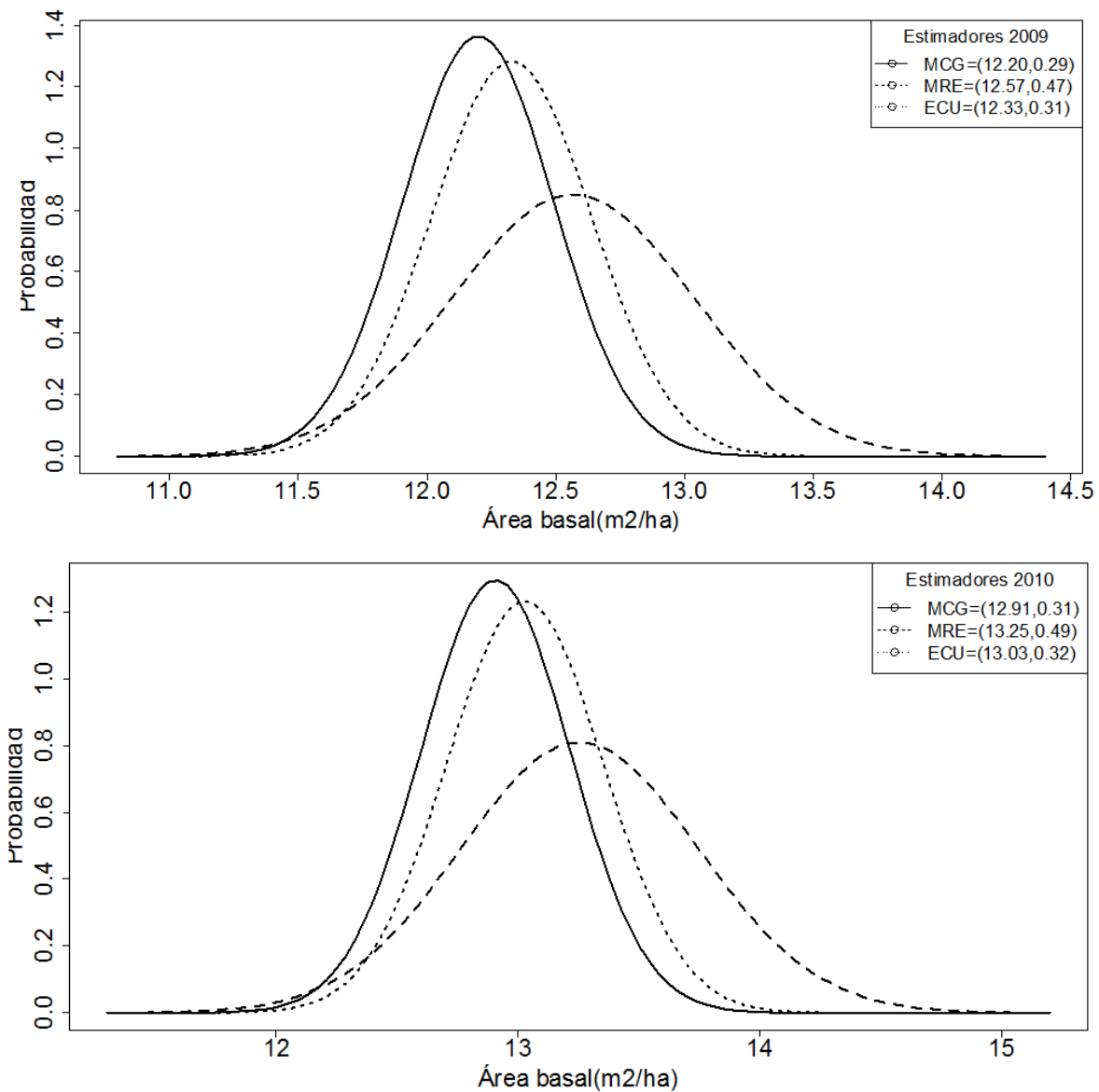


Figura 4.12c. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2009 a 2010 en la condición dos (MCG: línea sólida, MRE: línea punteada grande, ECU: línea punteada pequeña). El primer número entre paréntesis indica la media del estimador y el segundo su desviación estándar.

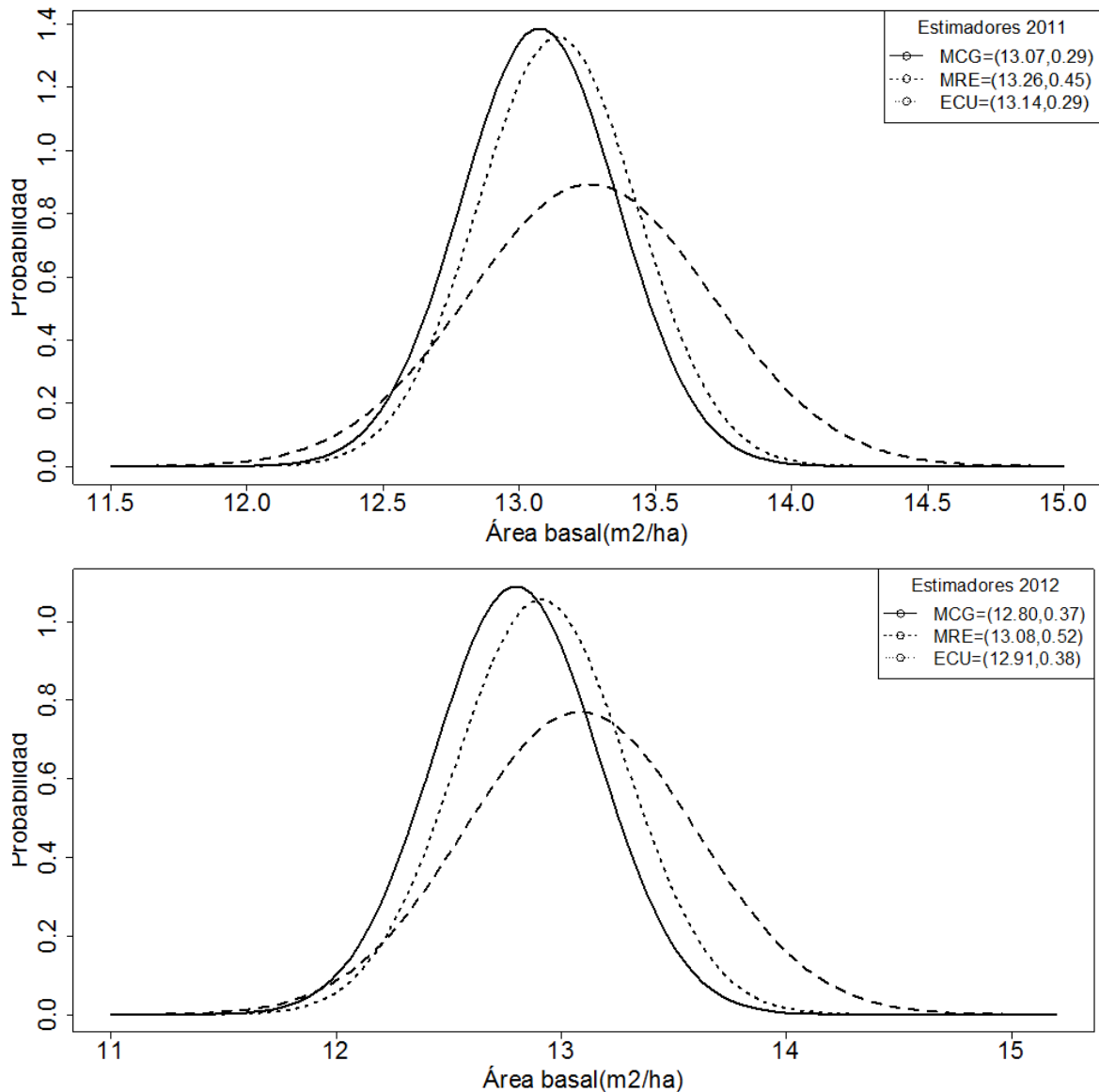


Figura 4.12d. Funciones de densidad normal modeladas de ABHA para las ocasiones de medición 2011 a 2012 en la condición dos (MCG: línea sólida, MRE: línea punteada grande, ECU: línea punteada pequeña). El primer número entre paréntesis indica la media del estimador y el segundo su desviación estándar.

En general se observa que la RaizCMEREL de ECU fue menor que MCG y MRE. En la condición uno los máximos valores de RaizCMEREL fueron de 10.32, 11.34 y 8.97% (2004) para MCG, MRE y ECU, respectivamente. En la condición

dos la RaizCMEREL fue 6.58% para MCG, 9.40% de MRE y 6% para ECU. En general, los valores de RaizCMEREL del ECU en la condición dos fueron menores que los correspondientes en la condición uno en cada ocasión. La proporción de RaizCMEREL del ECU de la condición dos respecto a la condición uno fue del 75%. Este hecho puede ser atribuido al mayor tamaño de muestras panelizadas usado en la condición dos. Esta característica, hace que ECU sea atractivo tanto desde un punto teórico como práctico.

El ligero desvío de la estimación respecto al parámetro en MCG puede deberse a que la matriz de varianzas-covarianzas está integrada por la información muestral de todos los subpaneles correspondientes a todas las ocasiones de medición, mientras que en MRE la información que se utiliza en el proceso de estimación es sólo de la ocasión de medición de interés. También la reducción de la varianza de MCG es producto de la combinación de la información de los subpaneles en el proceso de estimación. El hecho de que en la condición dos la varianza de ECU haya sido en promedio casi igual que MCG puede ser explicado porque en el ECU y su varianza se incluyó la covarianza entre los estimadores de MCG y MRE.

En los tres estimadores y en las dos condiciones estudiadas se observa que la varianza correspondiente a la ocasión 2004 siempre es mayor que el resto de las ocasiones, esto se debe al bajo número de observaciones utilizadas en la estimación. En la condición uno el número total promedio de pares 2004-2009 (04-10,04-11,04-12) fue de 26 y en la condición dos fue de 39. Por el contrario, el número total promedio de pares 2005-2009 (05-10,05-11,05-12) fue de 428 en la condición uno y de 640 en la condición dos.

En base a lo anterior, parece ser que ECU es el mejor de los tres estimadores para obtener ABHA en cada ocasión de medición. En promedio, el 91% de los intervalos de confianza al 95% generados por ECU atrapan el valor verdadero en la condición uno (Figura 4.13), y el 90% en la condición dos (Figura 4.14).

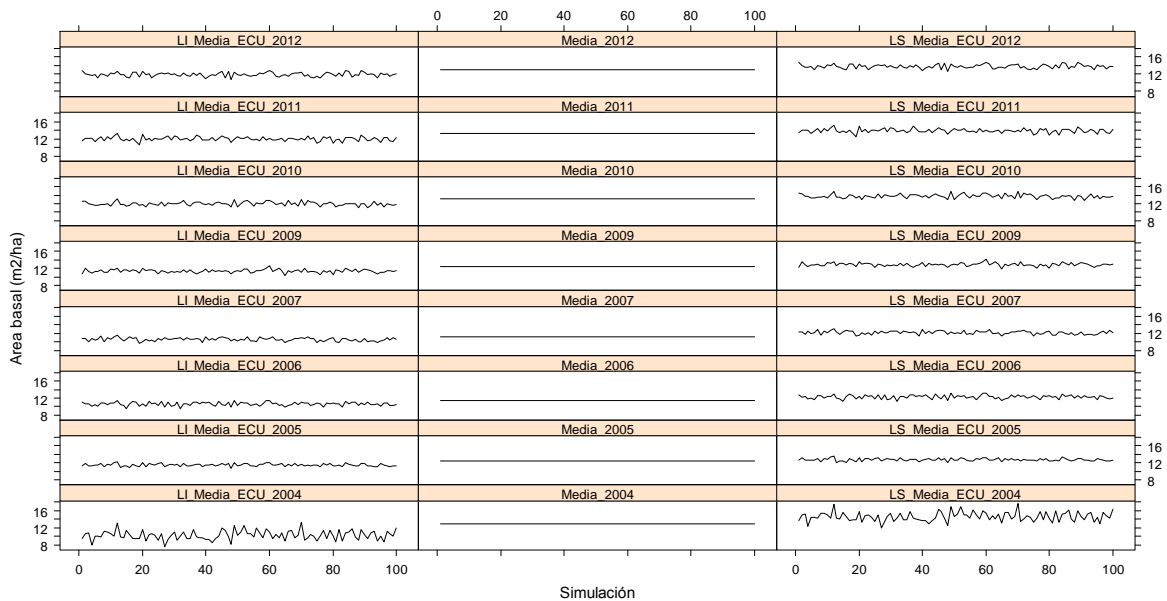


Figura 4.13. Intervalos de confianza al 95% para ABHA generados por ECU en la condición uno. El porcentaje de estos intervalos que atrapan el valor verdadero de ABHA fueron de 95, 83, 95, 93, 87,91,92 y 92% para 2004, 2005, 2006, 2007, 2009, 2010, 2011 y 2012, respectivamente. Los intervalos de confianza amplificados para cada ocasión en la condición se presentan el ANEXO 4.6.

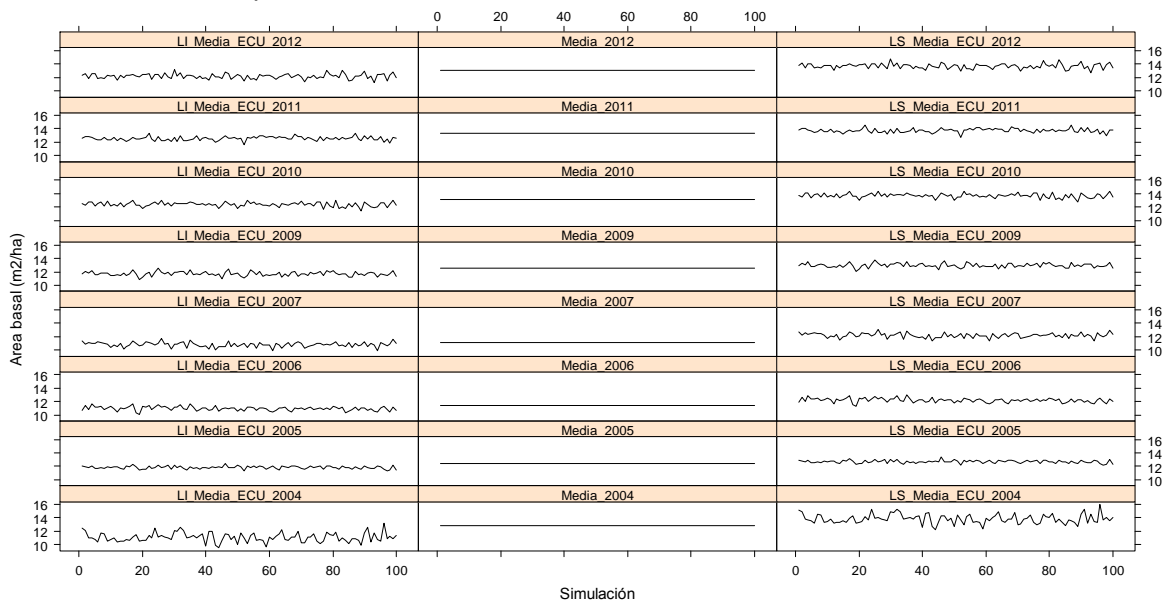


Figura 4.14. Intervalos de confianza al 95% para ABHA generados por ECU en la condición dos. El porcentaje de estos intervalos que atrapan el valor verdadero de ABHA fueron de 93, 83, 93, 86, 91, 93, 92 y 90 para 2004, 2005, 2006, 2007, 2009, 2010, 2011 y 2012, respectivamente. Los intervalos de confianza amplificados para cada ocasión en la condición dos se presentan el ANEXO 4.7.

4.7.2. Estimación de ABHA en la ocasión más reciente

En el Cuadro 4.4 se presentan los resultados de la estimación de ABHA correspondiente únicamente a 2012. Estos valores se obtuvieron por MCG, MRE y MM. En la condición uno el sesgo relativo mayor fue de -4.86%, mientras que en la condición dos fue de -4.24%, ambos correspondientes a MM-FIA0. En las dos condiciones MRE presentó el menor sesgo relativo en comparación con MCG y MM. El sesgo relativo de MCG fue intermedio entre MM-BIA3 y MM-BIA4 con valores de -2.85 y -2.12% en la condición uno y dos, respectivamente. Para los estimadores MM se nota una reducción gradual del sesgo relativo conforme se asigna mayor peso a la ocasión más reciente (2012), pasando de -4.86 a -1.95% en la ocasión uno y de -4.24 a -1.41% en la ocasión dos. Sin embargo, esta tendencia no se observa en el caso de la varianza, de hecho los valores son prácticamente iguales independientemente de los pesos asignados.

En todos los casos la varianza de los estimadores MM resultó menor que MCG; en promedio la varianza de MM fue tan sólo el 35% de la varianza de MCG. Aun cuando los estimadores MM-BIA1, MM-BIA2 y MM-BIA3 presentan menor varianza que MCG, tienen el inconveniente de presentar mayor sesgo que éste. Para los tres estimadores, los valores de la RaizCMEREL de la condición dos son menores que aquellos de la condición uno. En ambas condiciones la RaizCMEREL de MM disminuye a medida que los pesos se incrementan. Asimismo, la RaizCMEREL de MAFIA0 supera a la RaizCMEREL de MCG y de MRE. Resulta interesante observar que en ambas condiciones la RaizCMEREL de MM-BIA4 tuvo el menor valor en comparación con los otros estimadores. De hecho, la RaizCMEREL de MM-BIA4 fue un poco más del 60% respecto al MCG.

Por otra parte, los factores de ponderación de los estimadores MM-BIA fueron totalmente arbitrarios sobre la base de que las estimaciones de las ocasiones previas aportan mayor información a la estimación de la ocasión actual. El estimador MM-FIA0, también caracterizado por presentar varianza pequeña, resultó ser el más sesgado de los estimadores en ambas condiciones. MM-FIA0, al

considerar factores de ponderación iguales, conduce a ser un estimador sencillo e intuitivamente insesgado de la parte media del periodo de evaluación. Aunque sesgado del parámetro de la ocasión actual, la sencillez y reducción de la varianza de MA-FIA0, ha justificado que sea el estimador por *default* que usa el FIA de USA (Eskelson *et al.*, 2009). En estas condiciones, los estimadores MCG y MRE siguen siendo los preferidos, y su combinación mediante ECU trae consigo una estimación casi insesgada y de mínima varianza.

Los estimadores MM manifiestan su sesgo en una subestimación. Sus varianzas pequeñas se reflejan en que los valores se concentran alrededor de su valor esperado (Figura 4.15 y Figura 4.16).

Cuadro 4.4. Estimación de área basal por hectárea (ABHA) ($m^2 ha^{-1}$) en 2012 mediante MCG, MRE y MM bajo dos condiciones diferentes.

	MCG- 2012	MRE- 2012	MM- FIA0	MM- BIA1	MM- BIA2	MM- BIA3	MM- BIA4
Condición 1							
Estimación	12.70	13.07	12.44	12.54	12.58	12.65	12.82
Parámetro	13.08	13.08	13.08	13.08	13.08	13.08	13.08
Sesgo	-0.37	-0.01	-0.64	-0.53	-0.50	-0.43	-0.25
Sesgo relativo (%)	-2.85	-0.04	-4.86	-4.07	-3.82	-3.27	-1.95
Varianza	0.26	0.39	0.09	0.08	0.07	0.08	0.09
CME	0.40	0.39	0.49	0.36	0.32	0.26	0.15
RaizCME	0.63	0.63	0.70	0.60	0.56	0.51	0.39
RaizCMEREL (%)	4.85	4.80	5.35	4.61	4.30	3.89	2.99
Condición 2							
Estimación	12.80	13.08	12.52	12.63	12.66	12.72	12.89
Parámetro	13.08	13.08	13.08	13.08	13.08	13.08	13.08
Sesgo	-0.28	0.01	-0.55	-0.45	-0.42	-0.36	-0.18
Sesgo relativo (%)	-2.12	0.05	-4.24	-3.44	-3.22	-2.72	-1.41
Varianza	0.13	0.27	0.06	0.05	0.04	0.05	0.05
CME	0.21	0.27	0.36	0.25	0.22	0.17	0.09
RaizCME	0.46	0.52	0.60	0.50	0.47	0.42	0.30
RaizCMEREL (%)	3.51	3.96	4.62	3.86	3.57	3.19	2.26

* Condición 1 se refiere a tamaños de muestras panelizadas de 1000 conglomerados.

**Condición 2 se refiere a tamaños de muestras panelizadas de 1500 conglomerados.

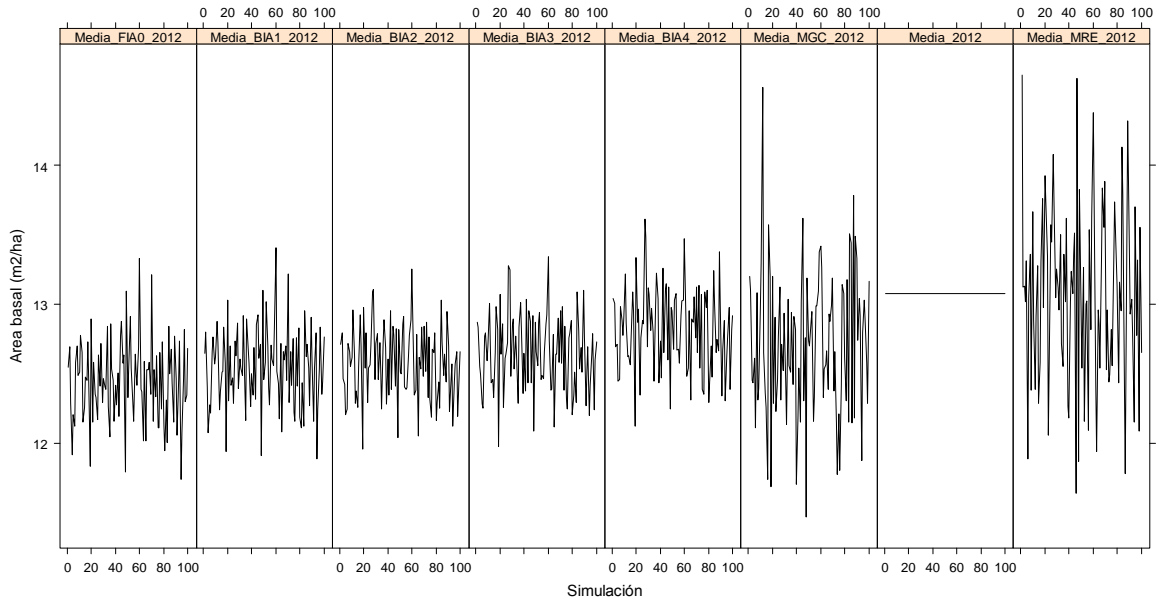


Figura 4.15. Estimaciones de ABHA obtenidas a través de MCG, MRE y MM en la condición uno. Los valores de MA-FIA0 presentan mayor lejanía respecto al valor verdadero.

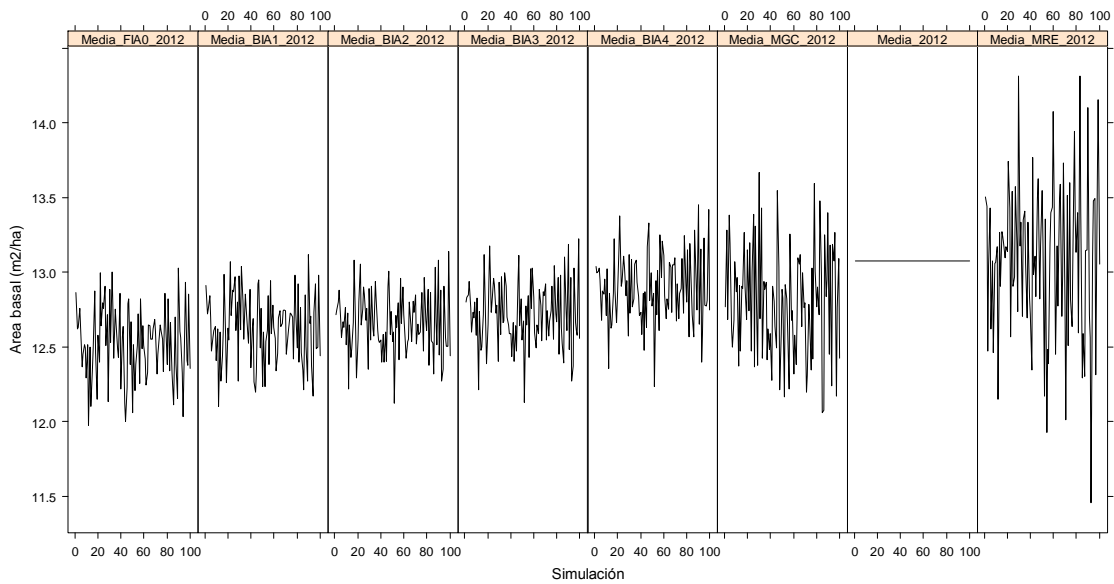


Figura 4.16. Estimaciones de ABHA obtenidas a través de MCG, MRE y MM en la condición dos. Los valores de MM-FIA0 presentan mayor lejanía respecto al valor verdadero.

4.7.3. Estimación de cambio de ABHA entre puntos medios de periodo

Las estimaciones del cambio de ABHA entre puntos medios de periodo para las condiciones uno y dos se reportan en el Cuadro 4.5. Las medias verdaderas fueron de 11.95 y 13 m² ha⁻¹, para el periodo uno y dos, respectivamente. El valor verdadero del cambio entre los puntos medios de los dos periodos fue de 1.05 m² ha⁻¹, equivalente a 0.21 m² ha⁻¹ año⁻¹. En las dos condiciones estudiadas el estimador MCG resultó medianamente sesgado; en la condición uno el sesgo relativo superó el 25% y en la condición dos el 20%. El estimador MRE es prácticamente insesgado; el mayor sesgo relativo se observó en la condición uno con 2.3%. El sesgo relativo de ECU es menor que MCG, pero significativamente mayor que MRE. Tanto en la condición uno como en la condición dos el sesgo de MCG y ECU se refleja en una subestimación, es decir, existe una tendencia hacia abajo del estimador, aunque ésta es de menor intensidad en ECU. Mientras que en MRE se observa una distribución equitativa de las estimaciones por arriba y por debajo del parámetro, en MCG se nota una marcada subestimación (Figura 4.17 y 4.18). Dado que en la condición dos los tres estimadores presentan menor sesgo que en la condición uno, puede decirse que estos estimadores son asintóticamente insesgados.

Cuadro 4.5. Estimaciones del cambio de ABHA entre periodos para las condiciones uno y dos.

	Condición 1			Condición 2		
	MCG	MRE	ECU	MCG	MRE	ECU
Estimación	0.78	1.02	0.87	0.83	1.04	0.90
Parámetro	1.05	1.05	1.05	1.05	1.05	1.05
Sesgo	-0.27	-0.02	-0.18	-0.22	-0.01	-0.15
Sesgo relativo (%)	-25.87	-2.30	-16.96	-21.23	-1.01	-14.18
Varianza	0.08	0.15	0.06	0.03	0.12	0.04
CME	0.15	0.15	0.09	0.08	0.12	0.06
RaizCME	0.39	0.39	0.31	0.29	0.34	0.25
RaizCMEREL (%)	36.77	37.49	29.22	27.70	32.50	23.66

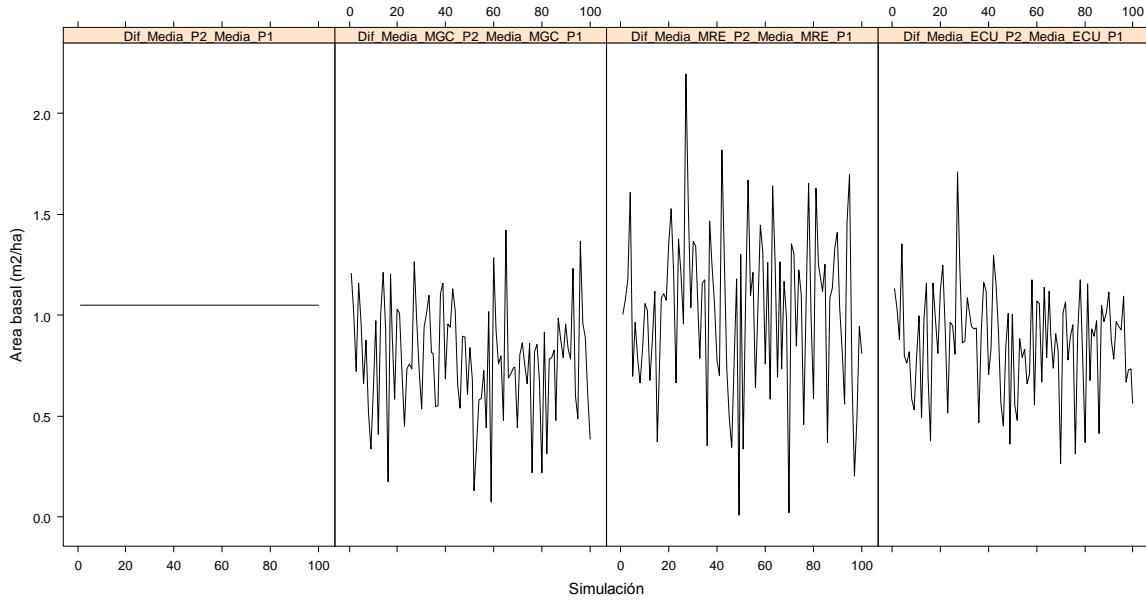


Figura 4.17. Estimación del cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición uno. El cambio se obtuvo mediante estimadores MCG, MRE y ECU.

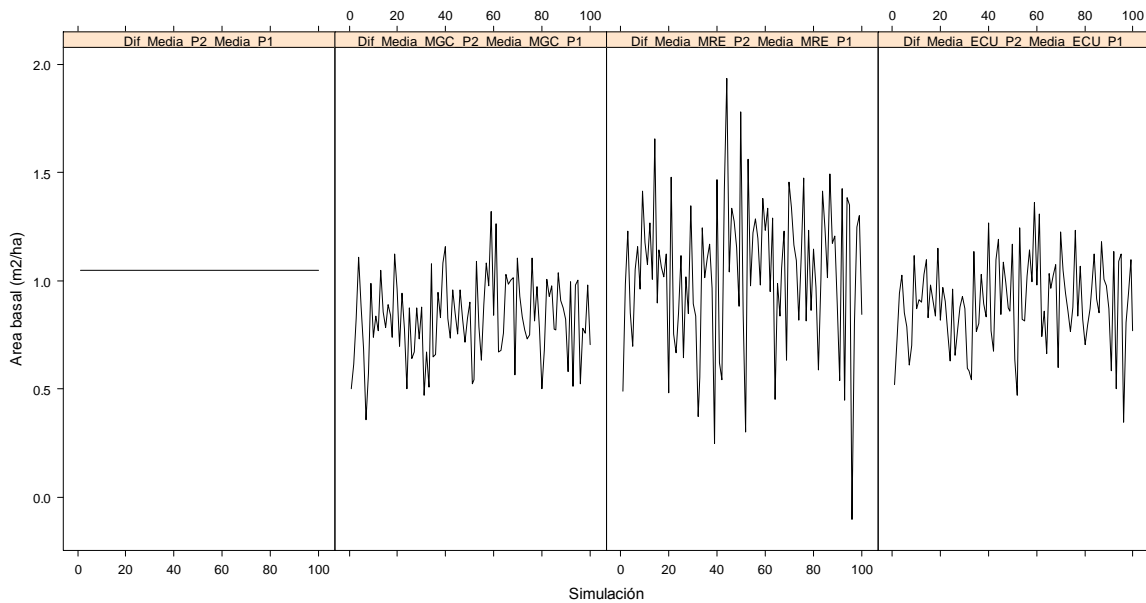


Figura 4.18. Estimación del cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición dos. El cambio se obtuvo mediante estimadores MCG, MRE y ECU.

Aun cuando el estimador MCG es medianamente sesgado tiene como ventaja presentar menor varianza que MRE; en la condición uno la varianza de MCG fue el 53% respecto a MRE, y en la condición dos del 25%. Por otra parte, en las dos condiciones la varianza de ECU es casi la misma que MCG; no obstante, la varianza de ECU respecto a MRE es en promedio del 36% al considerar ambas

condiciones. El ECU tiene como propiedad reducir el sesgo de MCG y reducir varianza de MRE. Entonces, ECU puede considerarse un estimador de varianza mínima y moderadamente sesgado, aunque se espera que este sesgo disminuya conforme el tamaño de muestras panelizadas se incrementa. De los tres estimadores, ECU presentó la menor RaízCMEREL, la cual fue arriba del 29% para la condición uno, y cerca del 24% para la condición dos.

En la Figura 4.19 se exponen las funciones de densidad normal modeladas para las estimaciones de cambio de ABHA correspondientes a la condición uno y en la Figura 4.20 a la condición dos. En ambas se observa que la media de ECU se localiza entre la media de MCG y MRE, además, aunque la campana de Gauss de ECU es tan angosta como la curva normal de MCG, también es notablemente más estrecha en comparación con la de MRE.

En base a estas consideraciones, parece ser que ECU es el mejor de los tres métodos para estimar el cambio de área basal por hectárea entre puntos medios de periodo. El porcentaje de los intervalos de confianza ECU, al 95%, que atrapan el valor verdadero de cambio de ABHA fue de 87% en la condición uno y de 88% en la condición dos (Figura 4.21 y 4.22).

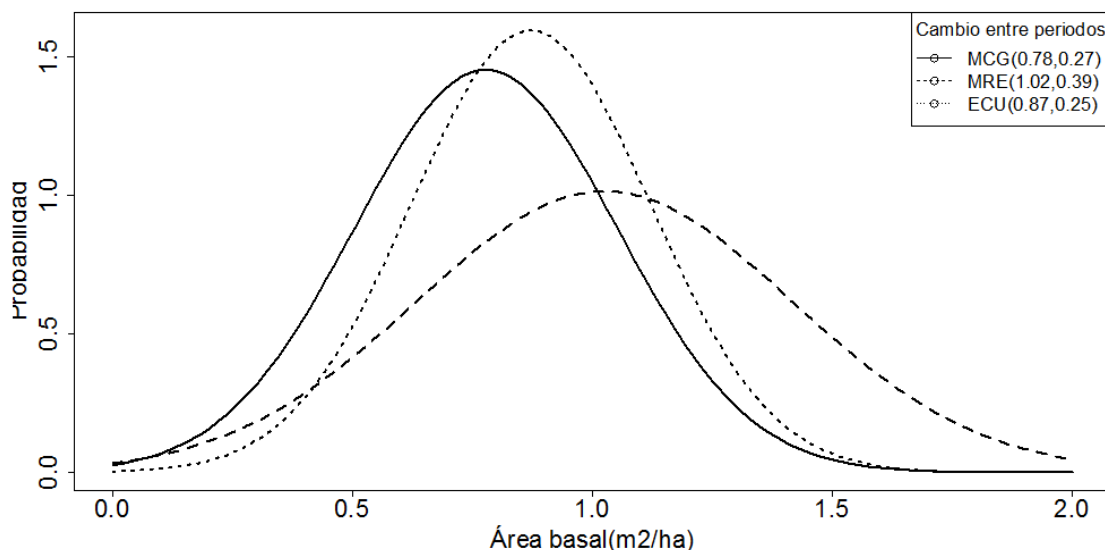


Figura 4.19. Funciones de densidad normal modeladas del cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición uno. El cambio se obtuvo mediante estimadores MCG, MRE y ECU.

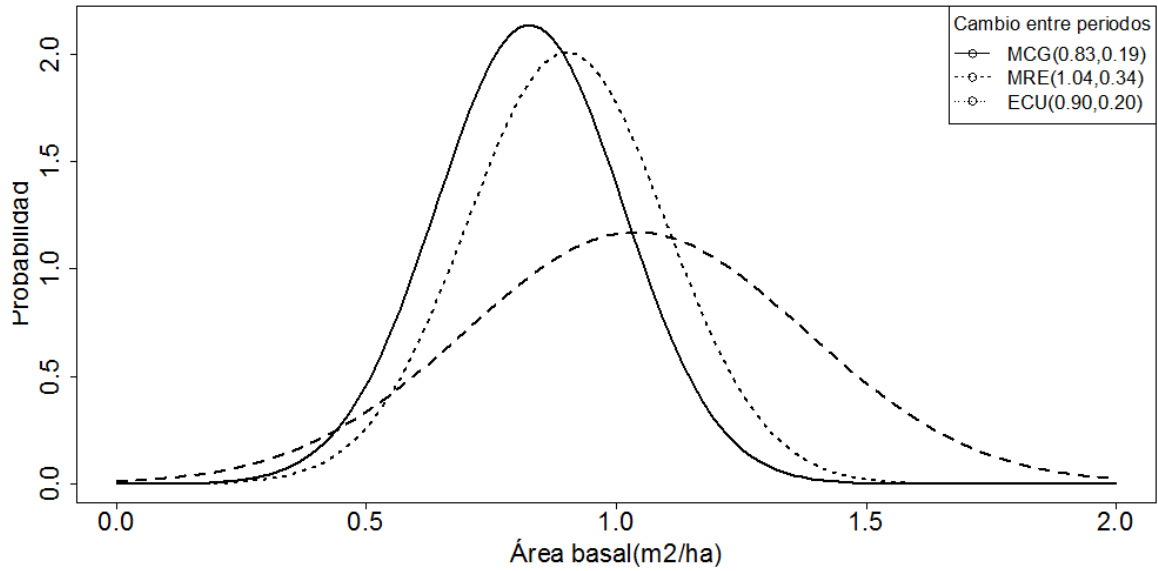


Figura 4.20. Funciones de densidad normal modeladas del cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición dos. El cambio se obtuvo mediante estimadores MCG, MRE y ECU.

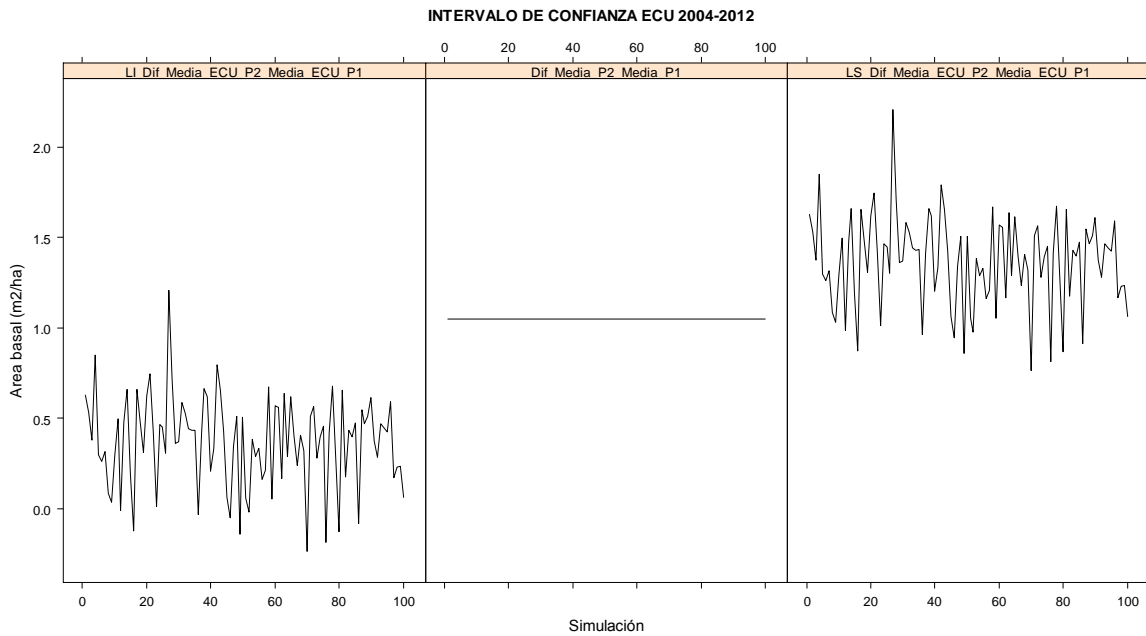


Figura 4.21. Intervalos de confianza al 95% para cambio de ABHA generados por ECU en la condición uno. El porcentaje de estos intervalos que atrapan el valor verdadero de ABHA fue de 87%.

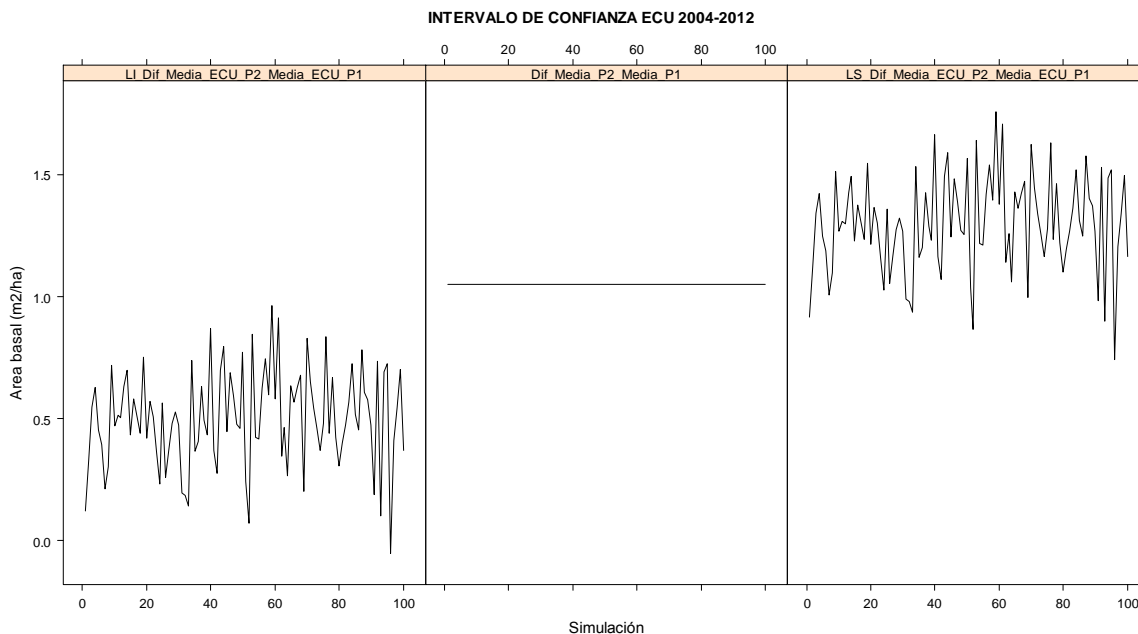


Figura 4.22. Intervalos de confianza al 95% para cambio de ABHA generados por ECU en la condición dos. El porcentaje de estos intervalos que atrapan el valor verdadero de ABHA fue de 88%.

4.7.4. Estimación de la tasa de cambio de ABHA entre puntos medios de periodo

En el Cuadro 4.6 se muestran los resultados correspondientes a las tasas de cambio entre puntos medios de periodo en las dos condiciones estudiadas. La tasa de cambio se refiere a la razón del promedio de ABHA del punto medio del periodo dos respecto al promedio de ABHA del punto medio del periodo uno. Las medias verdaderas fueron de 11.95 y 13 m² ha⁻¹, para el periodo uno y dos, respectivamente. La tasa de cambio verdadera fue de 1.09 e indica un cambio de ABHA entre puntos medios de periodo de aproximadamente 1.07 m² ha⁻¹.

Resulta interesante observar que en las dos condiciones estudiadas MCG y ECU conducen a resultados similares en cuanto a sesgo relativo y varianza. Al considerar los tres estimadores, el sesgo relativo en ningún caso superó el 2%; asimismo, la RaizCMEREL en ningún caso fue mayor al 3.5%. No obstante, ECU presenta la RaizCMEREL ligeramente menor que MCG y MRE. Por lo que ECU puede considerarse un estimador apropiado cuando el interés es la estimación de la tasa de cambio entre puntos medios de periodo.

Las RaizCMEREL para estimación del cambio (Cuadro 4.5.) superan en gran magnitud los correspondientes para la estimación de la tasa de cambio. En aquel caso se alcanzaron valores hasta del 37%, mientras que en la estimación de la tasa de cambio el máximo valor encontrado fue de 3.21%. Cifras que por cierto pertenecen al estimador MRE en la condición uno.

Cuadro 4.6. Estimaciones de la tasa de cambio de ABHA entre puntos medios de periodo para las condiciones uno y dos.

	Condición 1			Condición 2		
	MCG	MRE	ECU	MCG	MRE	ECU
Estimación	1.0667	1.0868	1.0740	1.0695	1.0872	1.0756
Parámetro	1.0877	1.0877	1.0877	1.0877	1.0877	1.0877
Sesgo	-0.0210	-0.0009	-0.0137	-0.0182	-0.0005	-0.0122
Sesgo relativo (%)	-1.9316	-0.0827	-1.2631	-1.6722	-0.0476	-1.1192
Varianza	0.0006	0.0012	0.0005	0.0003	0.0009	0.0003
CME	0.0011	0.0012	0.0007	0.0006	0.0009	0.0005
RaizCME	0.0325	0.0350	0.0264	0.0246	0.0303	0.0214
RaizCMEREL (%)	2.9899	3.2134	2.4295	2.2615	2.7887	1.9689

Las funciones de densidad normal de la tasa de cambio modeladas para los tres estimadores se presentan en la Figura 4.23 para la condición uno y en la Figura 4.24 para la condición dos. En las dos condiciones se aprecia gran similitud, en cuanto a su amplitud, de las curvas normales entre MCG y ECU. La curva normal MRE refleja mayor variabilidad con respecto a MCG y ECU.

Finalmente, el porcentaje de intervalos de confianza al 95% que atrapan a la tasa de cambio verdadera de ABHA fue de 89% tanto en la condición uno como en la condición dos (Figura 4.25 y 4.26). Como efecto de la varianza pequeña, estos intervalos de confianza son más estrechos comparativamente con aquéllos obtenidos en la estimación del cambio.

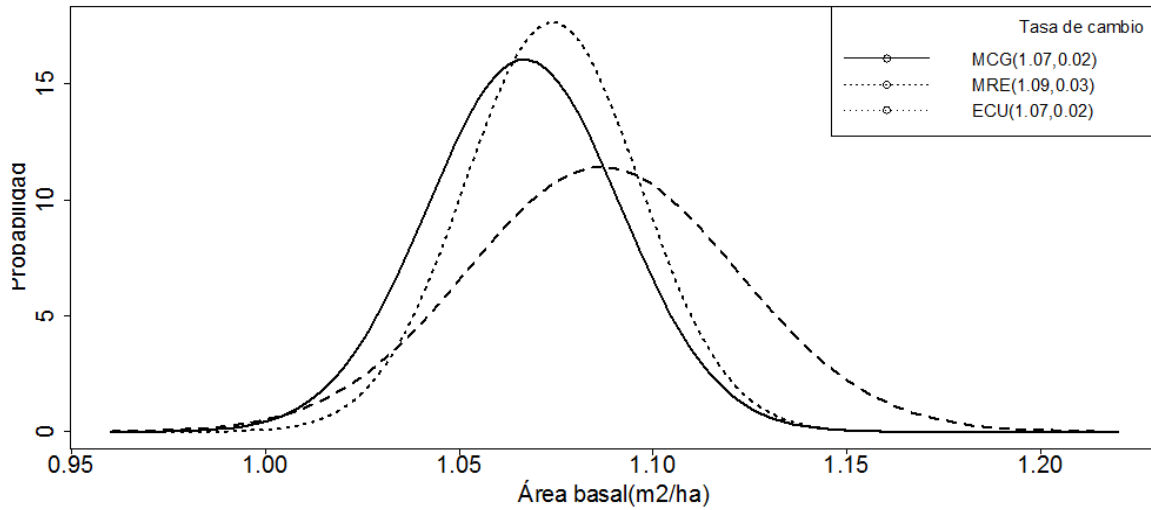


Figura 4.23. Funciones de densidad normal modeladas de la tasa cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición uno. El cambio se obtuvo mediante estimadores MCG, MRE y ECU.

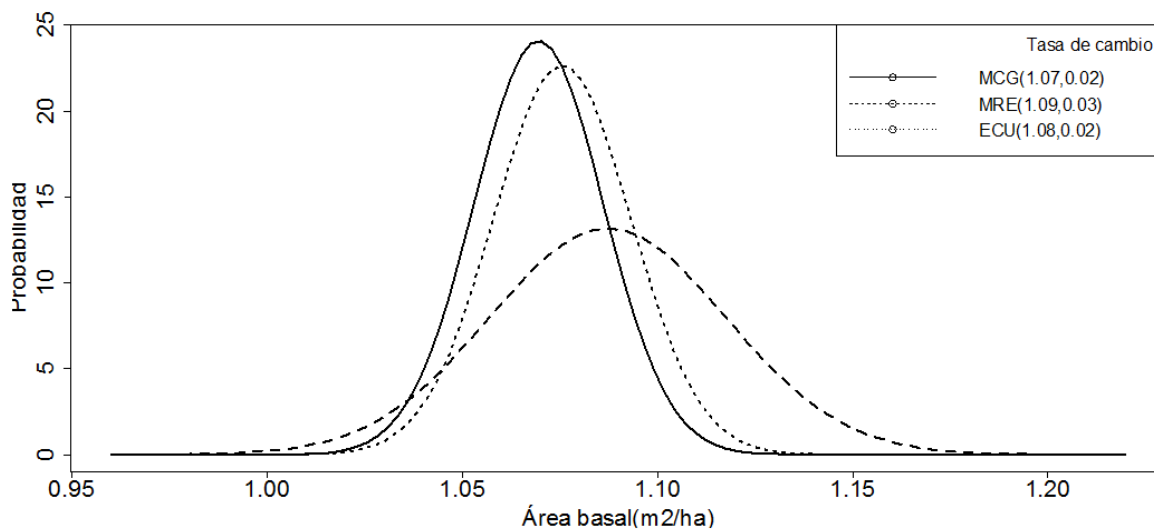


Figura 4.24. Funciones de densidad normal modeladas de la tasa cambio de ABHA entre periodos (2004-2007 y 2009-2012) de la condición dos. El cambio se obtuvo mediante estimadores MCG, MRE y ECU.

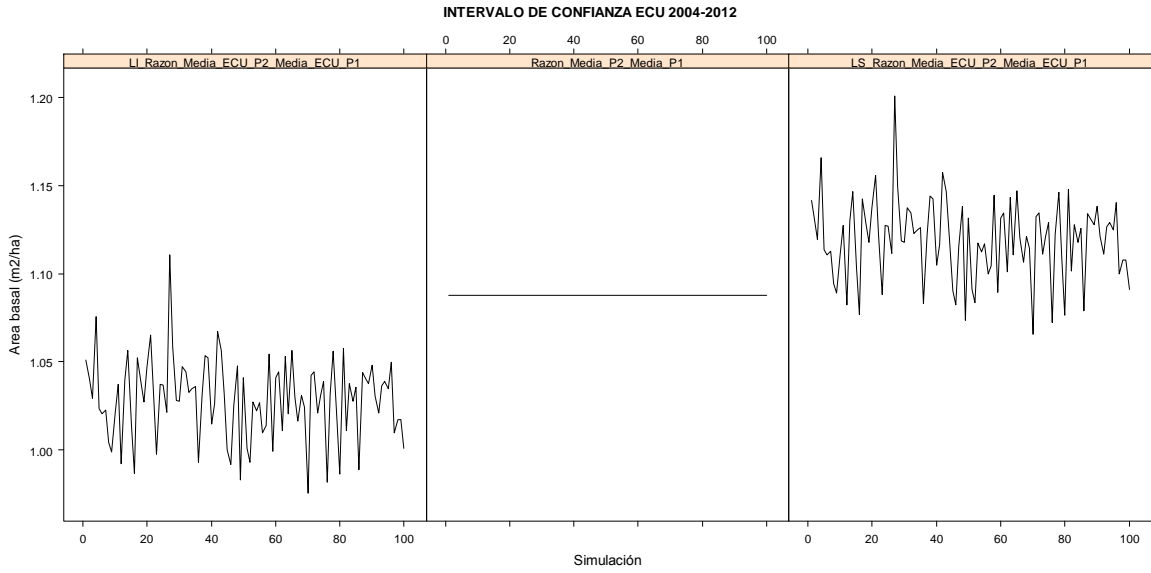


Figura 4.25. Intervalos de confianza al 95% para la tasa de cambio de ABHA generados por ECU en la condición uno. El porcentaje de estos intervalos que atrapan el valor verdadero de la tasa de cambio de ABHA fue de 89.

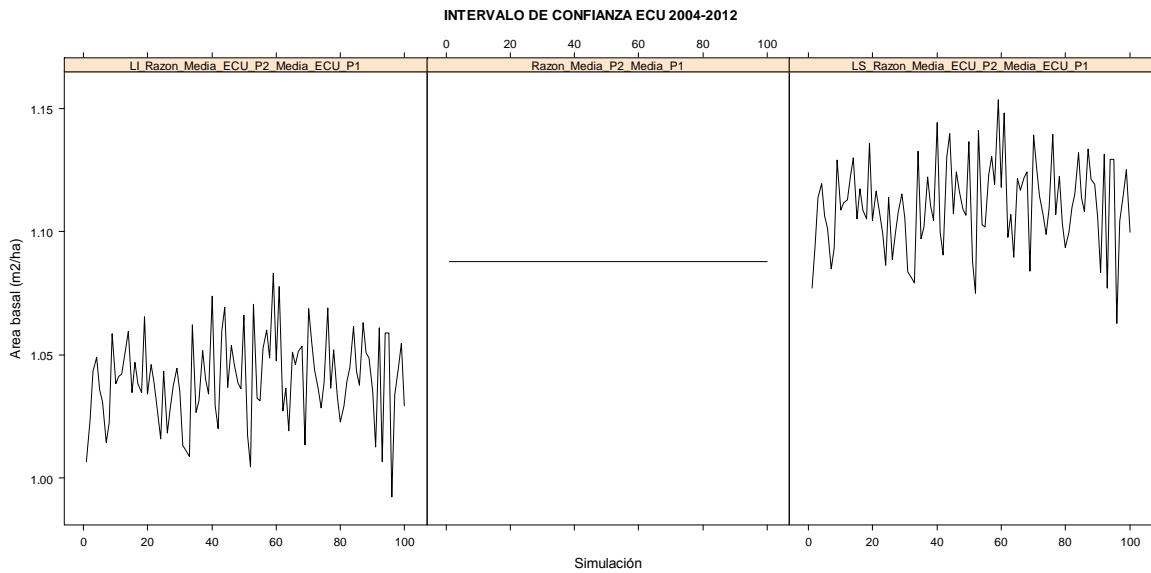


Figura 4.26. Intervalos de confianza al 95% para la tasa de cambio de ABHA generados por ECU en la condición dos. El porcentaje de estos intervalos que atrapan el valor verdadero de la tasa de cambio de ABHA fue de 89.

4.8. CONCLUSIONES

En las dos condiciones de estudio y para cada una de las ocho ocasiones de medición se observó que la estimación de ABHA correspondiente a MCG fue ligeramente sesgada en comparación con el ABHA de MRE. Dado que el sesgo relativo de MCG fue menor al 5.5% se consideró insignificante. Aunque MCG subestimó ligeramente el parámetro, en general presentó una varianza menor respecto a MRE.

El sesgo relativo de MRE en ningún caso fue superior al 1%. Así, las estimaciones de ABHA obtenidas mediante MRE se consideraron insesgadas. En promedio, la varianza de MRE fue casi el doble de la varianza de MCG.

El ECU generó en cada una de las ocasiones de medición un sesgo relativo intermedio entre MCG y MRE. El ECU además de reducir el sesgo respecto al MCG también presentó menor varianza comparado con MRE. De esta manera, ECU se considera un estimador insesgado (o cuasinsesgado) de mínima varianza.

El estimador MM-FIA0, caracterizado por presentar varianza pequeña, resultó ser el estimador más sesgado —en ambas condiciones— en la ocasión más reciente. MM-FIA0, al considerar factores de ponderación iguales, conduce a ser un estimador sencillo e intuitivamente insesgado de la parte media del periodo de evaluación.

En la estimación de la tasa de cambio entre puntos medios de periodo, en las dos condiciones estudiadas, MCG y ECU conducen a resultados similares en cuanto a sesgo relativo y varianza. No obstante, ECU presenta la RaizCMEREL ligeramente menor que MCG y MRE. Entonces, ECU puede considerarse un estimador apropiado cuando el interés es la estimación de la tasa de cambio.

Para la implementación de ECU en el marco del INFyS, además de las varianzas de MCG y MRE, se requiere de la covarianza entre ellos (ecuación 4.118). No obstante, en la práctica únicamente se tiene una muestra, por lo que

una forma eficiente de estimar estas varianzas y covarianzas es mediante el método de Bootstrap, es decir, replicar un número grande de veces la muestra original panelizada. Esta forma de proceder trae consigo muchas estimaciones de MCG, MRE, y en consecuencia de ECU, en cada ocasión de medición.

Dado que además del área basal puede existir interés en estudiar otros parámetros forestales como volumen, carbono, biomasa o densidad por unidad de superficie, el Bootstrap debe implementarse desde la perspectiva multitemporal y multivariada, de manera tal que se conserve la correlación temporal y contemporánea (entre variables del rodal). Con el fin de asegurar que la matriz de varianzas y covarianzas de MCG pueda ser invertida, se sugiere que la muestra original incluya ocasiones de medición que contengan por lo menos 10 conglomerados.

Los métodos estadísticos aquí esbozados permitirán estimar de forma eficiente los parámetros forestales agregados en cada una de las diferentes ocasiones de medición conforme los datos de campo se vayan colectando, es decir, no es necesario esperar hasta el fin del periodo. De esta manera, se podrán construir series temporales promedio y evaluar la tendencia de los recursos forestales en el corto y mediano plazo.

El estimador MCG se hará potente conforme la correlación entre las parcelas remedidas sea alta, de aquí la importancia de que la Comisión Nacional Forestal siga un protocolo estricto de control de calidad con respecto a la visita de las parcelas después de cinco años.

4.9. LITERATURA CITADA

- Arner S.L., J. A. Westfall and C.T. Scott. 2004. Comparison of annual inventory designs using forest inventory and analysis data. *Forest Science*. 50(2):188-203.
- Binder D. A. and M. A. Hidiroglou. Sampling time. 1988. In: *Handbook of statistics, Volumen 6 Sampling*. Edited by P.K. Krishnaiah and C.R. Rao. Elsevier Science Publishers B.V. Netherlands. pp:187-211.
- Bromiley, P.A. 2004. *Products and Convolutions of Gaussian Probability Density Functions*. University of Manchester Manchester, UK. 13 p.
- Cochran, W. G. 1993. *Técnicas de Muestreo*. Compañía Editorial Continental. México. 513 p.
- Czaplewski, R. L. 2010. Complex sample survey estimation in static state-space. Gen. Tech. Rep. RMRS-GTR-239. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 124 p.
- Czaplewski, R. L. and Thompson, M. T. 2009. Opportunities to improve monitoring of temporal trends with FIA panel data. In: McWilliams, Will; Moisen, Gretchen; Czaplewski, Ray, comps. *Forest Inventory and Analysis (FIA) Symposium 2008; October 21-23, 2008; Park City, UT*. Proc. RMRS-P-56CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 55 p.
- Dixon B. L. and R. E. Howitt. 1979. Continuous forest inventory using a linear filter. *Forest Science*. 25(4):675-689.
- Eskelson B.N.I., H. Temesgen and T.M. Barret. 2009. Estimating current forest attributes from paneled inventory data using plot-level imputation: A study from the Pacific Northwest. *Forest Science*. 55(1):64-71.
- Esteban G., M.V. y M. Regúlez C. 2010. *Análisis de Datos: un Enfoque Econométrico*. Universidad de País Vasco. España. 272 p.
- Fuller, W.A. 2009. *Sampling Statistics*. John Wiley & Sons, Inc. New Jersey. USA. 454 p.
- Greene W.H. 2012. *Econometric Analysis*. Seventh Edition. Prentice Hall. USA. 1189 p.
- Gregoire, T.G. and H.T. Valentine. 2008. *Sampling Strategies for Natural Resources and the Environment*. Chapman & Hall/CRC. New York, USA. 474 p.

- Hoff, P.D. 2009. A First Course in Bayesian Statistical Methods. Springer Science + Business Media. New York. USA. 267.
- Johnson D. S., M. S. Williams y R. L. Czaplewski. 2003. Comparison of estimators for rolling samples using Forest Inventory and Analysis data. *Forest Science*. 49(1):50-63.
- Köhl, M., A. Lister, C.T. Scott, T. Baldauf and D. Plugge. 2011. Implications of sampling design and sample size for national carbon accounting systems. *Carbon Balance and Management*. 6(10):1-20.
- Maybeck, P.S. 1979. Stochastic Models, Estimation and Control. Academic Press. New York. USA. 423 p.
- McRoberts, R.E., W.A. Bechtold, P.L. Patterson, C.T. Scott and G.A. Reams. 2005. The enhanced Forest Inventory and Analysis program of the USDA Forest Service: Historical perspective and announcement of statistical documentation. *Journal of Forestry* 103(6): 304-308.
- Patterson and Reams. 2005. Combining panels for forest inventory and analysis estimation. In: The enhanced Forest Inventory and Analysis program-national sampling design and estimation procedures. W. A. Bechtold and P. L. Patterson (Eds.). General Technical Report SRS-80. Forest Service. United States Department of Agriculture. Southern Research Station, Asheville, North Carolina. USA. pp: 69-74.
- Rencher, A.C. and G. Bruce S. 2008. Linear Models in Statistics. Second Edition. Wiley-Interscience. USA. 672 p.
- Ross, S. (2006). A First Course in Probability. Seventh Edition. Pearson Prentice Hall. New Jersey. USA. 565 p.
- SAS Institute Inc. 2009. SAS/STAT 9.2 User's Guide. Second Edition. Cary, NC: SAS Institute Inc. North Carolina. USA. 7869 p.
- SAS Institute Inc. 2010. SAS/IML 9.22 User's Guide. Cary, NC: SAS Institute Inc. North Carolina. USA. 1099 p.
- Schreuder, H.T., T. G. Gregoire and G. B. Wood. 1993. Sampling Methods for Multiresource Forest Inventory. John Wiley & Sons, Inc. New York, USA. 446 p.
- Schreuder, H. T., R. Ernst and H. Ramírez M. 2004. Statistical Techniques for Sampling and Monitoring Natural Resources. Forest Service-USDA. General Technical Report RMRS-GTR-126. USA. 111 p.

- Scott, C. 1984. A new look at sampling with partial replacement. *Forest Science*. 30(1):157-166.
- Scott, C. T. and M. Köhl. 1994. Sampling with partial replacement and stratification. *Forest Science*. 40(1):30-46.
- Scott, C.T., M. Köhl and H.J. Schnellbacher. 1999. A comparison of periodic and annual forest surveys. *Forest Science*. 45(3):433-451.
- Theil, H. 1971. *Principles of Econometrics*. John Wiley & Sons, Inc.USA. 736 p.
- Van Deusen, P.C. 1989. Multiple-occasion partial replacement sampling for growth components. *Forest Science*. 35(2):388-400.

ANEXO 4.1. DETERMINACIÓN DE a Y d (EXPRESION (4.15)). MUESTREO CON REEMPLAZO PARCIAL EN DOS OCASIONES

Recuerde que (4.16) se expresa como

$$\begin{aligned} V(\hat{T}_{y,MRP}(t_2 : t_1)) &= a^2 (V(\hat{T}_{y\pi,rep}(t_1, m_{1.})) + V(\hat{T}_{y\pi,rep}(t_1, m_{12}))) \\ &\quad + (1-d)^2 V(\hat{T}_{y\pi,rep}(t_2, m_{2.})) + d^2 V(\hat{T}_{y\pi,rep}(t_2, m_{12})) \\ &\quad - 2adCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) \end{aligned} \quad (4.16)$$

Para encontrar los valores de a y d en (4.16) considere lo siguiente:

$$\begin{aligned} \frac{\partial V(\hat{T}_{y,MRP}(t_2 : t_1))}{\partial a} &= 2a(V(\hat{T}_{y\pi,rep}(t_1, m_{1.})) + V(\hat{T}_{y\pi,rep}(t_1, m_{12}))) \\ &\quad - 2dCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) \\ \frac{\partial V(\hat{T}_{y,MRP}(t_2 : t_1))}{\partial d} &= -2(1-d)V(\hat{T}_{y\pi,rep}(t_2, m_{2.})) + 2dV(\hat{T}_{y\pi,rep}(t_2, m_{12})) \\ &\quad - 2aCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) \end{aligned}$$

Igualando estas ecuaciones a cero:

$$\begin{aligned} 2a(V(\hat{T}_{y\pi,rep}(t_1, m_{1.})) + V(\hat{T}_{y\pi,rep}(t_1, m_{12}))) - 2dCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) &= 0 \\ -2(1-d)V(\hat{T}_{y\pi,rep}(t_2, m_{2.})) + 2dV(\hat{T}_{y\pi,rep}(t_2, m_{12})) - 2aCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) &= 0 \end{aligned}$$

Eliminando el 2 en todo el sistema

⇒

$$\begin{aligned} a(V(\hat{T}_{y\pi,rep}(t_1, m_{1.})) + V(\hat{T}_{y\pi,rep}(t_1, m_{12}))) - dCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) &= 0 \\ -(1-d)V(\hat{T}_{y\pi,rep}(t_2, m_{2.})) + dV(\hat{T}_{y\pi,rep}(t_2, m_{12})) - aCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) &= 0 \end{aligned}$$

Despejando a de la de la primera ecuación y acomodando términos en la segunda

⇒

$$a = \frac{dCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12}))}{V(\hat{T}_{y\pi,rep}(t_1, m_{12})) + V(\hat{T}_{y\pi,rep}(t_2, m_{12}))} - V(\hat{T}_{y\pi,rep}(t_2, m_{22})) + dV(\hat{T}_{y\pi,rep}(t_2, m_{22})) + dV(\hat{T}_{y\pi,rep}(t_2, m_{12})) - aCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) = 0$$

Sustituyendo a en la segunda ecuación

⇒

$$-V(\hat{T}_{y\pi,rep}(t_2, m_{22})) + dV(\hat{T}_{y\pi,rep}(t_2, m_{22})) + dV(\hat{T}_{y\pi,rep}(t_2, m_{12})) - \left(\frac{dCov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12}))}{V(\hat{T}_{y\pi,rep}(t_1, m_{12})) + V(\hat{T}_{y\pi,rep}(t_2, m_{12}))} \right) Cov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) = 0$$

Multiplicando las covarianzas en el cuarto término

⇒

$$-V(\hat{T}_{y\pi,rep}(t_2, m_{22})) + dV(\hat{T}_{y\pi,rep}(t_2, m_{22})) + dV(\hat{T}_{y\pi,rep}(t_2, m_{12})) - d \frac{\left(Cov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) \right)^2}{V(\hat{T}_{y\pi,rep}(t_1, m_{12})) + V(\hat{T}_{y\pi,rep}(t_2, m_{12}))} = 0$$

Factorizando d

⇒

$$-V(\hat{T}_{y\pi,rep}(t_2, m_{22})) + d \left(V(\hat{T}_{y\pi,rep}(t_2, m_{22})) + V(\hat{T}_{y\pi,rep}(t_2, m_{12})) - \frac{\left(Cov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) \right)^2}{V(\hat{T}_{y\pi,rep}(t_1, m_{12})) + V(\hat{T}_{y\pi,rep}(t_2, m_{12}))} \right) = 0$$

Despejando d

⇒

$$d = \frac{V(\hat{T}_{y\pi,rep}(t_2, m_{22}))}{V(\hat{T}_{y\pi,rep}(t_2, m_{22})) + V(\hat{T}_{y\pi,rep}(t_2, m_{12})) - \frac{\left(Cov(\hat{T}_{y\pi,rep}(t_1, m_{12}), \hat{T}_{y\pi,rep}(t_2, m_{12})) \right)^2}{V(\hat{T}_{y\pi,rep}(t_1, m_{12})) + V(\hat{T}_{y\pi,rep}(t_2, m_{12}))}}$$

O bien

$$d = \frac{V(\hat{T}_{y\pi,rep}(t_2, m_{22}))}{V(\hat{T}_{y\pi,rep}(t_2, m_{22})) + V(\hat{T}_{y\pi,rep}(t_2, m_{12})) - \gamma_{12}}$$

$$\text{Donde } \gamma_{12} = \frac{\left(\text{Cov}(\hat{T}_{y\pi,rep}(t_1, m_{1_2}), \hat{T}_{y\pi,rep}(t_2, m_{1_2})) \right)^2}{V(\hat{T}_{y\pi,rep}(t_1, m_{1_2})) + V(\hat{T}_{y\pi,rep}(t_2, m_{1_2}))}$$

Recordando que bajo muestreo replicado con m observaciones evaluadas en una sola ocasión se tiene que (ecuación (2.10), Capítulo II):

$$V(\hat{T}_{y\pi,rep}) = \frac{1}{m} V(\hat{T}_{y\pi}) = \frac{\phi}{m}$$

donde

$$V(\hat{T}_{y\pi}) = \phi = \sum_{k=1}^N y_k^2 \left(\frac{1 - \pi_k}{\pi_k} \right) + 2 \sum_{k=1}^N \sum_{k' > k}^N y_k y_{k'} \left(\frac{\pi_{kk'} - \pi_k \pi_{k'}}{\pi_k \pi_{k'}} \right)$$

Entonces, por analogía:

$$V(\hat{T}_{y\pi,rep}(t_2, m_{2_2})) = \frac{\phi_2}{m_2}; V(\hat{T}_{y\pi,rep}(t_2, m_{1_2})) = \frac{\phi_2}{m_{1_2}}; V(\hat{T}_{y\pi,rep}(t_1, m_{1_1})) = \frac{\phi_1}{m_{1_1}};$$

$$V(\hat{T}_{y\pi,rep}(t_1, m_{1_2})) = \frac{\phi_1}{m_{1_2}}$$

Sustituyendo estos valores en d y expresando la covarianza en términos del coeficiente de correlación y las varianzas se tiene

$$d = \frac{V(\hat{T}_{y\pi,rep}(t_2, m_{2_2}))}{V(\hat{T}_{y\pi,rep}(t_2, m_{2_2})) + V(\hat{T}_{y\pi,rep}(t_2, m_{1_2})) - \frac{\left(\text{Cov}(\hat{T}_{y\pi,rep}(t_1, m_{1_2}), \hat{T}_{y\pi,rep}(t_2, m_{1_2})) \right)^2}{V(\hat{T}_{y\pi,rep}(t_1, m_{1_1})) + V(\hat{T}_{y\pi,rep}(t_1, m_{1_2}))}}$$

$$d = \frac{\frac{\phi_2}{m_2}}{\frac{\phi_2}{m_2} + \frac{\phi_2}{m_{1_2}} - \frac{\rho_{12}^2 \phi_1 \phi_2}{m_1 m_{1_2}}} = \frac{\frac{\phi_2}{m_2}}{\frac{\phi_2}{m_2} + \frac{\phi_2}{m_{1_2}} - \frac{\rho_{12}^2 \phi_1 \phi_2}{m_1 m_{1_2}}} = \frac{\frac{\phi_2}{m_2}}{\frac{\phi_2}{m_2} + \frac{\phi_2}{m_{1_2}} - \frac{\rho_{12}^2 \phi_1 \phi_2}{m_1 m_{1_2}}} = \frac{\frac{\phi_2}{m_2}}{\frac{\phi_2}{m_2} + \frac{\phi_2}{m_{1_2}} - \frac{\rho_{12}^2 \phi_1 \phi_2}{m_1 m_{1_2}}}$$

$$\begin{aligned}
&= \frac{\frac{\phi_2}{m_2}}{\frac{\phi_2}{m_2} + \frac{\phi_2}{m_{12}} - \frac{\rho_{12}^2 \phi_1 \phi_2 m_1 m_{12}}{m_{12}^2 \phi_1 m_1}} = \frac{\frac{1}{m_2}}{\frac{1}{m_2} + \frac{1}{m_{12}} - \frac{\rho_{12}^2 m_1 m_{12}}{m_{12}^2 m_1}} = \frac{\frac{1}{m_2}}{\frac{m_{12} + m_2}{m_2 m_{12}} - \frac{\rho_{12}^2 m_1 m_{12}}{m_{12}^2 m_1}} = \frac{\frac{1}{m_2}}{\frac{m_2}{m_2 m_{12}} - \frac{\rho_{12}^2 u_1}{m_{12}}} \\
&= \frac{\frac{1}{m_2}}{\frac{m_2}{m_2 m_{12}} - \frac{m_2 \rho_{12}^2 u_1}{m_2 m_{12}}} = \frac{1}{\frac{m_2}{m_{12}} - \frac{m_2 \rho_{12}^2 u_1}{m_{12}}} = \frac{1}{\frac{m_2 - m_2 \rho_{12}^2 u_1}{m_{12}}} = \frac{m_{12}}{m_2 - u_1 m_2 \rho_{12}^2}
\end{aligned}$$

Entonces

$$d = \frac{m_{12}}{m_2 - u_1 m_2 \rho_{12}^2}$$

En esta derivación se usaron las siguientes expresiones:

$$m_1 = m_1 + m_{12}$$

$$m_2 = m_2 + m_{12}$$

$$u_1 = \frac{m_1}{m_1}$$

Operando de manera similar en a y sustituyendo el valor de d

$$\begin{aligned}
a &= \frac{d \text{Cov}(\hat{T}_{y\pi, \text{rep}}(t_1, m_{12}), \hat{T}_{y\pi, \text{rep}}(t_2, m_{12}))}{V(\hat{T}_{y\pi, \text{rep}}(t_1, m_1)) + V(\hat{T}_{y\pi, \text{rep}}(t_1, m_{12}))} \\
&= \left(\frac{m_{12}}{m_2 - u_1 m_2 \rho_{12}^2} \right) \frac{\rho_{12} \sqrt{\frac{\phi_1}{m_{12}}} \sqrt{\frac{\phi_2}{m_{12}}}}{\frac{\phi_1}{m_1} + \frac{\phi_1}{m_{12}}} = \left(\frac{m_{12}}{m_2 - u_1 m_2 \rho_{12}^2} \right) \frac{\rho_{12} \sqrt{\frac{\phi_1}{m_{12}}} \sqrt{\frac{\phi_2}{m_{12}}}}{\frac{\phi_1 m_{12} + \phi_1 m_1}{m_1 m_{12}}} \\
&= \left(\frac{m_{12}}{m_2 - u_1 m_2 \rho_{12}^2} \right) \frac{\rho_{12} \frac{\sqrt{\phi_1} \sqrt{\phi_2}}{m_{12}}}{\frac{\phi_1 m_1}{m_1 m_{12}}} = \left(\frac{m_{12}}{m_2 - u_1 m_2 \rho_{12}^2} \right) \frac{\rho_{12} \sqrt{\phi_1} \sqrt{\phi_2}}{\frac{\phi_1 m_1}{m_1}} = \left(\frac{m_{12}}{m_2 - u_1 m_2 \rho_{12}^2} \right) \frac{m_1 \rho_{12} \sqrt{\phi_1} \sqrt{\phi_2}}{\phi_1 m_1} \\
&= \left(\frac{m_{12}}{m_2 - u_1 m_2 \rho_{12}^2} \right) u_1 \rho_{12} \sqrt{\phi_1} \sqrt{\phi_2} \phi_1^{-1} = \left(\frac{m_{12}}{m_2 - u_1 m_2 \rho_{12}^2} \right) u_1 \rho_{12} \sqrt{\phi_2} \phi_1^{-\frac{1}{2}} = \left(\frac{m_{12}}{m_2 - u_1 m_2 \rho_{12}^2} \right) \frac{u_1 \rho_{12} \sqrt{\phi_2}}{\sqrt{\phi_1}}
\end{aligned}$$

$$= \left(\frac{m_{12}u_1}{m_2 - u_1 m_{22} \rho_{12}^2} \right) \rho_{12} \sqrt{\frac{\phi_2}{\phi_1}}$$

Entonces

$$a = \left(\frac{m_{12}u_1}{m_2 - u_1 m_{22} \rho_{12}^2} \right) \rho_{12} \sqrt{\frac{\phi_2}{\phi_1}} = \left(\frac{m_{12}}{m_2 - u_1 m_{22} \rho_{12}^2} \right) u_1 \rho_{12} \sqrt{\frac{\phi_2}{\phi_1}} = du_1 B_{21}$$

Donde

$$B_{21} = \rho_{12} \sqrt{\frac{\phi_2}{\phi_1}}$$

Permita que $\hat{T}_{y\pi,rep}(t_1)$ denote el estimador de $T_y(t_1)$, el cual está basado en todas los m_1 puntos de muestreo. Entonces la siguiente identidad es válida (la expresión del lado derecho del signo de la igualdad es un promedio ponderado

$\hat{T}_{y\pi,rep}(t_1)$):

$$\hat{T}_{y\pi,rep}(t_1) = \frac{m_1 \hat{T}_{y\pi,rep}(t_1, m_1) + m_{12} \hat{T}_{y\pi,rep}(t_1, m_{12})}{m_1 + m_{12}} = u_1 \hat{T}_{y\pi,rep}(t_1, m_1) + (1 - u_1) \hat{T}_{y\pi,rep}(t_1, m_{12})$$

$$= u_1 \hat{T}_{y\pi,rep}(t_1, m_1) + \hat{T}_{y\pi,rep}(t_1, m_{12}) - u_1 \hat{T}_{y\pi,rep}(t_1, m_{12})$$

\Rightarrow

$$\hat{T}_{y\pi,rep}(t_1) = u_1 \hat{T}_{y\pi,rep}(t_1, m_1) + \hat{T}_{y\pi,rep}(t_1, m_{12}) - u_1 \hat{T}_{y\pi,rep}(t_1, m_{12})$$

\Rightarrow

$$\hat{T}_{y\pi,rep}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12}) = u_1 \hat{T}_{y\pi,rep}(t_1, m_1) - u_1 \hat{T}_{y\pi,rep}(t_1, m_{12})$$

\Rightarrow

$$\hat{T}_{y\pi,rep}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12}) = u_1 (\hat{T}_{y\pi,rep}(t_1, m_1) - \hat{T}_{y\pi,rep}(t_1, m_{12}))$$

\Rightarrow

$$\hat{T}_{y\pi,rep}(t_1, m_1) - \hat{T}_{y\pi,rep}(t_1, m_{12}) = \frac{\hat{T}_{y\pi,rep}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12})}{u_1}$$

Donde se ha considerado que $m_1 = m_{1.} + m_{12}$, $1 = \frac{m_{1.}}{m_1} + \frac{m_{12}}{m_1}$, $u_1 = \frac{m_{1.}}{m_1}$ y $1 - u_1 = \frac{m_{12}}{m_1}$

Sustituyendo este resultado y $a = du_1 B_{21}$ en (4.15) produce

$$\begin{aligned} \hat{T}_{y,MRP}(t_2 : t_1) &= a \hat{T}_{y\pi,rep}(t_1, m_{1.}) - a \hat{T}_{y\pi,rep}(t_1, m_{12}) \\ &+ (1-d) \hat{T}_{y\pi,rep}(t_2, m_{2.}) + d \hat{T}_{y\pi,rep}(t_2, m_{12}) \end{aligned} \quad (4.15)$$

\Rightarrow

$$\begin{aligned} \hat{T}_{y,MRP}(t_2 : t_1) &= a (\hat{T}_{y\pi,rep}(t_1, m_{1.}) - \hat{T}_{y\pi,rep}(t_1, m_{12})) \\ &+ (1-d) \hat{T}_{y\pi,rep}(t_2, m_{2.}) + d \hat{T}_{y\pi,rep}(t_2, m_{12}) \end{aligned}$$

\Rightarrow

$$\begin{aligned} \hat{T}_{y,MRP}(t_2 : t_1) &= du_1 B_{21} \left(\frac{\hat{T}_{y\pi,rep}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12})}{u_1} \right) \\ &+ (1-d) \hat{T}_{y\pi,rep}(t_2, m_{2.}) + d \hat{T}_{y\pi,rep}(t_2, m_{12}) \end{aligned}$$

\Rightarrow

$$\begin{aligned} \hat{T}_{y,MRP}(t_2 : t_1) &= dB_{21} (\hat{T}_{y\pi,rep}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12})) \\ &+ (1-d) \hat{T}_{y\pi,rep}(t_2, m_{2.}) + d \hat{T}_{y\pi,rep}(t_2, m_{12}) \end{aligned}$$

Finalmente (4.15) queda expresada como

$$\begin{aligned} \hat{T}_{y,MRP}(t_2 : t_1) &= d (\hat{T}_{y\pi,rep}(t_2, m_{12}) + B_{21} (\hat{T}_{y\pi,rep}(t_1) - \hat{T}_{y\pi,rep}(t_1, m_{12}))) \\ &+ (1-d) \hat{T}_{y\pi,rep}(t_2, m_{2.}) \end{aligned}$$

O bien

$$\hat{T}_{y,\text{MRP}}(t_2 : t_1) = d \hat{T}_{y\pi,\text{reg}}(t_2 : t_1) + (1-d) \hat{T}_{y\pi,\text{rep}}(t_2, m_2)$$

Donde

$$\hat{T}_{y\pi,\text{reg}}(t_2 : t_1) = \hat{T}_{y\pi,\text{rep}}(t_2, m_{12}) + B_{21}(\hat{T}_{y\pi,\text{rep}}(t_1) - \hat{T}_{y\pi,\text{rep}}(t_1, m_{12}))$$

Derivación de la varianza de $\hat{T}_{y,\text{MRP}}(t_2 : t_1)$

$$\hat{T}_{y,\text{MRP}}(t_2 : t_1) = \frac{w(t_2 : t_1) \hat{T}_{y\pi,\text{reg}}(t_2 : t_1) + w(t_2) \hat{T}_{y\pi,\text{rep}}(t_2, m_2)}{w_{\text{MRP}}} \quad (4.30)$$

Entonces

$$\begin{aligned} V(\hat{T}_{y,\text{MRP}}(t_2 : t_1)) &= V\left(\frac{w(t_2 : t_1) \hat{T}_{y\pi,\text{reg}}(t_2 : t_1) + w(t_2) \hat{T}_{y\pi,\text{rep}}(t_2, m_2)}{w_{\text{MRP}}}\right) \\ &= \frac{(w(t_2 : t_1))^2 V(\hat{T}_{y\pi,\text{reg}}(t_2 : t_1)) + (w(t_2))^2 V(\hat{T}_{y\pi,\text{rep}}(t_2, m_2))}{(w_{\text{MRP}})^2} \end{aligned}$$

$$V(\hat{T}_{y,\text{MRP}}(t_2 : t_1)) = \left(\frac{1}{(w_{\text{MRP}})^2}\right) \left((w(t_2 : t_1))^2 V(\hat{T}_{y\pi,\text{reg}}(t_2 : t_1)) + (w(t_2))^2 V(\hat{T}_{y\pi,\text{rep}}(t_2, m_2)) \right)$$

Pero de (4.26) y (4.27)

$$w(t_2 : t_1) = \frac{1}{V(\hat{T}_{y\pi,\text{reg}}(t_2 : t_1))} \quad (4.26)$$

$$w(t_2) = \frac{1}{V(\hat{T}_{y\pi,\text{rep}}(t_2, m_2))} \quad (4.27)$$

Entonces

$$\begin{aligned}
&= \left(\frac{1}{(w_{\text{MRP}})^2} \right) \left(\frac{1}{(V(\hat{T}_{y\pi,reg}(t_2:t_1)))^2} V(\hat{T}_{y\pi,reg}(t_2:t_1)) + \frac{1}{(V(\hat{T}_{y\pi,rep}(t_2,m.2)))^2} V(\hat{T}_{y\pi,rep}(t_2,m.2)) \right) \\
&= \left(\frac{1}{(w_{\text{MRP}})^2} \right) \left(\frac{1}{V(\hat{T}_{y\pi,reg}(t_2:t_1))} + \frac{1}{V(\hat{T}_{y\pi,rep}(t_2,m.2))} \right)
\end{aligned}$$

Pero de (4.28')

$$w_{\text{MRP}} = \frac{1}{V(\hat{T}_{y\pi,reg}(t_2:t_1))} + \frac{1}{V(\hat{T}_{y\pi,rep}(t_2,m.2))} \quad (4.28')$$

Entonces

$$V(\hat{T}_{y,\text{MRP}}(t_2:t_1)) = \left(\frac{1}{(w_{\text{MRP}})^2} \right) w_{\text{MRP}} = \frac{1}{w_{\text{MRP}}}$$

\Rightarrow

$$V(\hat{T}_{y,\text{MRP}}(t_2:t_1)) = \frac{1}{w_{\text{MRP}}} \quad (4.31)$$

ANEXO 4.2. PROPIEDADES DE MÍNIMOS CUADRADOS GENERALIZADOS

De acuerdo con Theil (1971), Esteban y Regúlez (2010) y Greene (2012), MCG presenta las propiedades estadísticas que enseguida se señalan.

Supongamos el modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde $E(\mathbf{u}) = \mathbf{0}$ y $E(\mathbf{u}\mathbf{u}') = \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega}$, siendo $\boldsymbol{\Omega}$ conocida. Además, \mathbf{Y} es un vector columna con T observaciones de la variable dependiente, \mathbf{X} es una matriz de orden $T \times k$ y $\boldsymbol{\beta}$ es un vector columna de coeficientes. Por último, \mathbf{u} es un vector columna que contiene los errores para cada observación.

Si lo que se quiere es estimar los coeficientes $\boldsymbol{\beta}$ desconocidos de forma que el estimador sea eficiente, una manera adecuada de proceder es transformar el modelo tal que sus perturbaciones sean esféricas, es decir, de media cero, varianza constante y covarianzas cero. El estimador de Mínimos Cuadrados Ordinarios en el modelo así transformado será lineal, insesgado, de mínima varianza o eficiente y consistente.

Dado que $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega}$ es simétrica y semidefinida positiva, y se ha supuesto que $\boldsymbol{\Omega}$ es conocida, existe una matriz no singular \mathbf{P} tal que $\boldsymbol{\Omega} = \mathbf{P}\mathbf{P}'$. Además, \mathbf{P} es conocida y no estocástica. La inversa de la matriz \mathbf{P} se utiliza como matriz de transformación del modelo original dado que

$$\boldsymbol{\Omega} = \mathbf{P}\mathbf{P}'$$

$$\boldsymbol{\Omega}^{-1} = (\mathbf{P}\mathbf{P}')^{-1} = (\mathbf{P}')^{-1}\mathbf{P}^{-1}$$

$$\mathbf{P}^{-1}\boldsymbol{\Omega}(\mathbf{P}')^{-1} = \mathbf{P}^{-1}\mathbf{P}\mathbf{P}'(\mathbf{P}')^{-1} = \mathbf{I}$$

Premultiplicando el modelo por \mathbf{P}^{-1} obtenemos el siguiente modelo transformado

$$\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}^{-1}\mathbf{u} \Leftrightarrow \mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u}^*$$

El modelo transformado tiene perturbaciones, \mathbf{u}^* , esféricas, es decir, de media cero, varianza constante y covarianzas cero.

$$E(\mathbf{u}^*) = E(\mathbf{P}^{-1}\mathbf{u}) = \mathbf{P}^{-1}E(\mathbf{u}) = \mathbf{0}$$

$$\begin{aligned} E(\mathbf{u}^*\mathbf{u}^{*\prime}) &= E(\mathbf{P}^{-1}\mathbf{u}\mathbf{u}'(\mathbf{P}^{-1})') = \mathbf{P}^{-1}E(\mathbf{u}\mathbf{u}')(\mathbf{P}^{-1})' = \mathbf{P}^{-1}\sigma^2\boldsymbol{\Omega}(\mathbf{P}^{-1})' = \sigma^2\mathbf{P}^{-1}\boldsymbol{\Omega}(\mathbf{P}^{-1})' \\ &= \sigma^2\mathbf{P}^{-1}\mathbf{P}\mathbf{P}'(\mathbf{P}')^{-1} = \sigma^2\mathbf{I} = \sigma^2\mathbf{I} \end{aligned}$$

Por lo tanto, en el modelo transformado se cumplen las hipótesis básicas, y el estimador MCO tendrán las propiedades habituales de linealidad, insesgadez y de varianza mínima.

Si definimos el estimador de MCO en el modelo transformado, y sustituimos las matrices transformadas por su expresión en términos de las variables originales del modelo, se obtiene la expresión del estimador de Mínimos Cuadrados Generalizados, MCG, en el modelo de interés.

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{MCO} &= (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}(\mathbf{X}^{*\prime}\mathbf{Y}^*) = (\mathbf{X}'(\mathbf{P}^{-1})'\mathbf{P}^{-1}\mathbf{X})^{-1}(\mathbf{X}'(\mathbf{P}^{-1})'\mathbf{P}^{-1}\mathbf{Y}) \\ &= (\mathbf{X}'(\mathbf{P}')^{-1}\mathbf{P}^{-1}\mathbf{X})^{-1}(\mathbf{X}'(\mathbf{P}')^{-1}\mathbf{P}^{-1}\mathbf{Y}) = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Y}) \\ &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}) = MCG \end{aligned}$$

Ya que,

$$\mathbf{Y}^* = \mathbf{P}^{-1}\mathbf{Y}, \mathbf{X}^* = \mathbf{P}^{-1}\mathbf{X}, (\mathbf{P}^{-1})' = (\mathbf{P}')^{-1}, \text{ y } \boldsymbol{\Omega}^{-1} = (\mathbf{P}\mathbf{P}')^{-1} = (\mathbf{P}')^{-1}\mathbf{P}^{-1}$$

Si $\boldsymbol{\Omega}$ (o $\boldsymbol{\Sigma}$ en su caso) es conocida el estimador es inmediatamente calculable.

Propiedades del estimador MCG

Linealidad

Dado que \mathbf{X} y $\boldsymbol{\Omega}$ son no estocásticas el estimador MCG es lineal en la perturbación \mathbf{u} , ya que se puede expresar como una combinación lineal del vector de perturbaciones \mathbf{u} y la matriz de constantes no estocásticas $\mathbf{C} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}$, siendo \mathbf{u} lo único aleatorio de su expresión:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{MCG} &= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{Y}) = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u})) \\ &= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})\boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{u} = \boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{u} \\ &= \boldsymbol{\beta} + \mathbf{C}\mathbf{u} \end{aligned}$$

Insesgadez

Dado que \mathbf{X} y $\mathbf{\Omega}$ son no estocásticas y $E(\mathbf{u}) = \mathbf{0}$ el estimador MCG es insesgado:

$$E(\hat{\boldsymbol{\beta}}_{MCG}) = E(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{u}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}E(\mathbf{u}) = \boldsymbol{\beta}$$

Varianza mínima

$$\begin{aligned} V(\hat{\boldsymbol{\beta}}_{MCG}) &= E[(\hat{\boldsymbol{\beta}}_{MCG} - E(\hat{\boldsymbol{\beta}}_{MCG}))(\hat{\boldsymbol{\beta}}_{MCG} - E(\hat{\boldsymbol{\beta}}_{MCG}))'] = E[(\hat{\boldsymbol{\beta}}_{MCG} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_{MCG} - \boldsymbol{\beta})'] \\ &= E[((\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{u})((\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{u})'] \\ &= E[((\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{u})(\mathbf{u}'\mathbf{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1})] \\ &= E[(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{u}\mathbf{u}'\mathbf{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}E(\mathbf{u}\mathbf{u}')\mathbf{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\sigma^2\mathbf{\Omega}\mathbf{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{\Omega}\mathbf{\Omega}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \end{aligned}$$

Ya que,

$$\hat{\boldsymbol{\beta}}_{MCG} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{u} \text{ y } E(\mathbf{u}\mathbf{u}') = \boldsymbol{\Sigma} = \sigma^2\mathbf{\Omega}$$

El Teorema de Gauss-Markov aplicado en el modelo transformado garantiza que el estimador MCG de $\boldsymbol{\beta}$ es el de varianza mínima entre los estimadores lineales e insesgados.

Distribución en muestras finitas

Bajo el supuesto de normalidad de las perturbaciones $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{\Omega})$, se tiene que

$$\hat{\boldsymbol{\beta}}_{MCG} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1})$$

Consistencia

Se puede demostrar la consistencia del estimador MCG mediante las condiciones suficientes de consistencia.

Sea $\lim_{T \rightarrow \infty} \frac{\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}}{T} = \mathbf{G}$ finita, definida positiva y no singular

$$1). \lim_{T \rightarrow \infty} E(\hat{\boldsymbol{\beta}}_{MCG}) = \boldsymbol{\beta}$$

$$E(\hat{\boldsymbol{\beta}}_{MCG}) = E(\boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{u}) = \boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}E(\mathbf{u}) = \boldsymbol{\beta}$$

$$\lim_{T \rightarrow \infty} E(\hat{\boldsymbol{\beta}}_{MCG}) = \lim_{T \rightarrow \infty} (\boldsymbol{\beta}) = \boldsymbol{\beta}$$

$$2). \lim_{T \rightarrow \infty} V(\hat{\boldsymbol{\beta}}_{MCG}) = 0$$

$$V(\hat{\boldsymbol{\beta}}_{MCG}) = \sigma^2 (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} = \frac{\sigma^2}{T} \left(\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{T} \right)^{-1} = 0\mathbf{G}^{-1} = 0$$

Ya que, $(k\mathbf{A})^{-1} = k^{-1}\mathbf{A}^{-1}$

Luego $p \lim \hat{\boldsymbol{\beta}}_{MCG} = \boldsymbol{\beta}$, con lo que MCG es consistente.

Distribución asintótica

i) Sea $\mathbf{u}^* \sim iid(\mathbf{0}, \sigma^2\mathbf{I})$

ii) $E(\mathbf{X}^*\mathbf{u}^*) = \mathbf{X}^* E(\mathbf{u}^*) = \mathbf{0}$, ya que \mathbf{X} es no estocástica.

iii) $\lim_{T \rightarrow \infty} \left(\frac{1}{T} \mathbf{X}^{*'} \mathbf{X}^* \right) = \mathbf{Q}^* = \mathbf{G}$ finita, simétrica y definida positiva.

Entonces, se cumplen los dos resultados siguientes:

$$a) \quad p \lim \left(\frac{\mathbf{X}^* \mathbf{u}^*}{T} \right) = \mathbf{0} \quad \text{y} \quad \frac{\mathbf{X}^* \mathbf{u}^*}{\sqrt{T}} \xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{G})$$

Por lo tanto,

$$\begin{aligned} \sqrt{T}(\hat{\boldsymbol{\beta}}_{MCG} - \boldsymbol{\beta}) &= \left(\frac{\mathbf{X}^{*'} \mathbf{X}^*}{T} \right)^{-1} \frac{\mathbf{X}^* \mathbf{u}^*}{\sqrt{T}} \xrightarrow{d} N\left(\mathbf{0}, \lim_{T \rightarrow \infty} \left(\frac{1}{T} \mathbf{X}^{*'} \mathbf{X}^* \right)^{-1} \sigma^2 \mathbf{G} \lim_{T \rightarrow \infty} \left(\frac{1}{T} \mathbf{X}^{*'} \mathbf{X}^* \right)\right) \\ &\xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{G}^{-1} \mathbf{G} \mathbf{G}^{-1}) \\ &\xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{G}^{-1}) \end{aligned}$$

De donde,

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{MCG} - \boldsymbol{\beta}) = \left(\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{T} \right)^{-1} \frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{u}}{\sqrt{T}} \xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{G}^{-1})$$

Ya que

$$\boldsymbol{\Omega}^{-1} = (\mathbf{P}')^{-1} \mathbf{P}^{-1}$$

$$\mathbf{X}^* = \mathbf{P}^{-1}\mathbf{X} \Rightarrow \mathbf{X}^{*'} = (\mathbf{P}^{-1}\mathbf{X})' = \mathbf{X}'(\mathbf{P}^{-1})' = \mathbf{X}'(\mathbf{P}')^{-1}$$

$$\mathbf{X}'^* \mathbf{X}^* = \mathbf{X}'(\mathbf{P}')^{-1} \mathbf{P}^{-1} \mathbf{X} = \mathbf{X}'\mathbf{\Omega}^{-1} \mathbf{X}$$

$$\mathbf{X}'^* \mathbf{u}^* = \mathbf{X}'(\mathbf{P}')^{-1} \mathbf{P}^{-1} \mathbf{u} = \mathbf{X}'\mathbf{\Omega}^{-1} \mathbf{u}$$

Método de Mínimos Cuadrados Generalizados Factibles (MCGF)

Hasta ahora hemos supuesto que se conoce $\mathbf{\Omega}$ ó $\mathbf{\Sigma}$ en $E(\mathbf{uu}') = \mathbf{\Sigma} = \sigma^2 \mathbf{\Omega}$, pero en la práctica la mayoría de las veces $\mathbf{\Omega}$ ó $\mathbf{\Sigma}$ son desconocidas. En este caso el estimador MCG no es directamente calculable, ya que en su expresión aparecen estas matrices. La solución habitual es sustituir $\mathbf{\Omega}$ ó $\mathbf{\Sigma}$ por una estimación suya en la expresión del estimador de MCG. Este es el estimador de Mínimos Cuadrados Generalizados Factibles, MCGF:

$$\hat{\boldsymbol{\beta}}_{MCGF} = (\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{Y}) = (\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{Y})$$

Propiedades del estimador de MCGF

Bajo el supuesto de que las varianzas de las perturbaciones se han modelado correctamente, el estimador MCGF en muestras finitas es un estimador no lineal y sesgado en general, su distribución en muestras finitas no es conocida. En muestras grandes, bajo ciertas condiciones de regularidad, y si $\hat{\mathbf{\Omega}}$ es un estimador consistente de $\mathbf{\Omega}$, el estimador MCGF es consistente, asintóticamente eficiente y tiene distribución asintótica conocida.

Linealidad

El estimador MCGF no es lineal en \mathbf{u} .

$$\hat{\boldsymbol{\beta}}_{MCGF} = (\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{Y}) = \boldsymbol{\beta} + (\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{u}$$

En esta expresión aparecen \mathbf{u} y $\hat{\mathbf{\Omega}}$, ambas son variables aleatorias, ya que $\hat{\mathbf{\Omega}}$ es un estimador de $\mathbf{\Omega}$, luego es una matriz estocástica. Por lo tanto, $\hat{\boldsymbol{\beta}}_{MCGF}$ es una función no lineal de \mathbf{u} y $\hat{\mathbf{\Omega}}$.

Insesgadez

En general, el estimador de MCGF es sesgado.

$$E(\hat{\beta}_{MCGF}) = \beta + E[(\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\Omega}^{-1}\mathbf{u})] \neq \beta$$

Para determinar $E[(\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\Omega}^{-1}\mathbf{u})]$ se necesita conocer la distribución conjunta de las variables aleatorias \mathbf{u} y $\hat{\Omega}$. En general, $E[(\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\Omega}^{-1}\mathbf{u})] \neq \mathbf{0}$, y por lo tanto $E(\hat{\beta}_{MCGF}) \neq \beta$.

Varianza del estimador MCG factible

Además, $Var(\hat{\beta}_{MCGF}) = E[(\hat{\beta}_{MCGF} - E(\hat{\beta}_{MCGF}))(\hat{\beta}_{MCGF} - E(\hat{\beta}_{MCGF}))']$ es una expresión difícil de obtener analíticamente. Las propiedades en muestras finitas y la distribución exacta del estimador MCGF son desconocidas. Por lo tanto, en muestras finitas su comportamiento es difícil de comparar con el de MCO e incluso puede ser peor.

Consistencia

Si $\hat{\Omega}$ es un estimador consistente de Ω , el estimador MCGF es consistente, $p \lim \hat{\beta}_{MCGF} = \beta$.

Distribución asintótica

Si las siguientes condiciones:

$$p \lim_{T \rightarrow \infty} \frac{1}{T} \mathbf{X}'(\hat{\Omega}^{-1} - \Omega^{-1})\mathbf{X} = \mathbf{0}$$

$$p \lim_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \mathbf{X}'(\hat{\Omega}^{-1} - \Omega^{-1})\mathbf{u} = \mathbf{0}$$

son todas satisfechas, entonces el estimador MCGF es asintóticamente equivalente al estimador MCG y su distribución asintótica coincide. Así se puede demostrar que

$$\sqrt{T}(\hat{\beta}_{MCGF} - \beta) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{G}^{-1}) \text{ con } \lim_{T \rightarrow \infty} \frac{\mathbf{X}'\Omega^{-1}\mathbf{X}}{T} = \mathbf{G}$$

ANEXO 4.3. DESARROLLO DEL PRODUCTO DE DOS FUNCIONES DE DENSIDAD NORMAL

Según Bromiley (2004) el producto de dos FDP Gaussianas es una FDP Gaussiana escalada. Para demostrar esto, considérese el siguiente desarrollo.

Sea $f(x)$ y $g(x)$ FDP Gaussianas con medias μ_f y μ_g y desviaciones estándar σ_f y σ_g , entonces,

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_f} e^{-\frac{(x-\mu_f)^2}{2\sigma_f^2}} \quad (1)$$

y

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\frac{(x-\mu_g)^2}{2\sigma_g^2}} \quad (2)$$

Su producto es

$$f(x)g(x) = \frac{1}{\sqrt{2\pi}\sigma_f} e^{-\frac{(x-\mu_f)^2}{2\sigma_f^2}} \times \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\frac{(x-\mu_g)^2}{2\sigma_g^2}} = \frac{1}{2\pi\sigma_f\sigma_g} e^{-\left(\frac{(x-\mu_f)^2}{2\sigma_f^2} + \frac{(x-\mu_g)^2}{2\sigma_g^2}\right)} \quad (3)$$

Sea

$$\begin{aligned} B &= \frac{(x-\mu_f)^2}{2\sigma_f^2} + \frac{(x-\mu_g)^2}{2\sigma_g^2} = \frac{x^2 - 2x\mu_f + \mu_f^2}{2\sigma_f^2} + \frac{x^2 - 2x\mu_g + \mu_g^2}{2\sigma_g^2} = \frac{\sigma_g^2(x^2 - 2x\mu_f + \mu_f^2) + \sigma_f^2(x^2 - 2x\mu_g + \mu_g^2)}{2\sigma_f^2\sigma_g^2} \\ &= \frac{\sigma_g^2x^2 - 2\sigma_g^2x\mu_f + \sigma_g^2\mu_f^2 + \sigma_f^2x^2 - 2\sigma_f^2x\mu_g + \sigma_f^2\mu_g^2}{2\sigma_f^2\sigma_g^2} = \frac{\sigma_g^2x^2 + \sigma_f^2x^2 - 2\sigma_g^2x\mu_f - 2\sigma_f^2x\mu_g + \sigma_g^2\mu_f^2 + \sigma_f^2\mu_g^2}{2\sigma_f^2\sigma_g^2} \\ &= \frac{x^2(\sigma_g^2 + \sigma_f^2) - 2(\sigma_g^2\mu_f + \sigma_f^2\mu_g)x + \sigma_g^2\mu_f^2 + \sigma_f^2\mu_g^2}{2\sigma_f^2\sigma_g^2} = \frac{x^2 - 2\frac{\mu_f\sigma_g^2 + \mu_g\sigma_f^2}{\sigma_g^2 + \sigma_f^2}x + \frac{\mu_f^2\sigma_g^2 + \mu_g^2\sigma_f^2}{\sigma_g^2 + \sigma_f^2}}{\frac{2\sigma_f^2\sigma_g^2}{\sigma_g^2 + \sigma_f^2}} \\ &= \frac{x^2 - 2\frac{\mu_f\sigma_g^2 + \mu_g\sigma_f^2}{\sigma_g^2 + \sigma_f^2}x + \left(\frac{\mu_f\sigma_g^2 + \mu_g\sigma_f^2}{\sigma_g^2 + \sigma_f^2}\right)^2 + \frac{\mu_f^2\sigma_g^2 + \mu_g^2\sigma_f^2}{\sigma_g^2 + \sigma_f^2} - \left(\frac{\mu_f\sigma_g^2 + \mu_g\sigma_f^2}{\sigma_g^2 + \sigma_f^2}\right)^2}{\frac{2\sigma_f^2\sigma_g^2}{\sigma_g^2 + \sigma_f^2}} \\ &= \frac{x^2 - 2x\frac{\mu_f\sigma_g^2 + \mu_g\sigma_f^2}{\sigma_g^2 + \sigma_f^2} + \left(\frac{\mu_f\sigma_g^2 + \mu_g\sigma_f^2}{\sigma_g^2 + \sigma_f^2}\right)^2 + \frac{\mu_f^2\sigma_g^2 + \mu_g^2\sigma_f^2}{\sigma_g^2 + \sigma_f^2} - \left(\frac{\mu_f\sigma_g^2 + \mu_g\sigma_f^2}{\sigma_g^2 + \sigma_f^2}\right)^2}{\frac{2\sigma_f^2\sigma_g^2}{\sigma_g^2 + \sigma_f^2}} \quad (4) \end{aligned}$$

En (4), el primer término puede desarrollarse de la siguiente manera:

$$\frac{x^2 - 2x \frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{\sigma_g^2 + \sigma_f^2} + \left(\frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{\sigma_g^2 + \sigma_f^2} \right)^2}{\frac{2\sigma_f^2 \sigma_g^2}{\sigma_g^2 + \sigma_f^2}} = \frac{\left(x - \frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{\sigma_g^2 + \sigma_f^2} \right)^2}{\frac{2\sigma_f^2 \sigma_g^2}{\sigma_g^2 + \sigma_f^2}} = \frac{(x - \mu_{fg})^2}{2\sigma_{fg}^2} \quad (5)$$

Donde, $\mu_{fg} = \frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{\sigma_g^2 + \sigma_f^2}$ y $\sigma_{fg}^2 = \frac{\sigma_f^2 \sigma_g^2}{\sigma_g^2 + \sigma_f^2}$

El segundo término de (4) se puede desarrollar de la siguiente manera:

$$\begin{aligned} & \frac{\frac{\mu_f^2 \sigma_g^2 + \mu_g^2 \sigma_f^2}{\sigma_g^2 + \sigma_f^2} - \left(\frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{\sigma_g^2 + \sigma_f^2} \right)^2}{\frac{2\sigma_f^2 \sigma_g^2}{\sigma_g^2 + \sigma_f^2}} = \frac{\frac{\mu_f^2 \sigma_g^2 + \mu_g^2 \sigma_f^2}{\sigma_g^2 + \sigma_f^2} - \frac{(\mu_f \sigma_g^2 + \mu_g \sigma_f^2)^2}{(\sigma_g^2 + \sigma_f^2)^2}}{\frac{2\sigma_f^2 \sigma_g^2}{\sigma_g^2 + \sigma_f^2}} = \frac{(\sigma_g^2 + \sigma_f^2)(\mu_f^2 \sigma_g^2 + \mu_g^2 \sigma_f^2) - (\mu_f \sigma_g^2 + \mu_g \sigma_f^2)^2}{\frac{2\sigma_f^2 \sigma_g^2}{\sigma_g^2 + \sigma_f^2}} \\ & = \frac{(\mu_f^2 \sigma_g^2 \sigma_g^2 + \mu_g^2 \sigma_f^2 \sigma_g^2 + \mu_f^2 \sigma_f^2 \sigma_g^2 + \mu_g^2 \sigma_f^2 \sigma_f^2) - \mu_f^2 \sigma_g^2 \sigma_g^2 - \mu_g^2 \sigma_f^2 \sigma_f^2 - 2\mu_f \mu_g \sigma_f^2 \sigma_g^2}{\frac{2\sigma_f^2 \sigma_g^2}{\sigma_g^2 + \sigma_f^2}} \\ & = \frac{\mu_g^2 \sigma_f^2 \sigma_g^2 + \mu_f^2 \sigma_f^2 \sigma_g^2 - 2\mu_f \mu_g \sigma_f^2 \sigma_g^2}{\frac{2\sigma_f^2 \sigma_g^2}{\sigma_g^2 + \sigma_f^2}} = \frac{(\sigma_g^2 + \sigma_f^2)(\mu_g^2 \sigma_f^2 \sigma_g^2 + \mu_f^2 \sigma_f^2 \sigma_g^2 - 2\mu_f \mu_g \sigma_f^2 \sigma_g^2)}{(\sigma_g^2 + \sigma_f^2)^2 2\sigma_f^2 \sigma_g^2} \\ & = \frac{\mu_g^2 \sigma_f^2 \sigma_g^2 + \mu_f^2 \sigma_f^2 \sigma_g^2 - 2\mu_f \mu_g \sigma_f^2 \sigma_g^2}{(\sigma_g^2 + \sigma_f^2) 2\sigma_f^2 \sigma_g^2} = \frac{\sigma_f^2 \sigma_g^2 (\mu_g^2 + \mu_f^2 - 2\mu_f \mu_g)}{(\sigma_g^2 + \sigma_f^2) 2\sigma_f^2 \sigma_g^2} = \frac{\mu_g^2 + \mu_f^2 - 2\mu_f \mu_g}{2(\sigma_g^2 + \sigma_f^2)} = \frac{(\mu_f - \mu_g)^2}{2(\sigma_f^2 + \sigma_g^2)} \quad (6) \end{aligned}$$

Sustituyendo (5) y (6) en (4)

$$B = \frac{(x - \mu_f)^2}{2\sigma_f^2} + \frac{(x - \mu_g)^2}{2\sigma_g^2} = \frac{(x - \mu_{fg})^2}{2\sigma_{fg}^2} + \frac{(\mu_f - \mu_g)^2}{2(\sigma_f^2 + \sigma_g^2)} \quad (7)$$

Sustituyendo (7) en (3)

$$\begin{aligned} f(x)g(x) &= \frac{1}{2\pi\sigma_f\sigma_g} e^{-\left(\frac{(x-\mu_f)^2}{2\sigma_f^2} + \frac{(x-\mu_g)^2}{2\sigma_g^2}\right)} = \frac{1}{2\pi\sigma_f\sigma_g} e^{-\left(\frac{(x-\mu_{fg})^2}{2\sigma_{fg}^2} + \frac{(\mu_f-\mu_g)^2}{2(\sigma_f^2+\sigma_g^2)}\right)} \\ &= \frac{1}{2\pi\sigma_f\sigma_g} e^{-\frac{(x-\mu_{fg})^2}{2\sigma_{fg}^2}} e^{-\frac{(\mu_f-\mu_g)^2}{2(\sigma_f^2+\sigma_g^2)}} \end{aligned}$$

Pero,

$$\sigma_{fg}^2 = \frac{\sigma_f^2 \sigma_g^2}{\sigma_g^2 + \sigma_f^2} \Rightarrow \sigma_f^2 \sigma_g^2 = \sigma_{fg}^2 (\sigma_g^2 + \sigma_f^2) \Rightarrow \sigma_f \sigma_g = \sigma_{fg} \sqrt{(\sigma_g^2 + \sigma_f^2)}$$

Entonces,

$$f(x)g(x) = \frac{1}{\sqrt{2\pi}\sigma_{fg}} e^{-\frac{(x-\mu_{fg})^2}{2\sigma_{fg}^2}} \frac{1}{\sqrt{2\pi(\sigma_g^2 + \sigma_f^2)}} e^{-\frac{(\mu_f - \mu_g)^2}{2(\sigma_f^2 + \sigma_g^2)}}$$

Por lo tanto, el producto de dos FDP Gaussianas $f(x)$ y $g(x)$ es una PDF Gaussiana escalada:

$$f(x)g(x) = \frac{S_{fg}}{\sqrt{2\pi}\sigma_{fg}} e^{-\frac{(x-\mu_{fg})^2}{2\sigma_{fg}^2}} \alpha \frac{1}{\sqrt{2\pi}\sigma_{fg}} e^{-\frac{(x-\mu_{fg})^2}{2\sigma_{fg}^2}}$$

Donde,

$$S_{fg} = \frac{1}{\sqrt{2\pi(\sigma_g^2 + \sigma_f^2)}} e^{-\frac{(\mu_f - \mu_g)^2}{2(\sigma_f^2 + \sigma_g^2)}}, \mu_{fg} = \frac{\mu_f \sigma_g^2 + \mu_g \sigma_f^2}{\sigma_g^2 + \sigma_f^2} \text{ y } \sigma_{fg}^2 = \frac{\sigma_f^2 \sigma_g^2}{\sigma_g^2 + \sigma_f^2}$$

ANEXO 4.4. ESTIMACIONES DE ABHA OBTENIDAS MEDIANTE MCG, MRE Y ECU EN LA CONDICIÓN UNO

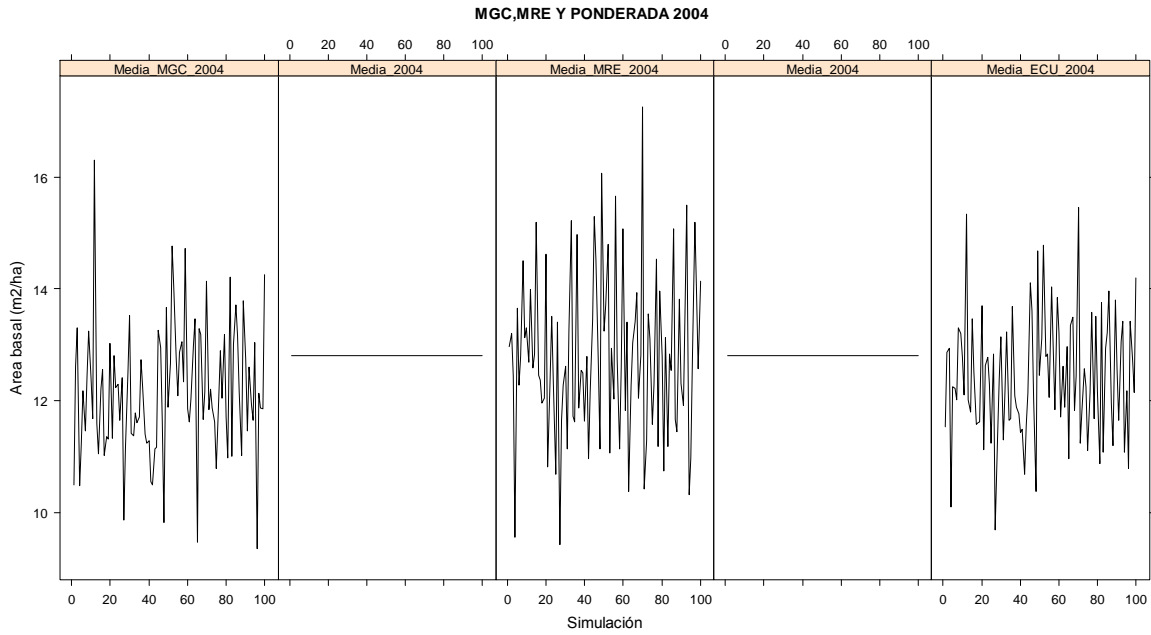


Figura 4.4.1. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2004 de la condición uno.

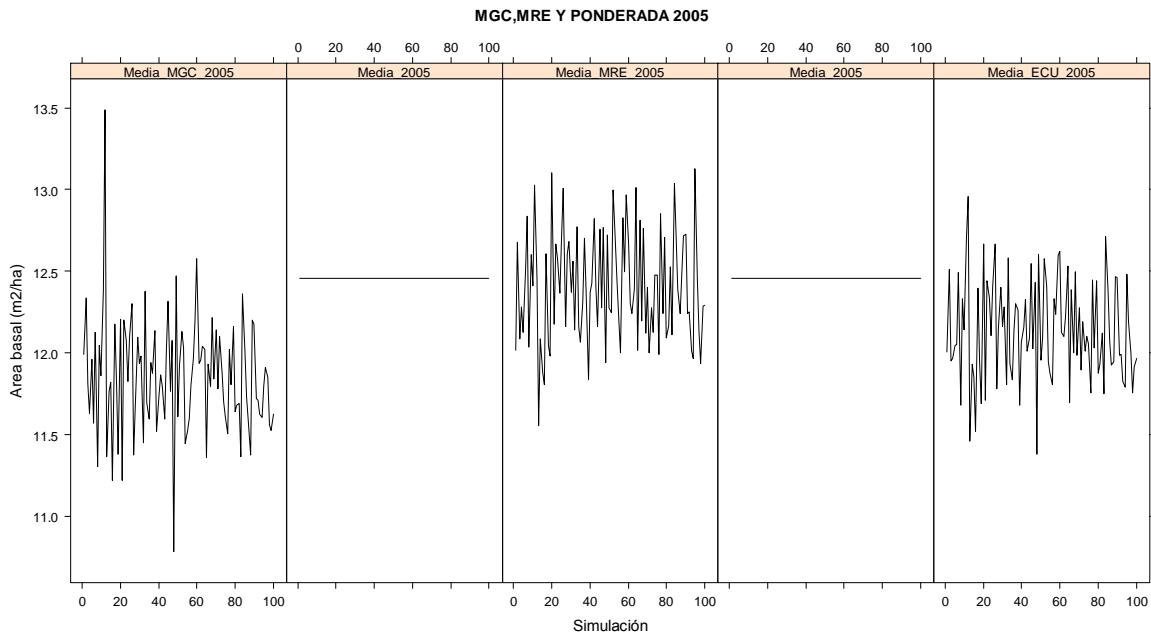


Figura 4.4.2. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2005 de la condición uno.

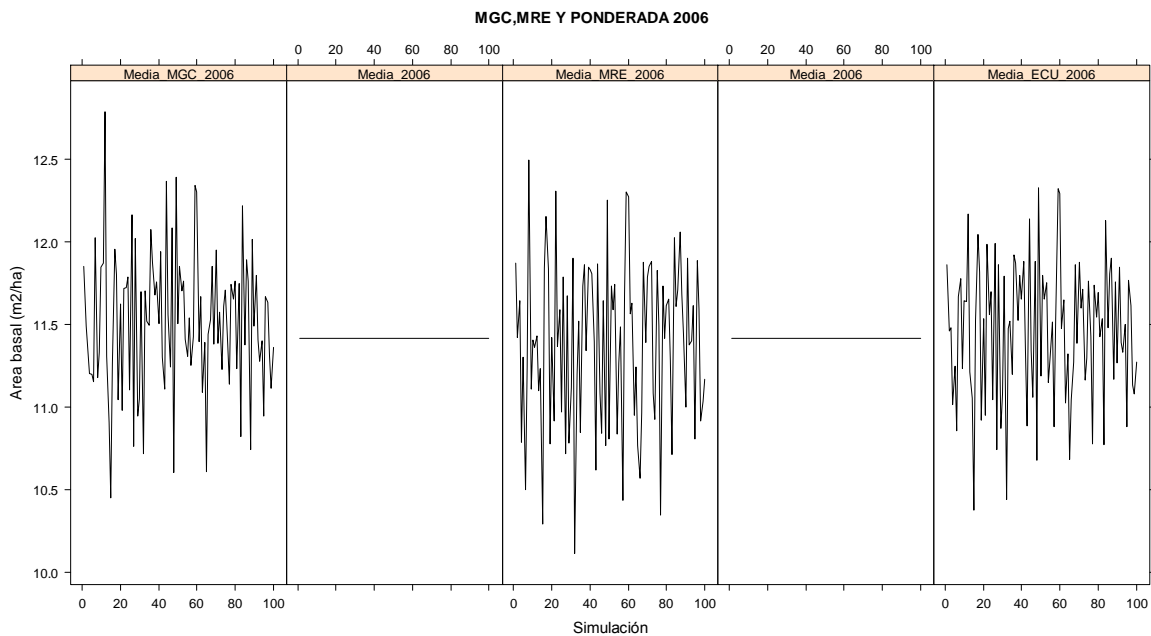


Figura 4.4.3. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2006 de la condición uno.

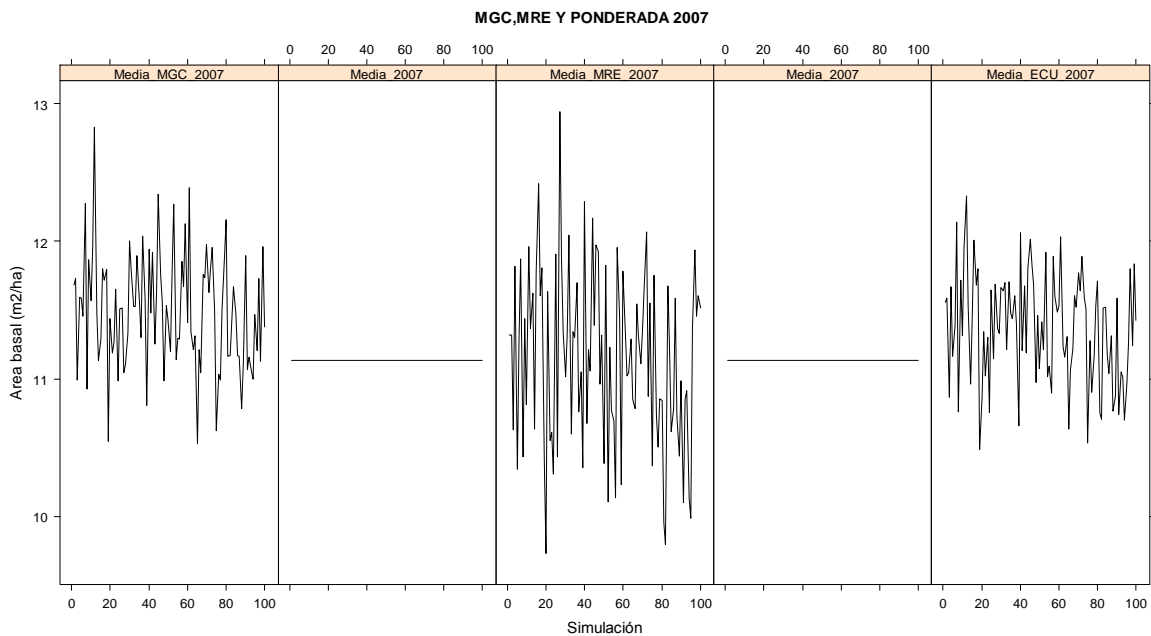


Figura 4.4.4. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2007 de la condición uno.

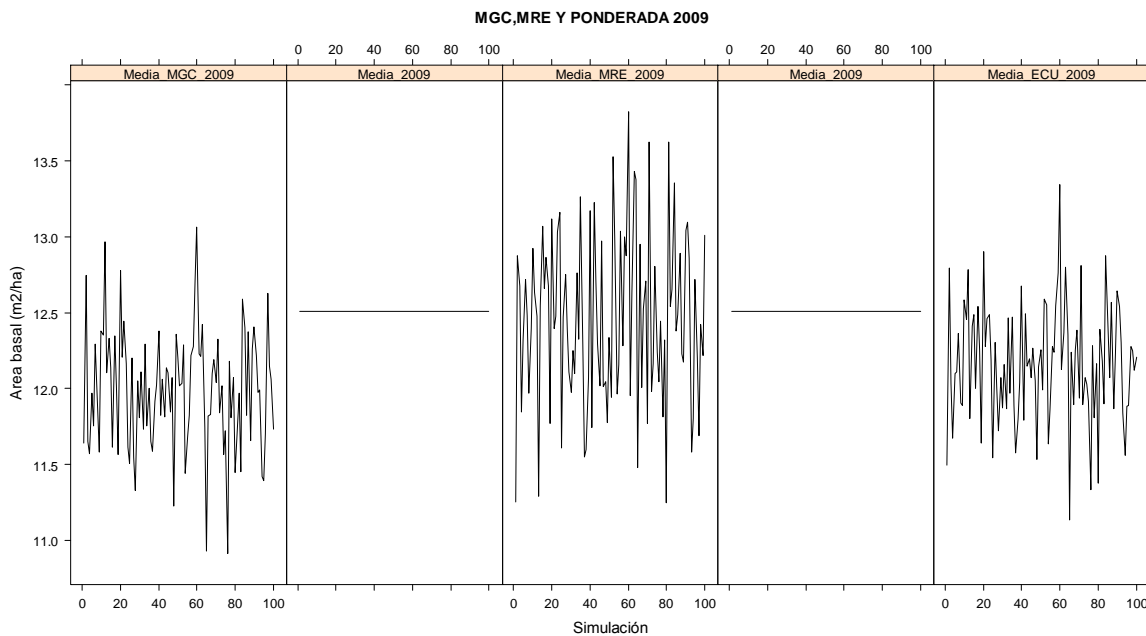


Figura 4.4.5. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2009 de la condición uno.

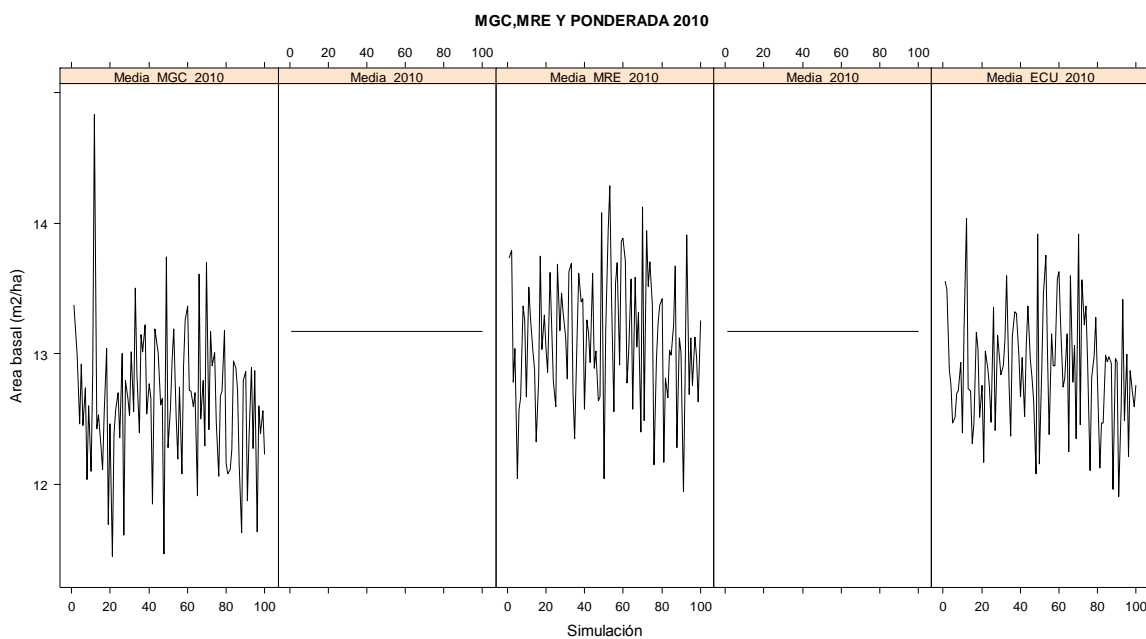


Figura 4.4.6. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2010 de la condición uno.

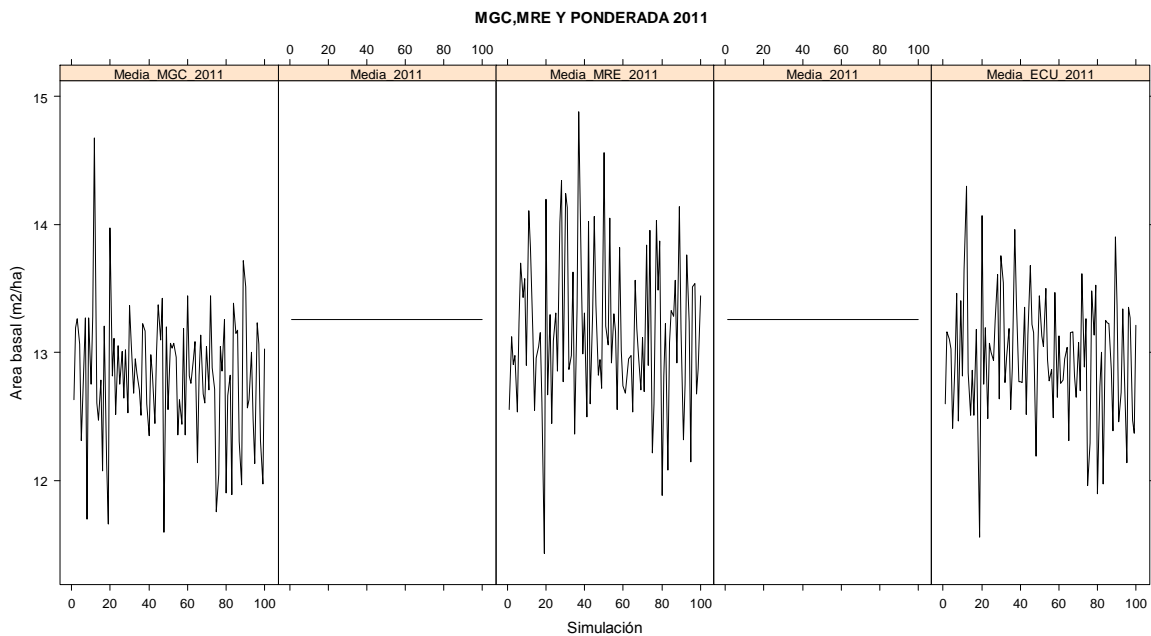


Figura 4.4.7. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2011 de la condición uno.

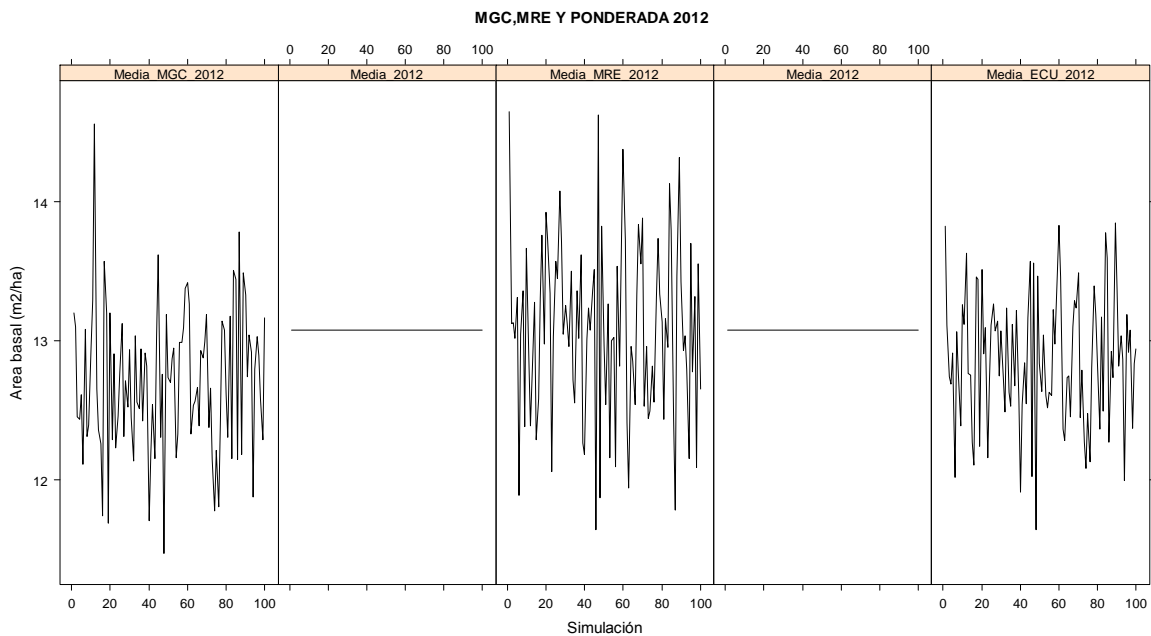


Figura 4.4.8. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2012 de la condición uno.

ANEXO 4.5. ESTIMACIONES DE ABHA OBTENIDAS MEDIANTE MCG, MRE Y ECU EN LA CONDICIÓN DOS

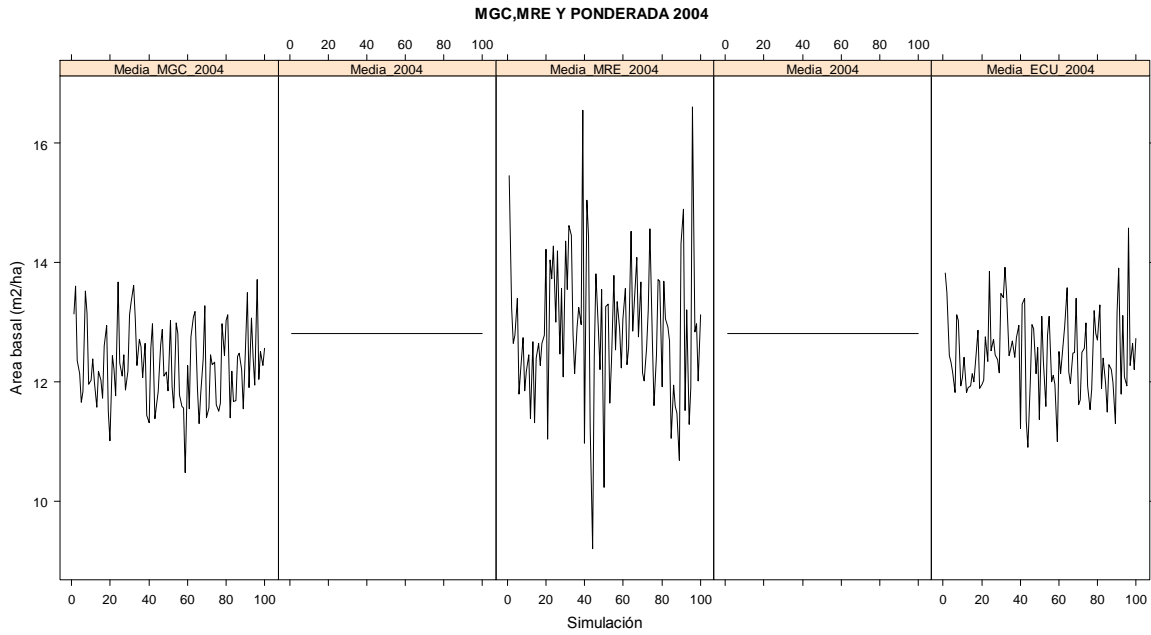


Figura 4.5.1. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2004 de la condición dos.

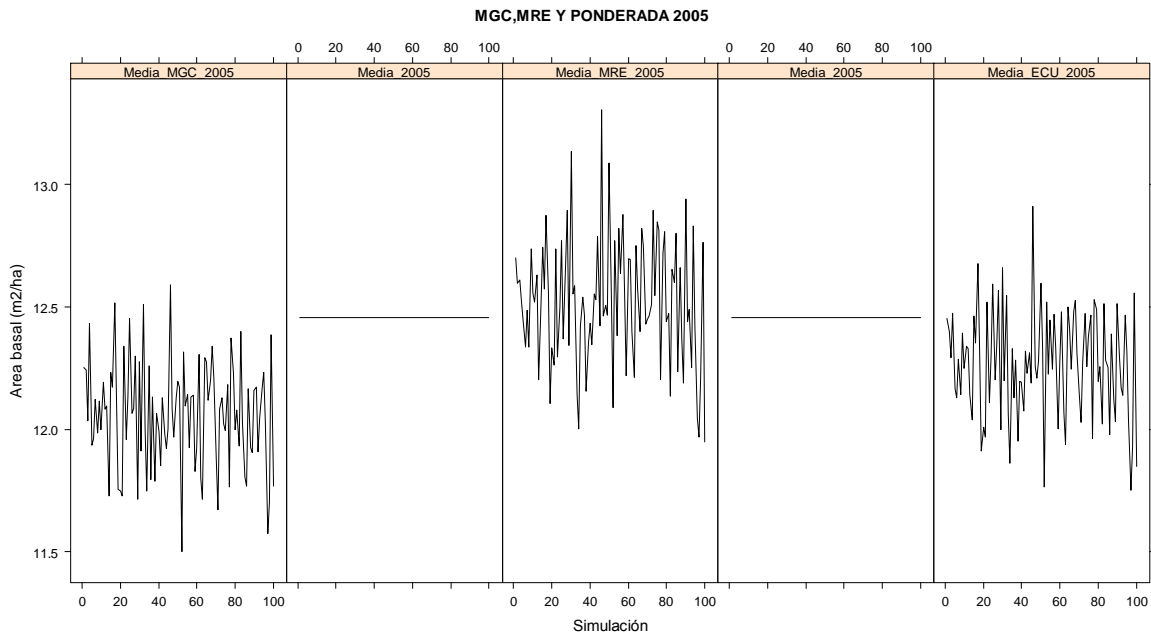


Figura 4.5.2. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2005 de la condición dos.

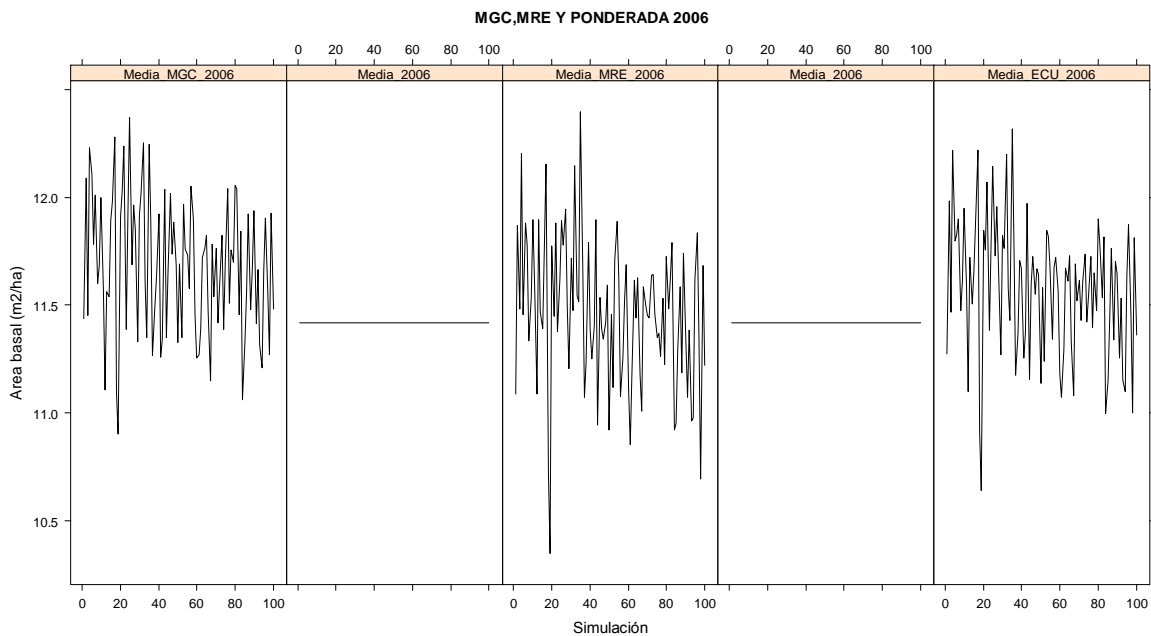


Figura 4.5.3. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2006 de la condición dos.

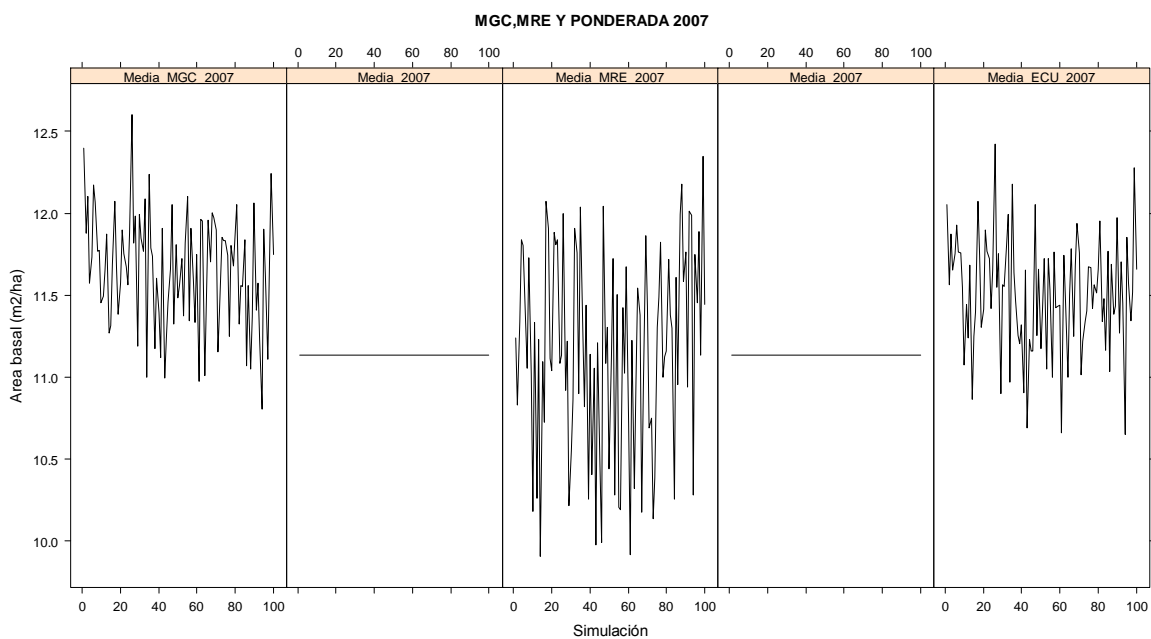


Figura 4.5.4. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2007 de la condición dos.

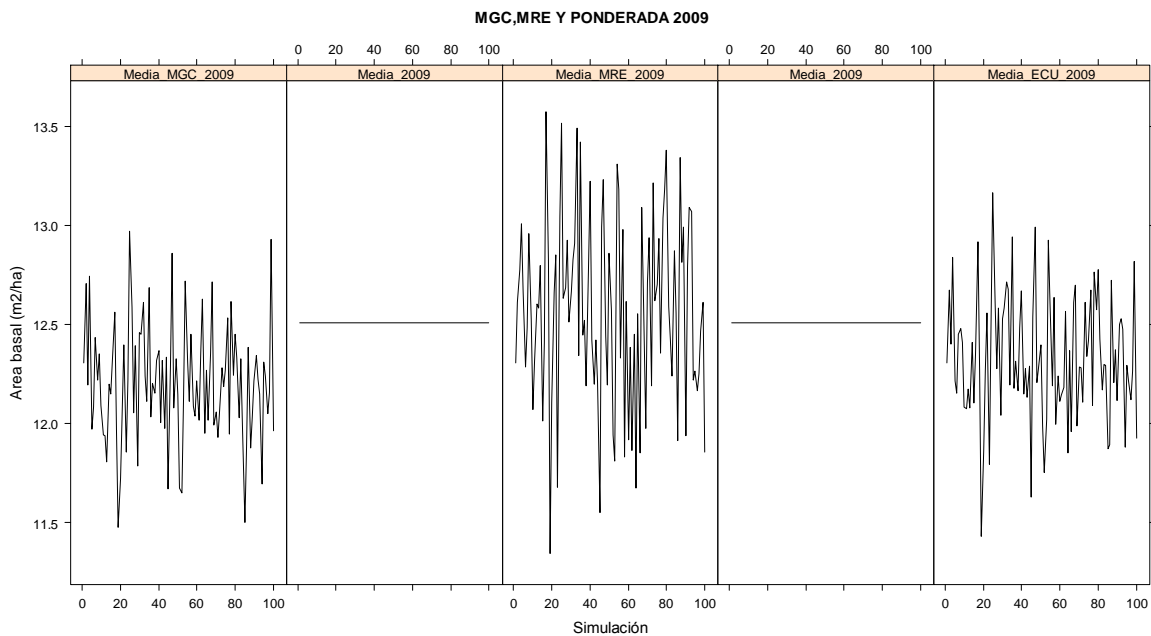


Figura 4.5.5. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2009 de la condición dos.

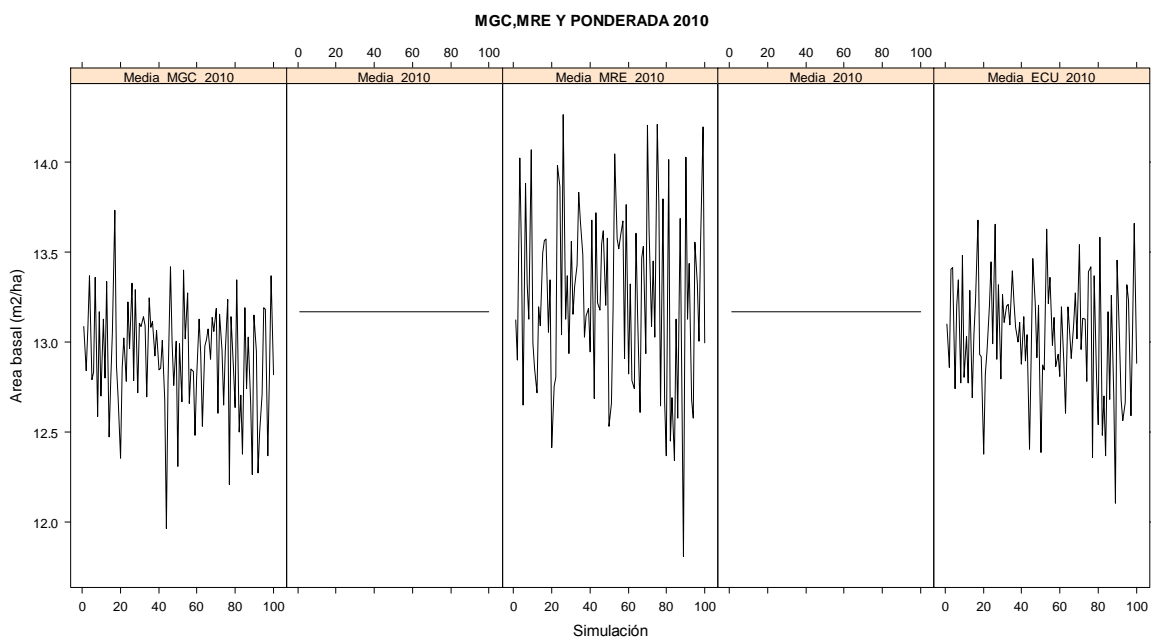


Figura 4.5.6. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2010 de la condición dos.

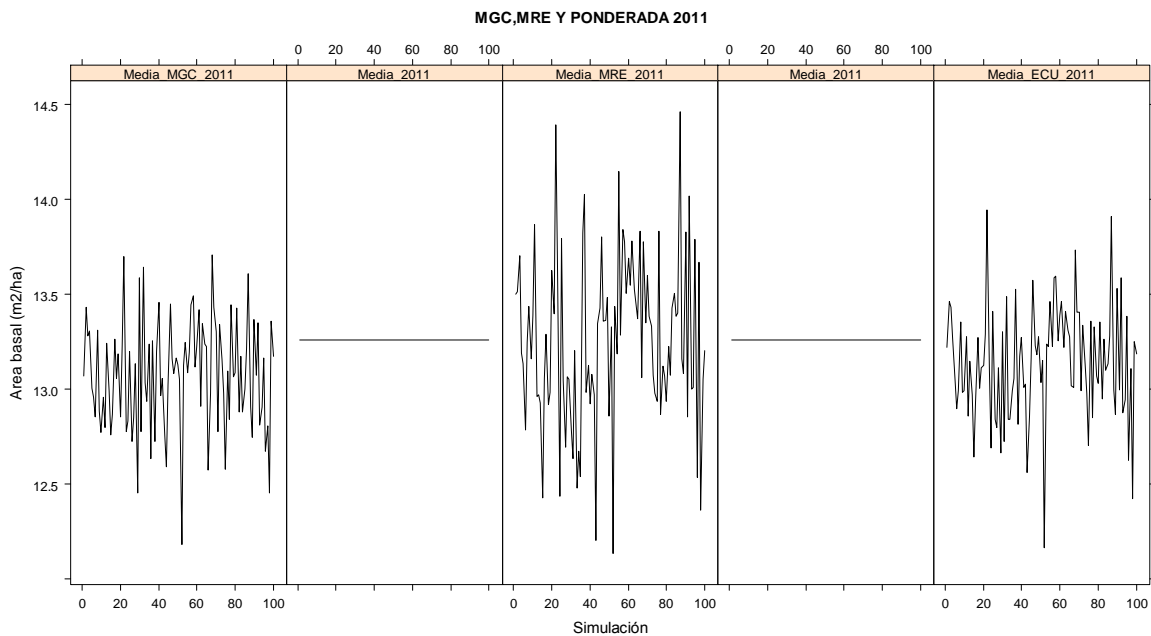


Figura 4.5.7. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2011 de la condición dos.

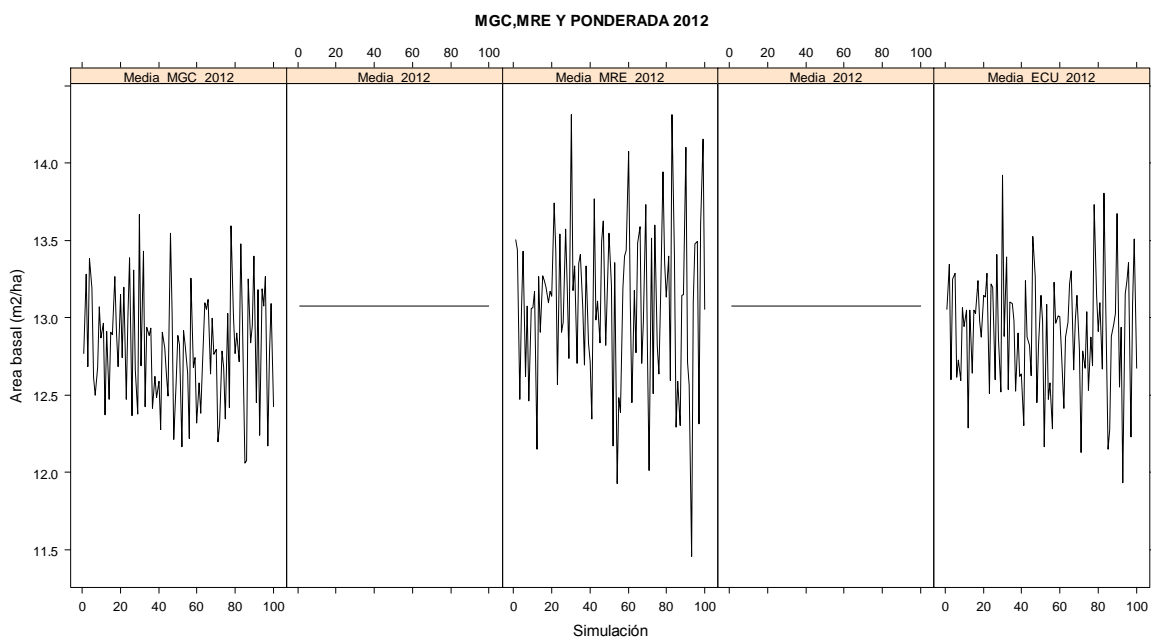


Figura 4.5.8. Estimaciones de ABHA obtenidas mediante MGC, MRE y ECU en la ocasión de medición 2012 de la condición dos.

ANEXO 4.6. INTERVALOS DE CONFIANZA AL 95% PARA ABHA GENERADOS POR ECU PARA TODAS LAS OCASIONES DE MEDICIÓN EN LA CONDICIÓN UNO

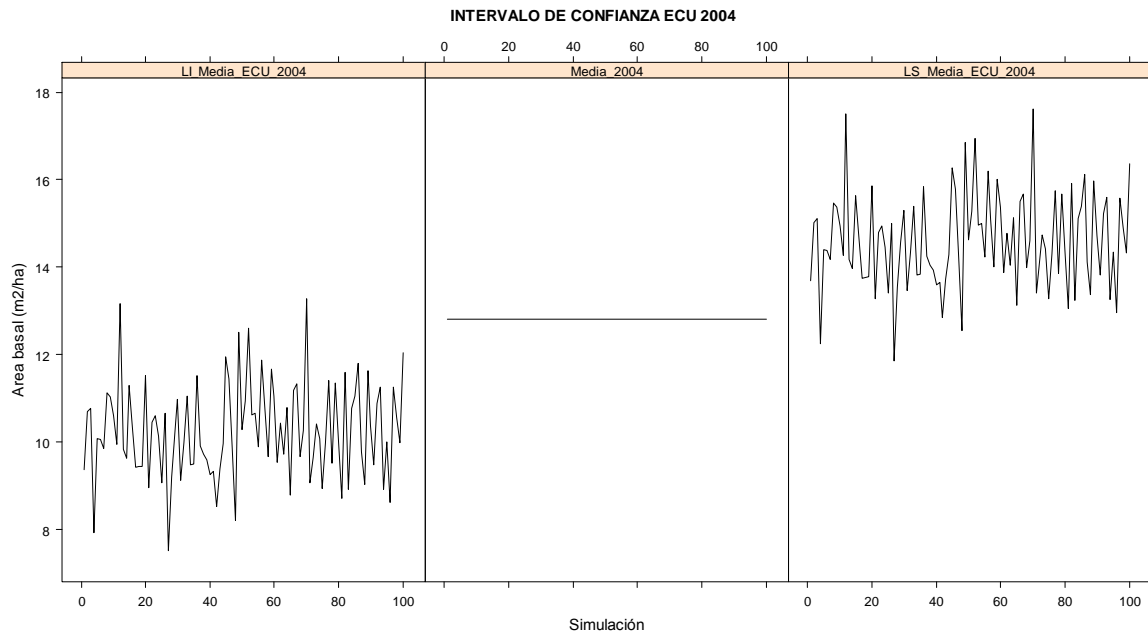


Figura 4.6.1. Intervalos de confianza ECU 2004 de la condición uno. El 95% de los intervalos de confianza para ABHA atrapan al parámetro.

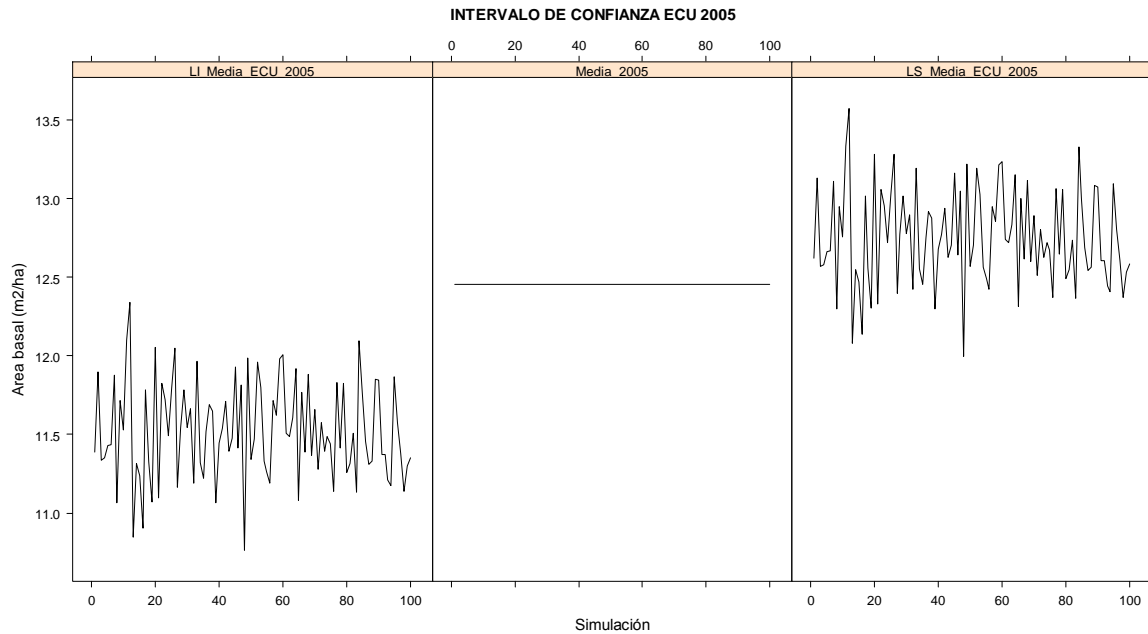


Figura 4.6.2. Intervalos de confianza ECU 2005 de la condición uno. El 83% de los intervalos de confianza para ABHA atrapan al parámetro.

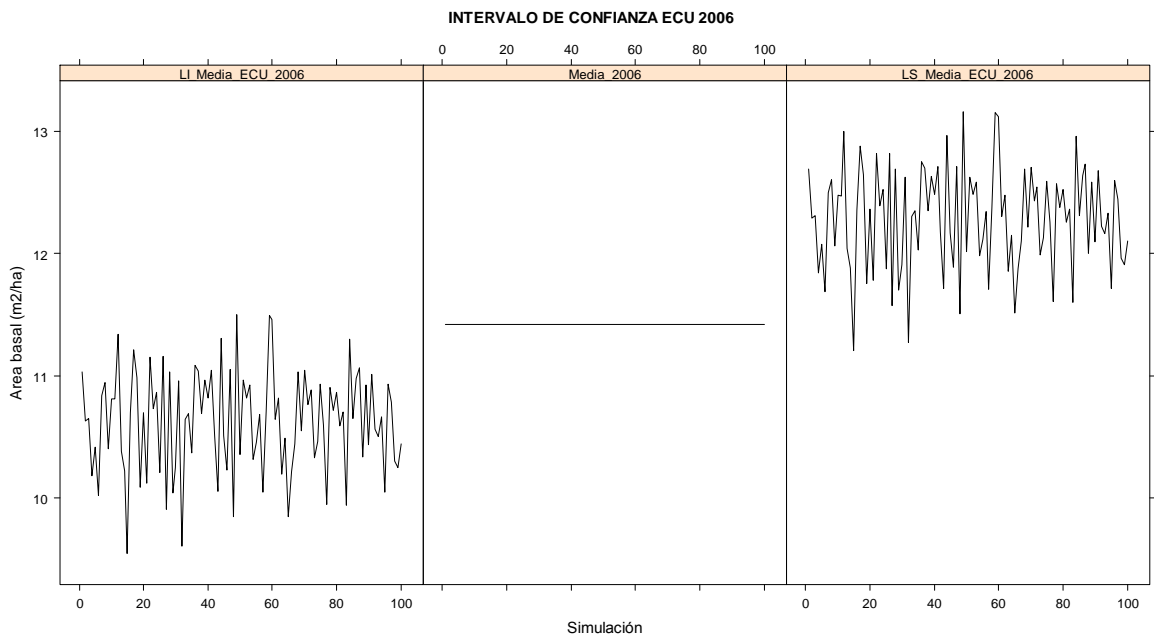


Figura 4.6.3. Intervalos de confianza ECU 2006 de la condición uno. El 95% de los intervalos de confianza para ABHA atrapan al parámetro.

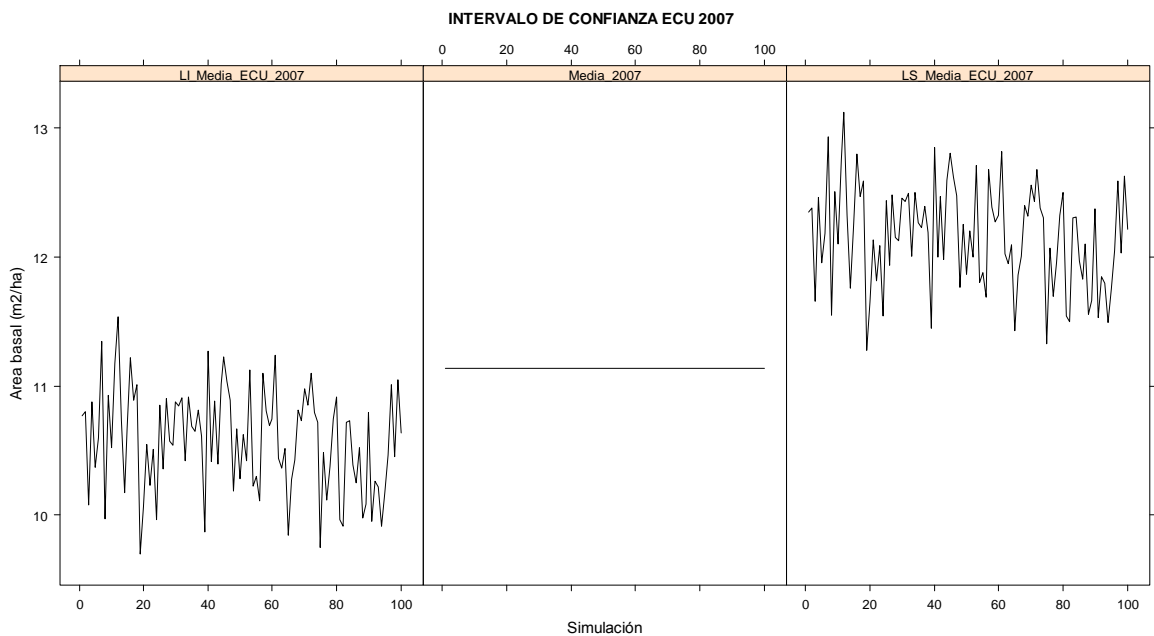


Figura 4.6.4. Intervalos de confianza ECU 2007 de la condición uno. El 93% de los intervalos de confianza para ABHA atrapan al parámetro.

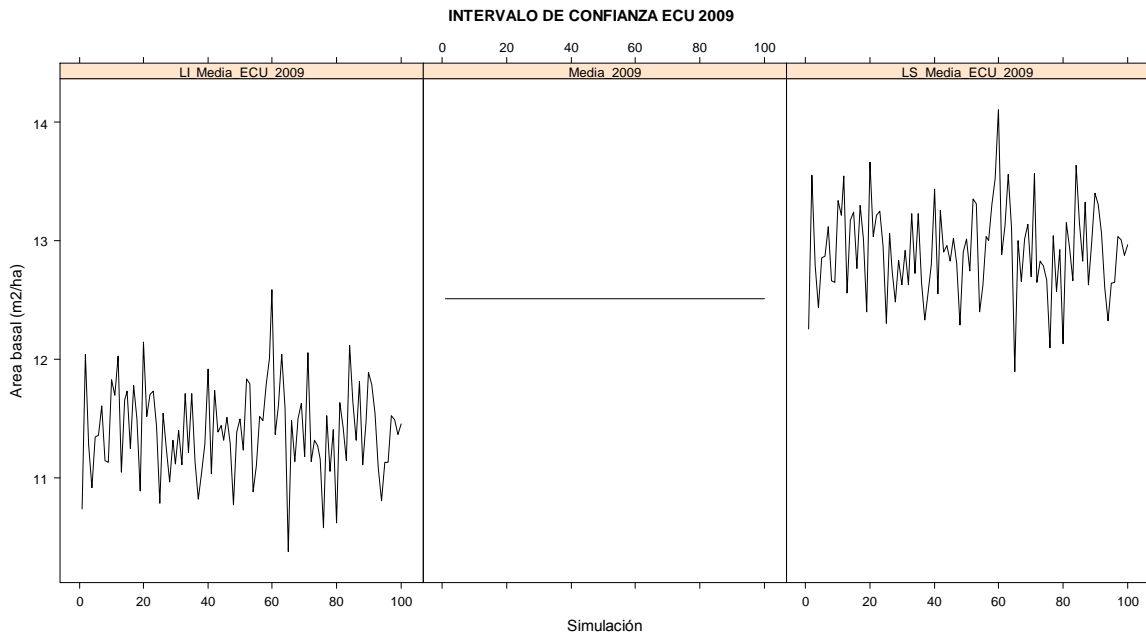


Figura 4.6.5. Intervalos de confianza ECU 2009 de la condición uno. El 87% de los intervalos de confianza para ABHA atrapan al parámetro.

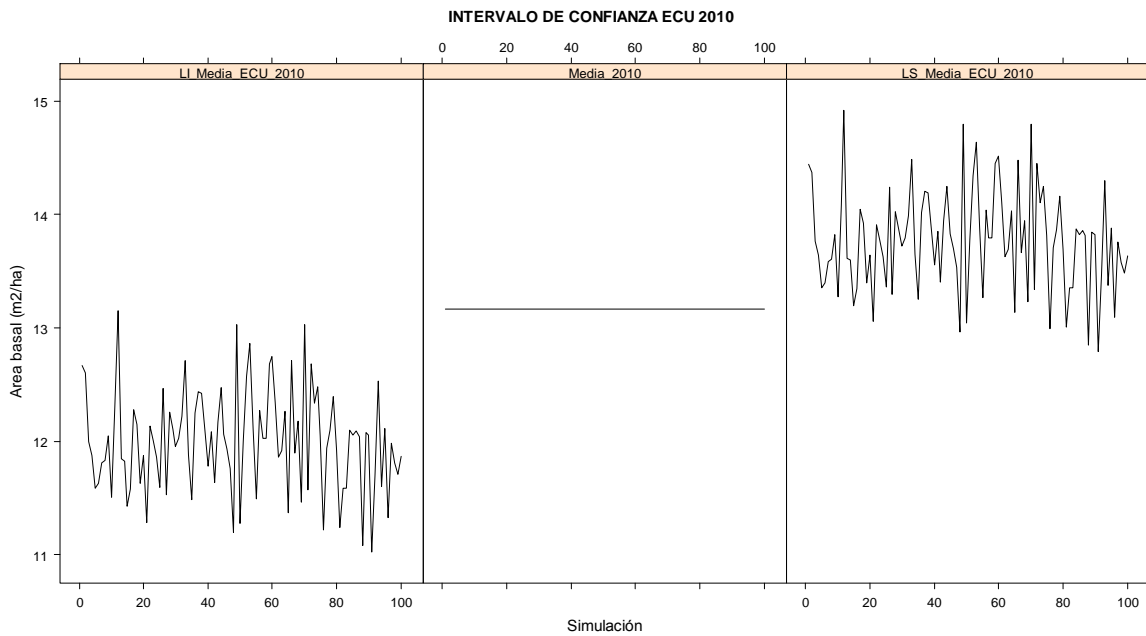


Figura 4.6.6. Intervalos de confianza ECU 2010 de la condición uno. El 91% de los intervalos de confianza para ABHA atrapan al parámetro.

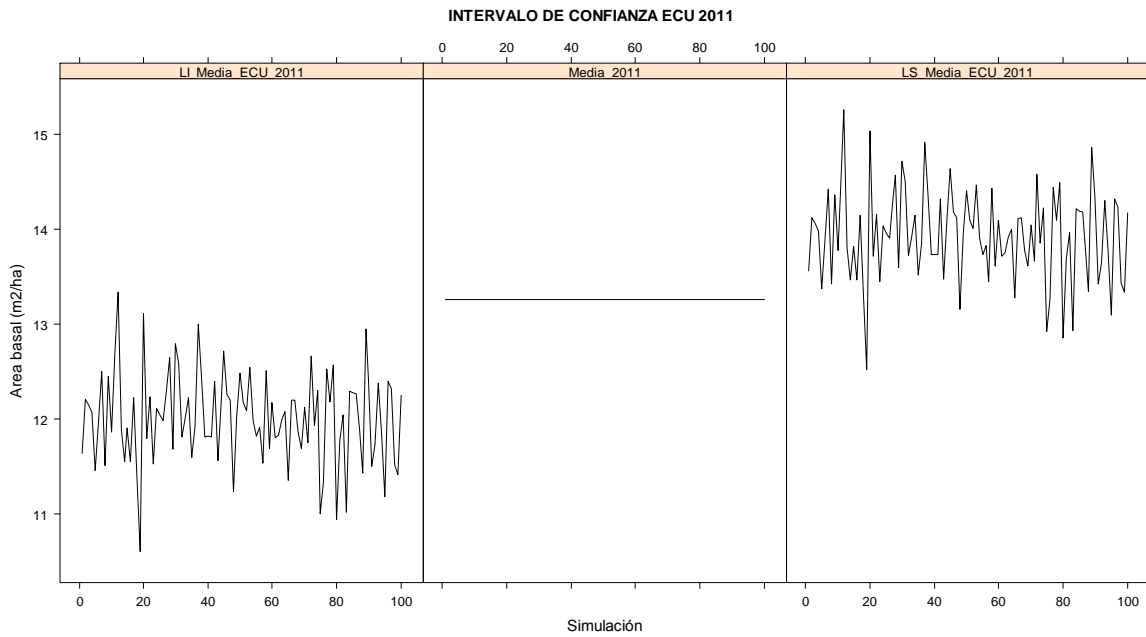


Figura 4.6.7. Intervalos de confianza ECU 2011 de la condición uno. El 92% de los intervalos de confianza para ABHA atrapan al parámetro.

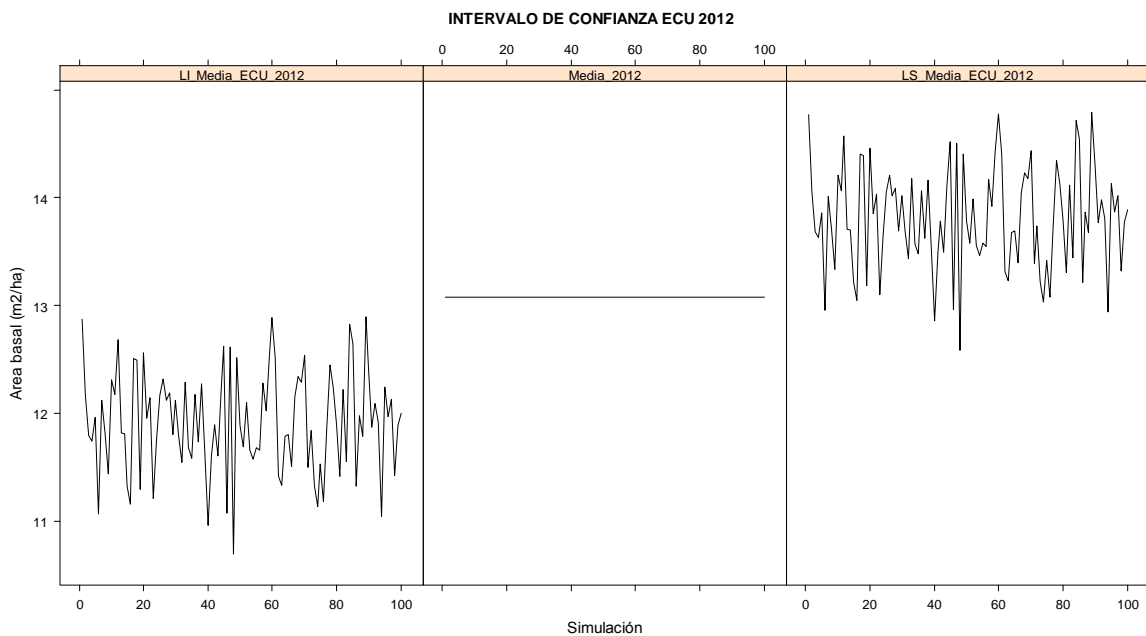


Figura 4.6.8. Intervalos de confianza ECU 2012 de la condición uno. El 92% de los intervalos de confianza para ABHA atrapan al parámetro.

ANEXO 4.7. INTERVALOS DE CONFIANZA AL 95% PARA ABHA GENERADOS POR ECU PARA TODAS LAS OCASIONES DE MEDICIÓN EN LA CONDICIÓN DOS

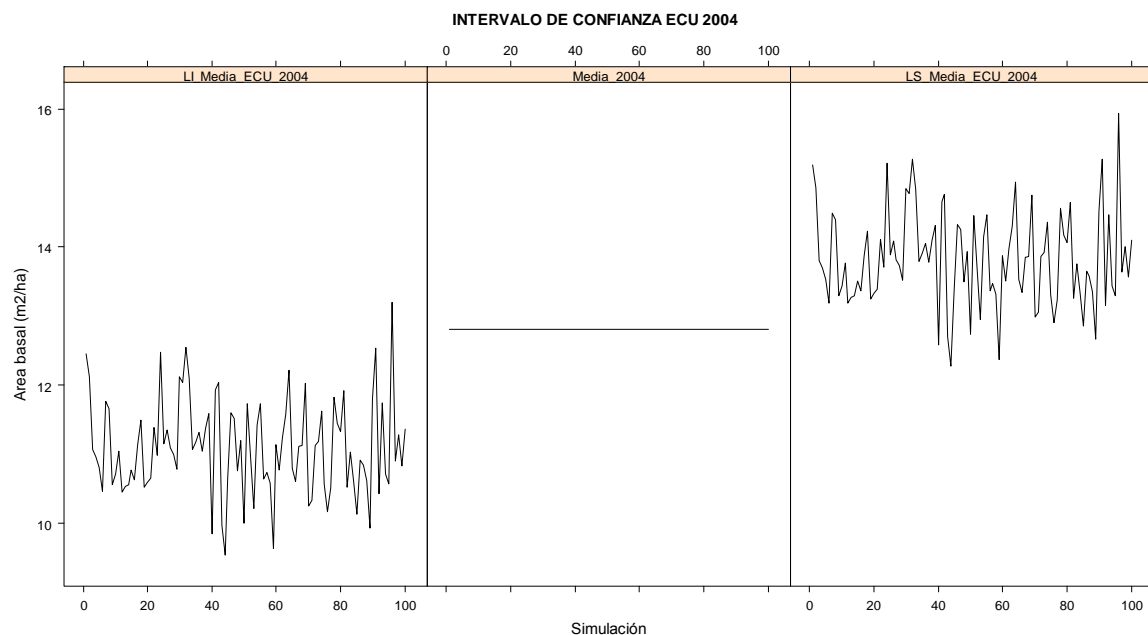


Figura 4.7.1. Intervalos de confianza ECU 2004 de la condición dos. El 93% de los intervalos de confianza para ABHA atrapan al parámetro.

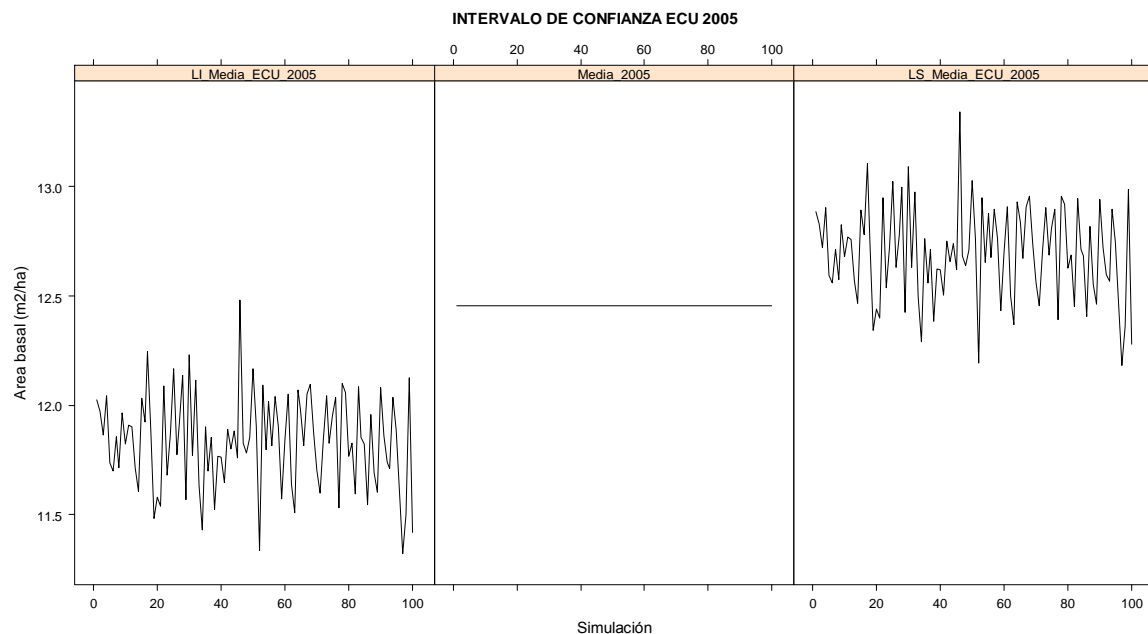


Figura 4.7.2. Intervalos de confianza ECU 2005 de la condición dos. El 83% de los intervalos de confianza para ABHA atrapan al parámetro.

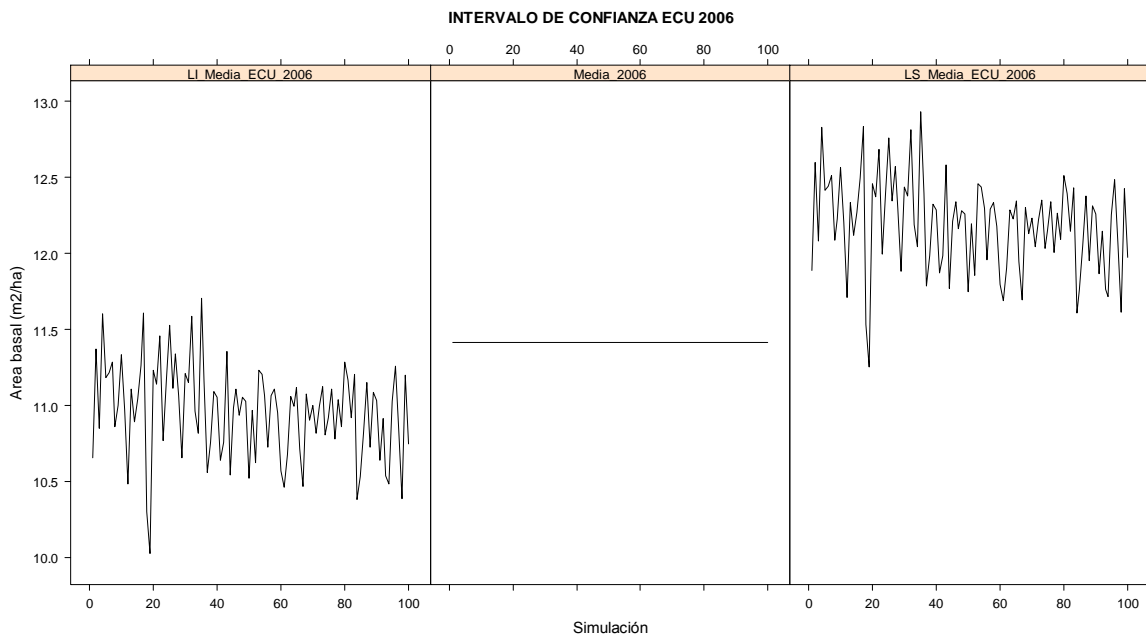


Figura 4.7.3. Intervalos de confianza ECU 2006 de la condición dos. El 93% de los intervalos de confianza para ABHA atrapan al parámetro.

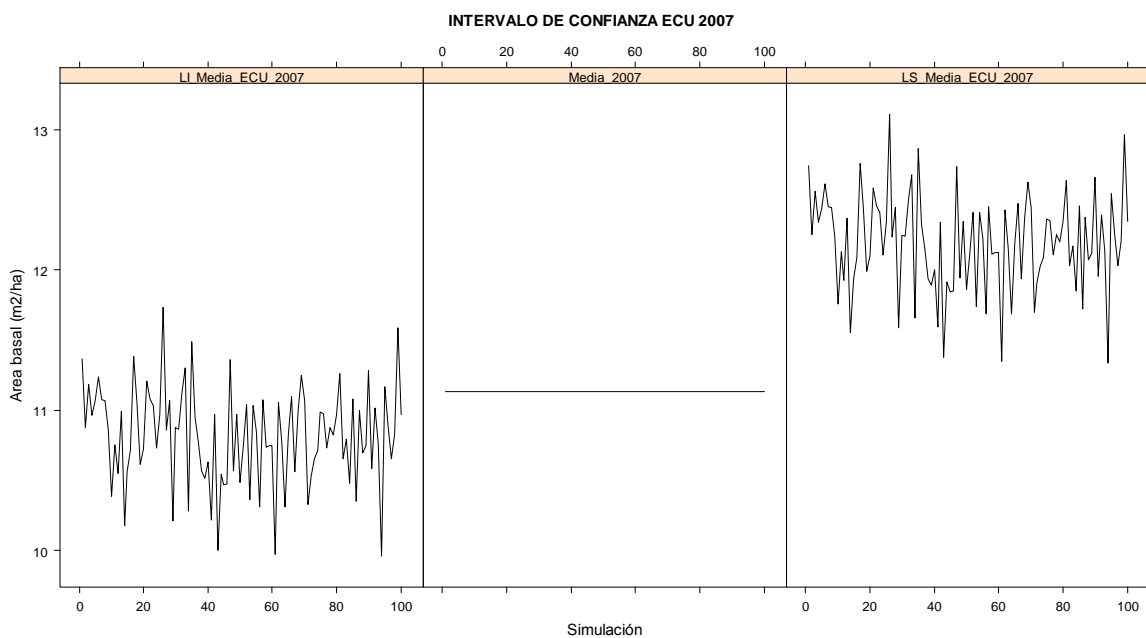


Figura 4.7.4. Intervalos de confianza ECU 2007 de la condición dos. El 86% de los intervalos de confianza para ABHA atrapan al parámetro.

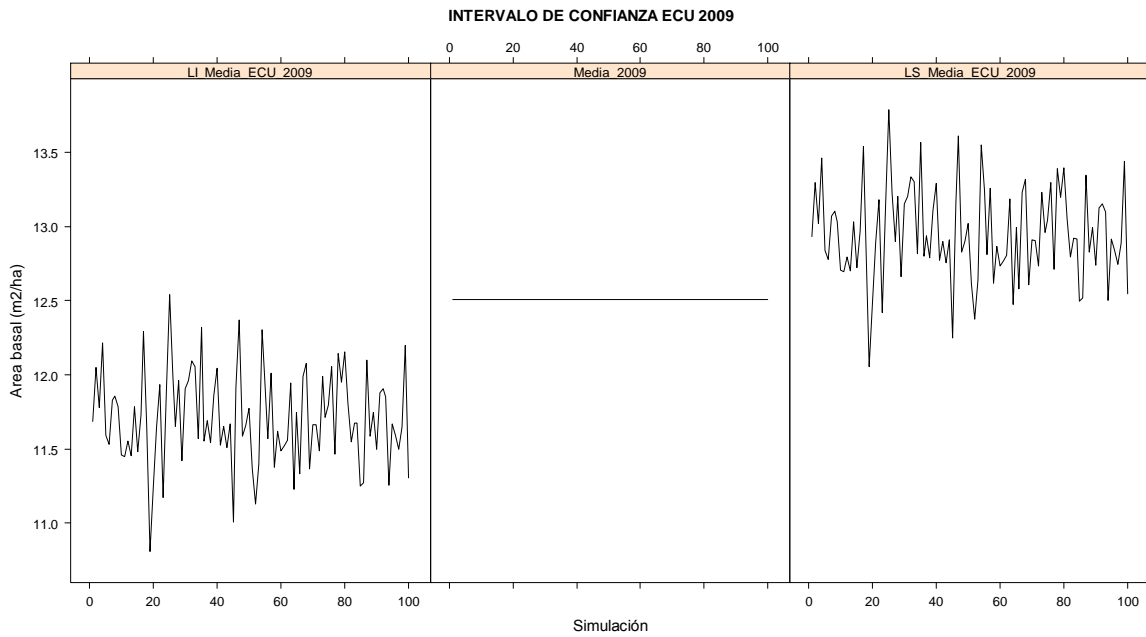


Figura 4.7.5. Intervalos de confianza ECU 2009 de la condición dos. El 91% de los intervalos de confianza para ABHA atrapan al parámetro.

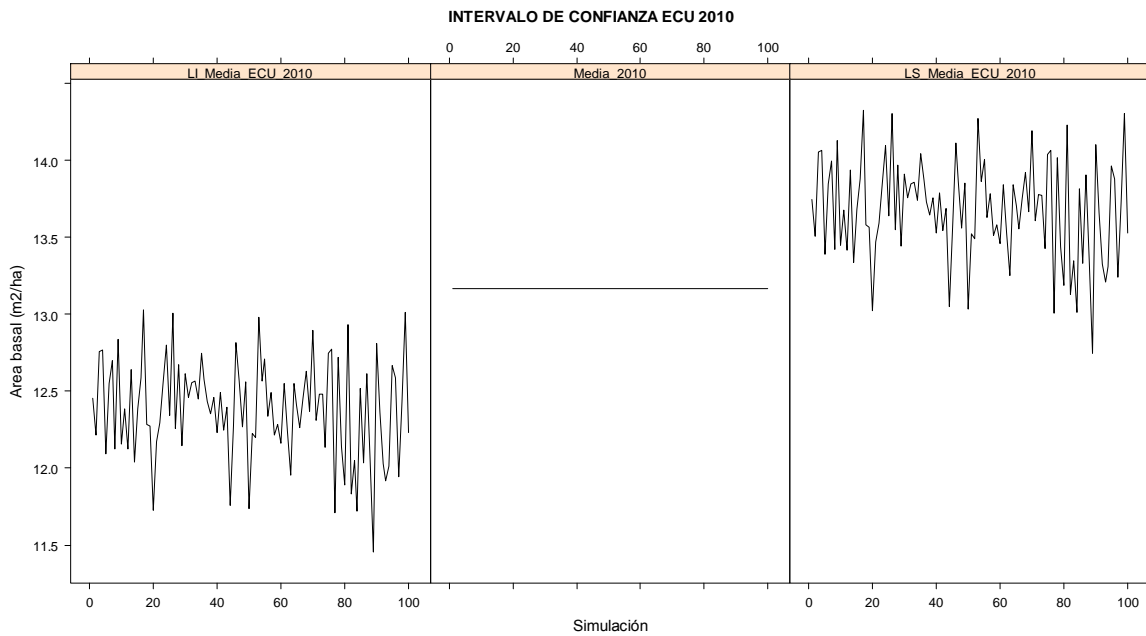


Figura 4.7.6. Intervalos de confianza ECU 2010 de la condición dos. El 93% de los intervalos de confianza para ABHA atrapan al parámetro.

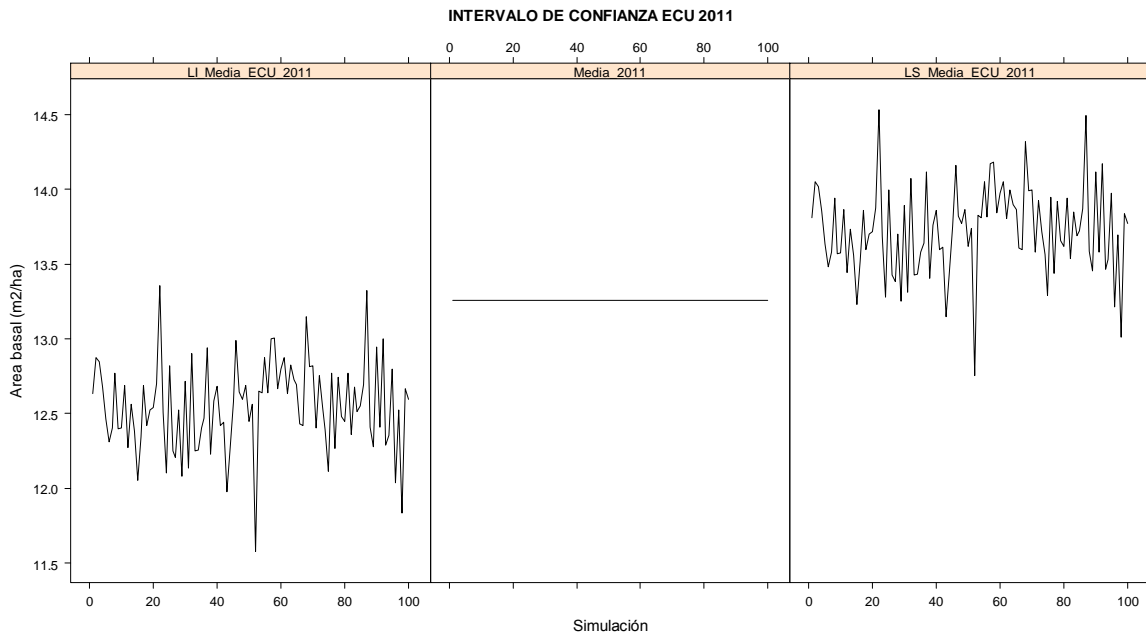


Figura 4.7.7. Intervalos de confianza ECU 2011 de la condición dos. El 92% de los intervalos de confianza para ABHA atrapan al parámetro.

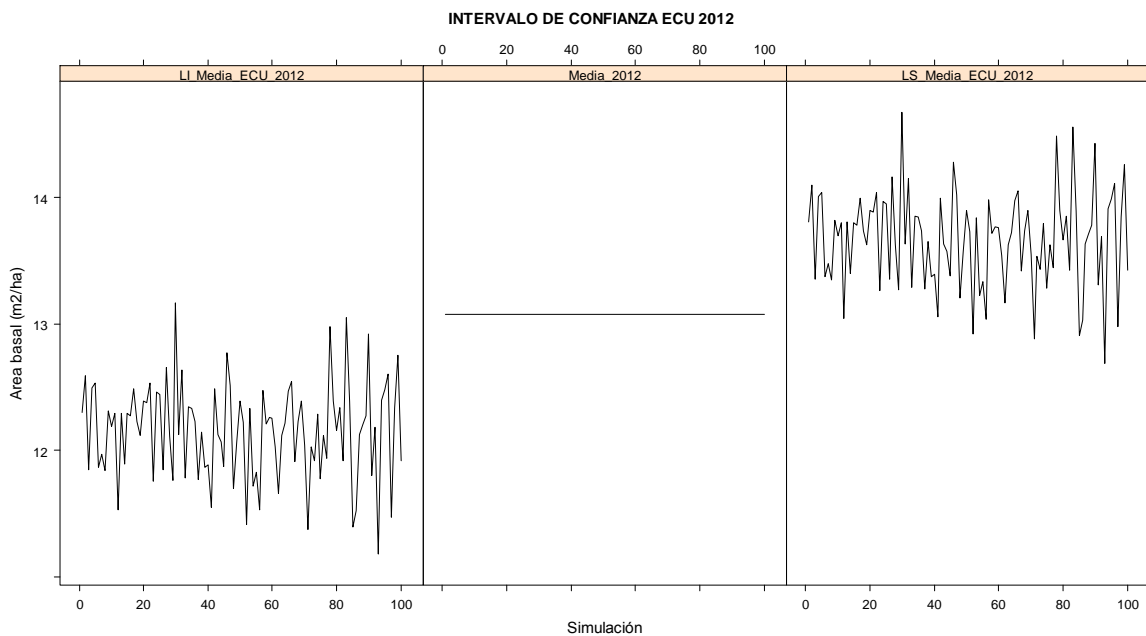


Figura 4.7.8. Intervalos de confianza ECU 2012 de la condición dos. El 90% de los intervalos de confianza para ABHA atrapan al parámetro.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES GENERALES

La gran cantidad de información recabada en campo por el Inventario Nacional Forestal y Suelos de México permite estimar parámetros de tipo agregado y no agregado. Con respecto a los primeros sobresalen el carbono, la biomasa, el volumen maderable, el área basal y la densidad arbórea por unidad de superficie, y en los segundos, destacan los índices de biodiversidad y de degradación forestal. En la presente investigación se abordaron, tanto desde el punto de vista teórico como práctico, métodos estadísticos para la estimación de parámetros de tipo agregado.

El diseño de muestreo del INFyS, desde su origen se ha conceptualizado como “muestreo estratificado sistemático por conglomerados en dos etapas”. El término *conglomerados en dos etapas* implica dos marcos de muestreo; así en una primera etapa se seleccionarían conglomerados de una hectárea, y en una segunda, se escogerían las cuatro subparcelas de 400 m² dentro de cada conglomerado, y en consecuencia, existirían dos fuentes de variabilidad. En términos estrictos esta selección en dos etapas no se realiza en la práctica del INFyS. La configuración geométrica de las cuatro subparcelas queda establecida una vez que se define la posición geográfica del punto central del conglomerado. Esta reflexión permite establecer lo siguiente:

Aun cuando un conglomerado comprende dos o más subparcelas disjuntas es tratado como una sola parcela debido a que es atado a un solo punto muestral. En efecto, el conglomerado de subparcelas simplemente sirve para identificar aquellos árboles cuyas zonas de inclusión espacialmente disjuntas incluyen al punto muestral. De hecho, un árbol pertenece a un conglomerado de subparcelas si el punto muestral cae en cualquier lugar de la zona de inclusión de ese árbol.

La aplicación del estimador media de razones (ó muestreo replicado) supone que las subzonas de inclusión de los árboles estén completas. Este supuesto permite la cuantificación directa del área del conglomerado efectivamente muestreada en campo. Aunque en campo el mapeo de las subparcelas es una actividad tediosa, la cuantificación del área efectivamente muestreada en campo permitirá conocer de manera exacta el área de inclusión de los árboles en el conglomerado.

Aun cuando teóricamente es posible e interpretable trazar las zonas de inclusión de los árboles cuando las subparcelas son circulares o rectangulares, la delimitación de las parcelas rectangulares en campo es complicada. Mediante el análisis de datos de una población simulada se observó que para las variables densidad y volumen maderable no existen diferencias significativas entre las poblaciones generadas por conglomerados de subparcelas circulares y rectangulares. Este hecho sugiere que en la práctica del INFyS en las selvas de México también pueden utilizarse subparcelas circulares. Varios países, incluyendo Suecia, Alemania y Estados Unidos, levantan sus datos de inventario utilizando parcelas circulares.

El estimador media de razones es tan eficiente como el estimador razón de medias, actualmente aplicado por el INFyS. Mientras que en razón de medias se obtiene un valor por hectárea una vez que en el numerador se han adicionado los valores de la variable de interés y en el denominador se han sumado las áreas efectivamente muestreadas en los conglomerados que integran la muestra, en media de razones los valores por hectárea se obtienen de manera inmediata. Así, el estimador media de razones es preferible cuando sea desea representar la variabilidad de la variable de interés a lo largo del territorio. Media de razones, visto como Muestreo Replicado, también es apropiado cuando forma parte del estimador Mínimos Cuadrados Generalizados.

Al aplicar media de razones para la estimación de parámetros forestales agregados, en selvas en un solo punto en tiempo, se observó que cuando se tienen poblaciones con tamaños de muestras menores a 30 los errores de muestreo ascendieron hasta alrededor del 30%. No obstante, para poblaciones con tamaños de muestra alrededor de 60 observaciones los errores de muestreo fueron inferiores al 15%. Esto sugiere que para garantizar errores de muestreo de esta magnitud en otros ecosistemas forestales, cuando la estimación se hace en un solo punto en el tiempo, un tamaño de muestra con 100 observaciones puede ser apropiado. En poblaciones reducidas será difícil alcanzar muestras de 100 observaciones, por lo que una buena práctica es juntarlas.

En la decisión del tamaño de muestra debe considerarse que el objetivo del INFyS es obtener estimaciones en poblaciones de gran magnitud con errores de muestreo relativamente pequeños. La misma población forestal, por ejemplo bosques del Estado de México, puede ser observada desde la perspectiva del INFyS y desde la perspectiva del Inventario Forestal Estatal. Si en éste último caso la estimación de los parámetros requiere errores de muestreo más pequeños que el INFyS, el tamaño de muestra deberá incrementarse vía anidamiento de redes sistemáticas.

En la investigación se enfatizó el estudio del análisis de datos tomando en cuenta la estructura de correlación temporal, la cual surge cuando el diseño de muestreo a través de tiempo se considera como muestreo de panel rotante o muestreo panelizado. En esta situación, el Estimador Combinado Univariado presentó las mejores propiedades estadísticas en cuanto a sesgo y varianza. El estimador de Media Móvil —utilizado por el Forest Inventory and Analysis Program de USA— no se recomienda cuando el interés se centra en estimar un parámetro agregado en la ocasión de medición más reciente.

El Estimador Combinado Univariado integra al estimador de Mínimos Cuadrados Generalizados y al estimador de Muestreo Replicado, de manera tal que para la aplicación del estimador ECU se requiere de la matriz $\hat{\Sigma}$ – varianzas-covarianzas de las estimaciones a nivel de subpaneles– y de la matriz D –ceros y unos que indican la ocasión de medición–. La complejidad relativa de estas matrices queda justificada por la ganancia de precisión en la estimación simultánea de las medias correspondientes a las diferentes ocasiones de medición.

Al considerar que el INFyS es un proyecto a largo plazo implementado por el Gobierno Mexicano y dada la estructura de muestreo de panel rotante con que los datos se registran en los distintos ecosistemas forestales de México, el estimador de ECU se puede aplicar conforme los datos se van tomando a lo largo del tiempo, no es necesario esperar hasta que un ciclo se complete. La información disponible hoy día con casi dos ciclos de medición, permitirá evaluar la tendencia de los parámetros del bosque en las diferentes poblaciones forestales del país. La precisión del estimador MCG, y en consecuencia de ECU, se reflejará en plenitud cuando exista alta correlación entre las parcelas remedidas, de ahí la importancia del control de calidad en la toma de información por parte de la Gerencia de Geomática e Inventario de la CONAFOR.