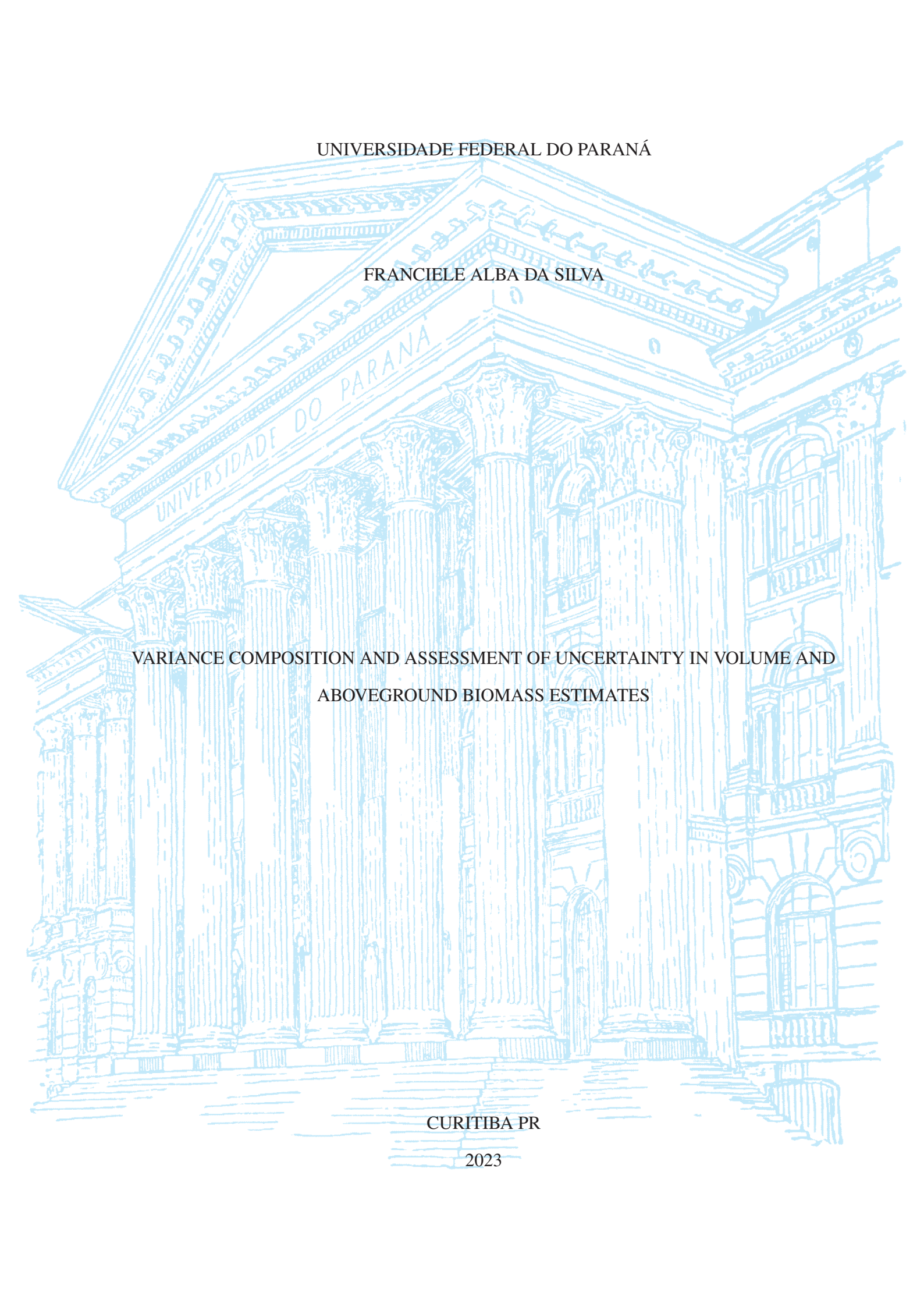




UNIVERSIDADE FEDERAL DO PARANÁ

FRANCIELE ALBA DA SILVA

VARIANCE COMPOSITION AND ASSESSMENT OF UNCERTAINTY IN VOLUME AND
ABOVEGROUND BIOMASS ESTIMATES

CURITIBA PR

2023

FRANCIELE ALBA DA SILVA

VARIANCE COMPOSITION AND ASSESSMENT OF UNCERTAINTY IN VOLUME AND
ABOVEGROUND BIOMASS ESTIMATES

Tese apresentada como requisito parcial à obtenção do grau de Doutora em Engenharia Florestal no Programa de Pós-Graduação em Engenharia Florestal, Setor de Ciências Agrárias, da Universidade Federal do Paraná.

Área de concentração: *Engenharia Florestal*.

Orientador: Sylvio Péllico Netto.

Coorientadores: Alexandre Behling, Rubén Manso.

CURITIBA PR

2023

Ficha catalográfica elaborada pela
Biblioteca de Ciências Florestais e da Madeira - UFPR

Silva, Franciele Alba da

Variance composition and assessment of uncertainty in volume and aboveground biomass estimates / Franciele Alba da Silva. - Curitiba, 2023.

1 recurso on-line : PDF.

Orientador: Prof. Dr. Sylvio Péllico Netto

Coorientadores: Prof. Dr. Alexandre Behling

Dr. Rubén Manso

Tese (Doutorado) - Universidade Federal do Paraná, Setor de Ciências Agrárias, Programa de Pós-Graduação em Engenharia Florestal. Defesa: Curitiba, 20/04/2023.

1. Análise de variância. 2. Incerteza de medição (Estatística). 3. Florestas - Medição. 4. Biomassa florestal - Medição. 5. Levantamentos florestais - Métodos estatísticos. 6. Levantamentos florestais - Amostragem. I. Péllico Netto, Sylvio. II. Behling, Alexandre. III. Manso, Rubén. IV. Universidade Federal do Paraná, Setor de Ciências Agrárias. V. Título.

CDD – 634.9285

CDU – 634.0.524.6

Bibliotecária: Berenice Rodrigues Ferreira – CRB 9/1160

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação ENGENHARIA FLORESTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **FRANCIELE ALBA DA SILVA** intitulada: **Variance Composition and Assessment of Uncertainty in Volume and Aboveground Biomass Estimates**, sob orientação do Prof. Dr. SYLVIO PÉLLICO NETTO, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutora está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 20 de Abril de 2023.

Assinatura Eletrônica

27/04/2023 23:29:26.0

SYLVIO PÉLLICO NETTO

Presidente da Banca Examinadora

Assinatura Eletrônica

24/04/2023 07:43:35.0

RENATA ALCARDE SERMARINI

Avaliador Externo (ESCOLA SUPERIOR DE AGRICULTURA LUIZ DE QUEIROZ)

Assinatura Eletrônica

23/04/2023 21:32:40.0

MAGDA LEA BOLZAN ZANON

Avaliador Externo (UNIVERSIDADE FEDERAL DE SANTA MARIA)

Assinatura Eletrônica

21/04/2023 09:47:08.0

JONATHAN WILLIAM TRAUTENMÜLLER

Avaliador Externo (UNIVERSIDADE FEDERAL RURAL DA AMAZÔNIA)

Assinatura Eletrônica

21/04/2023 20:01:53.0

HASSAN CAMIL DAVID

Avaliador Externo (SERVIÇO FLORESTAL BRASILEIRO)

Dedico esta tese ao meu pai Aldo (in memória), por ter iniciado esse sonho comigo e por ter compreendido minha ausência. A minha mãe Kika e ao meu irmão Álvaro por me apoiarem em todas as escolhas.

BIOGRAFIA DA AUTORA

Franciele Alba da Silva, filha de Aldomir da Silva e Clair Alba da Silva e irmã mais velha de Álvaro Alba da Silva, nasceu no município de Salto do Jacuí, no estado do Rio Grande do Sul, no dia 16 de novembro de 1993. Sua vida estudantil iniciou na primeira série na Escola Municipal de Ensino Fundamental Siegfried Heuser, Salto do Jacuí, no ano de 1999. Seus estudos foram continuados no Instituto Estadual de Educação Miguel Calmon, também pertencente ao mesmo município, onde concluiu o Ensino Médio em 2011.

Em 2012 mudou-se para Frederico Westphalen, Rio Grande do Sul, para ingressar no ensino de nível superior no curso de Engenharia Florestal, na Universidade Federal de Santa Maria, campus Frederico Westphalen. O período de graduação foi de integral dedicação à universidade. No quarto semestre de graduação iniciou os projetos de pesquisa científica na área de estatística e experimentação florestal e seguiu trabalhando como monitora de biometria e em projetos de manejo florestal.

Em 2016 mudou-se para Curitiba para realizar o estágio final de graduação na Embrapa Florestas, no Laboratório de Monitoramento Ambiental, realizando pesquisas na área de Inventário Florestal. No mesmo ano realizou o processo seletivo da Pós-Graduação em Engenharia Florestal, da Universidade Federal do Paraná, e em 2017 ingressou no mestrado nesta Universidade, na linha de pesquisa de Manejo Florestal, sob orientação do professor Dr. Sylvio Péllico Netto e coorientação do professor Dr. Alexandre Behling.

Em 2019 defendeu o mestrado e em seguida ingressou no doutorado, na linha de pesquisa de Manejo Florestal, sob orientação do professor Dr. Sylvio Péllico Netto e coorientação do professor Dr. Alexandre Behling e Dr. Rubén Manso. Em 2021 mudou-se para Edimburgo, Escócia, Reino Unido, para cursar doutorado sanduíche por seis meses. Neste período frequentou a Forest Research UK, onde desenvolveu parte de sua pesquisa de tese sob orientação do Pesquisador Ruben Manso. No último ano de doutorado concluiu sua pesquisa no Brasil, frequentando diariamente o laboratório de manejo de florestas nativas e plantadas, até o presente momento.

AGRADECIMENTOS

À Deus e as forças do universo por me permitirem ter chegado aqui até, por me guiarem pelos bons caminhos, e por sempre me protegerem.

À Universidade Federal do Paraná pela oportunidade de fazer parte da instituição e desenvolver minha pesquisa nesta grande universidade.

Ao Programa de Pós-Graduação em Engenharia Florestal, pela oportunidade de estudo.

À coordenação de Aperfeiçoamento de Pessoal de Nível Superior – (CAPES) pela concessão da bolsa de estudos de doutorado e doutorado sanduíche no exterior.

À Forest Research UK - Northern Research Station por me receberem durante o meu doutorado sanduíche.

À empresa *TANAC S.A* pela oportunidade de coletar e utilizar os dados das florestas de Acácia Negra.

Aos membros da banca examinadora, pela disponibilidade e contribuição para o aprimoramento deste trabalho: Dra. Magda Zanon, Dra. Renata Alcarde Sermarini, Dr. Hassan David, Dr. Jonathan William Trautenmuller, Dr. Allan Pelissari.

À minha família, em especial minha mãe Kika e meu irmão Álvaro, por serem minha base e meu porto seguro, por me apoiarem em todas as minhas decisões, por todo o amor e compreensão.

À minha segunda família de coração, minha Dinda Su e dindo Osiel e meus irmãos Maria Eduarda, Pedro e Luis, por todo amor e carinho.

Aos meus colegas e amigos de jornada do laboratório de manejo de florestas nativas, pela amizade e companheirismo de todos os dias: Antônio, Claiton, Jimmy, Kauana, Adriane, Lucas, e um agradecimento em especial ao Gabriel, João e Luan, por sempre estarem dispostos a me ajudar com dúvidas estatísticas.

Aos amigos que fiz durante esse período na pós-graduação, que mesmo à distância sempre estarão presentes em minha vida: Mirella, Maycon, Luciano, Jorge, Thiago, Lisandra, Cícero, Rudson, Cristine.

Aos amigos e colegas da pós-graduação dos outros laboratórios do ciploma, obrigada pela convivência diária e troca de experiências profissionais e pessoais: Kyvia, Felipe, Augusto, Darcy, Iaci, Theonize, Daiane, Thiago, Cibelle.

À minha rede de apoio feminina, formada por mulheres maravilhosas da pós-graduação: Kauana, Iaci Dandara, Darcy Laura, Adriane, Kyvia, Tarcila e Manu. Obrigada por me lembrarem diariamente da nossa força e coragem!

À Kauana Engel, minha amiga de longa data que ocupa muitos lugares importantes na minha vida. Obrigada por ser família e sempre estar ao meu lado.

À Fernanda Valente, minha amiga que tive a sorte grande em conhecer no doutorado sanduíche. Obrigada por ter compartilhado comigo não só a nossa casa em Edimburgo, nº 51 na Watson Crescent, mas muitos momentos de muita alegria e tristeza. Obrigada por ter sido luz em dias tão escuros.

À minha professora de inglês e amiga Andrea Carvalho, obrigada por todos os ensinamentos da língua inglesa que me permitiram realizar o sonho do doutorado sanduíche. Agradeço também pelas muitas reflexões sobre muitas questões que me fizeram evoluir muito nesses anos de doutorado.

Aos meus afilhados que com a inocência de crianças me mostram a verdadeira importância das coisas mais simples da vida, me ensinam e enchem meu coração de amor e esperança: Guilherme, Mateus, Miguel, Isabela, Luis Otávio, Cecília, Linda, Aurora e Anthonya. Aos amigos de longa data Franciel e Pâmela, por estarem presentes em todos os momentos.

Aos meus primos, por todo o suporte, carinho e amor.

À todas as pessoas que de alguma forma colaboraram e torceram para a realização desse trabalho, que entenderam a minha ausência e colocaram meu nome em suas orações. Muito obrigada!

AGRADECIMENTO AOS ORIENTADORES

Há quase sete anos eu tinha o sonho de entrar na pós-graduação, cursar o mestrado e o doutorado. Então, em uma tarde do mês de setembro de 2016, às vésperas da minha primeira candidatura ao processo seletivo para ingressar em um programa de pós-graduação, eu assisti uma palestra na semana acadêmica do curso de Engenharia Florestal da UFPR. O palestrante do evento era um professor muito renomado da Engenharia Florestal, uma pessoa extremamente brilhante, de uma inteligência e profissionalismo que admiro muito: Professor Sylvio Péllico Netto. Esse professor discursava sobre o inventário do Florestal do Parque Nacional do Iguaçu e contou sobre uma conversa que teve com o General Pinto da Luz, então Presidente do IBDF, para aprovar o relatório final deste inventário. E, assim, ele contou que o General, após ouvir seus argumentos e se sentir convencido das argumentações apresentadas, decidiu aprová-lo e falou a seguinte frase: “Lute pelo que você acredita, morra pelo que você acredita.” Esse foi o primeiro ensinamento que ele me deixou, sem ao menos me conhecer e sendo a primeira vez que eu o via. E assim eu soube que gostaria de ter uma oportunidade de trabalhar com ele. Ele me orientou antes mesmo de ser meu professor orientador. E assim trabalhamos seis anos juntos, lutando pelo que acreditamos, que é a ciência florestal.

Alguns dias passados no mesmo mês tive a oportunidade de conhecer um outro professor prodígio, recém concursado em duas universidades federais, tinha defendido seu doutorado havia poucos dias e estava planejando se mudar para Alagoas, o professor Alexandre Behling. Uma pessoa de uma humildade sem tamanho, que está sempre disposto a ajudar. Tomamos um café um dia antes da minha prova do processo seletivo para o mestrado e durante o café ele me disse “você precisa entrar na pós-graduação e estudar aqui na UFPR”, e assim eu fiquei ainda mais

empolgada para entrar. Algum tempo depois, quando eu já era aluna da UFPR tive a felicidade de encontrá-lo como professor na mesma instituição. Assim, tive a oportunidade de tê-lo como coorientador no mestrado. Um dia antes de uma apresentação eu estava muito nervosa e ansiosa e o professor Alexandre me fez a seguinte pergunta: “O que é o mais importante na vida?” e após me fazer refletir ele respondeu: “são as pessoas. Você pode fazer sua apresentação ficar menos nervosa e preocupada porque as pessoas sempre estarão lá por você.” Assim me tranquilizei e fiz minha apresentação, pois aprendi que onde temos pessoas boas sempre teremos segurança, seja no ambiente profissional, seja na vida.

Dois anos depois, no meu primeiro ano de doutorado, em uma das coincidências do universo, tivemos a imensa sorte de receber um e-mail de um pesquisador muito simpático e referência na área florestal de análise de incertezas em estimativas florestais, dizendo que estava vindo para o Brasil e gostaria de conversar sobre o assunto: o Dr Rubén Manso, que já de imediato me explicou muitas coisas sobre o assunto e se mostrou disposto a me receber na Forest Research-UK. Então, por conta de todo o apoio e incentivo desse pesquisador comecei a minha caminhada em busca do sonho de um doutorado sanduíche. Durante seis meses neste programa tive a oportunidade de conviver e trabalhar com o Rubén, uma das pessoas mais geniais e incríveis que tive a sorte de conhecer. Ele me mostrou que onde temos um amigo, temos casa, mesmo do outro lado do oceano. E se não fossem pelas conversas e histórias compartilhadas dos diferentes países, as longas caminhadas eu não teria aguentado meus dias mais difíceis longe de minha família. Ele me fez perceber o quanto somos capazes de “inventar a dança” diante dos problemas profissionais, buscando soluções.

O Professor Sylvio, o professor Alexandre e o Rubén me ensinaram muitas coisas profissionalmente, me guiaram até a finalização deste trabalho de tese e continuam me auxiliando no caminho da pesquisa florestal. Mas me deram algo mais, que vai além do aprimoramento profissional, são ensinamentos para a vida toda e que sempre me acompanharão. Eles foram muito mais que professores orientadores, eles foram meus grandes mestres e amigos. Eu deixo aqui toda minha gratidão por tudo que me propiciaram, com muita admiração e apreço.

“Somos assim. Sonhamos o voo, mas tememos as alturas. Para voar é preciso amar o vazio. Porque o voo só acontece se houver o vazio. O vazio é o espaço da liberdade, a ausência de certezas. Os homens querem voar, mas temem o vazio. Não podem viver sem certezas. Por isso trocam o voo por gaiolas. As gaiolas são o lugar onde as certezas moram.”

(RUBEM ALVES)

RESUMO

O inventário é a principal ferramenta para a quantificação florestal, pois permite avaliar o estado atual e o desenvolvimento das florestas. Os inventários florestais fornecem estimativas pontuais através de dois tipos diferentes de atributos: medições diretas de variáveis de fácil obtenção (por exemplo, diâmetro a altura do peito - dbh) e variáveis preditas (por exemplo, volume ou biomassa). As últimas são modeladas através de regressão que utilizam as variáveis de fácil medição como preditores, medir o volume e a biomassa de outra forma é um processo com alto custo e com grande demanda de tempo. O volume e a biomassa são altamente relevantes para fins de gerenciamento e tomada de decisões e para este objetivo, após a predição de árvores individuais, as estimativas em nível populacional são geralmente obtidas através de estimadores estatísticos. As estimativas em nível populacional incorrem em uma incerteza que está principalmente associada à amostragem e ao processo de modelagem, devido apenas parte da população ser medida de forma realista. Uma vez que a amostragem e os parâmetros do modelo têm erros desconhecidos, o que se pode quantificar é a incerteza em torno das estimativas. Entretanto, a confiabilidade nos resultados do inventário florestal é freqüentemente relatada sem incluir a incerteza relacionada ao modelo. Isto naturalmente leva a uma subestimativa da incerteza total das estimativas, com conseqüências potenciais para a tomada de decisões, pois introduz uma percepção enganosa da precisão dos resultados. A quantificação e compreensão da incerteza é, portanto, crucial para melhorar a confiabilidade nas estimativas reportadas no inventário florestal. Com isso, as principais perguntas científicas respondidas na pesquisa foram: Qual é a quantidade e a contribuição de incerteza decorrente do modelo e da amostragem em estimativas de volume e biomassa? Como quantificar a incerteza total, considerando estas fontes de incerteza de forma conjunta? Os dados utilizados neste estudo provêm dos povoados de *Acacia mearnsii* localizados no sul do Brasil. Os dados foram coletados em dois níveis: parcelas (nível de amostragem) e de parcelas com amostras destrutivas (ajuste dos modelos). Para estimar o volume e a biomassa médios da população, os modelos foram aplicados às árvores individuais, em que as predições foram agrupadas dentro das parcelas de inventário. Estes valores foram, então, inseridos no estimador apropriado da média por hectare. A incerteza devido à amostragem e ao modelo foi quantificada utilizando estimadores híbridos que consideram conjuntamente as principais fontes de incerteza nos inventários florestais. Os estimadores híbridos para analisar a incerteza foram o bootstrap e uma abordagem analítica. De acordo com os resultados obtidos, conclui-se que a avaliação da incerteza é essencial para a confiabilidade dos resultados que serão implementados nas estratégias de manejo e deve ser incluída como parte de informações de confiabilidade nas estimativas do inventário florestal.

Palavras-chave: incerteza; modelagem; amostragem; estimativas florestais; volume; biomassa acima do solo; estimativas híbridas de variância; propagação de erros.

ABSTRACT

Forest inventory is the main tool for forest quantification as it allows assessing the current state and development of forests at those scales. Forest inventories provide punctual estimates through two different types of attributes: direct measurements of easily obtainable variables (e.g. dbh) and predicted variables (e.g. volume or biomass). The latest are modelled through regression models that use the easily measured variables as predictors, which would otherwise be expensive and time-consuming to measure. Volume and biomass are highly relevant for management and decision-making purposes and towards this aim population-level estimates are usually obtained through statistical estimators. Population-level estimates incur an uncertainty that is mostly associated to the sampling and the modelling process, since only part of the population can realistically be measured. Since sampling and model parameters have unknown errors, all we can quantify is uncertainty around the estimates. However, the reliability in forest inventory results is often reported without including model-related uncertainty. This naturally leads to an underestimation of the total uncertainty of the estimates, with potential consequences for decision making, as it introduces a misleading perception of accuracy. Quantifying and understanding uncertainty is therefore crucial to improve the reliability in the estimates reported in forest inventory. Therefore, the main scientific questions I sought to answer were: *What is the amount and share of uncertainty arising from the model and sampling in volume and biomass estimates? How to quantify the total uncertainty, considering these sources of uncertainty jointly?* The data used in this study come from *Acacia mearnsii* stands located in southern Brazil. Data were collected at two levels: plot level (inventory purposes) and subplot level (model fitting purposes). To estimate the population-level mean volume and biomass, the models were applied to individual tree measurements, whose predictions are grouped within inventory plots. These values are then entered in the appropriate estimator of the mean per hectare. I quantified total uncertainty due to sampling and model, using hybrid estimators that jointly consider the main sources of uncertainty in forest inventories. I applied two hybrid estimators to analyse uncertainty: bootstrap and an analytical approach. In view of the results, I concluded that uncertainty assessment is essential for the reliability of the results that will be implemented in management strategies and should be included as part of the estimates of any forest inventory.

Keywords: uncertainty; modelling; sampling; forest estimates; volume; aboveground biomass; variance hybrid estimators; error propagation.

LIST OF FIGURES

1.1	SUMMARY OF THESIS STRUCTURE.	24
2.1	FLUX DIAGRAM ERROR PROPAGATION IN VOLUME AND BIOMASS ESTIMATES IN FOREST INVENTORIES.. . . .	36
3.1	BOX-PLOT OF THE VOLUME AND BIOMASS DATASETS ACCORDING TO SAMPLE SIZE.. . . .	62
3.2	RESIDUALS RESULTS FOR VOLUME EQUATIONS BY SAMPLE SIZE GROUPS.. . . .	64
3.3	RESIDUALS RESULTS FOR BIOMASS EQUATIONS BY SAMPLE SIZE GROUPS.. . . .	66
3.4	ABSOLUTE AND RELATIVE CONTRIBUTION OF SOURCES OF ERROR FOR THE TOTAL UNCERTAINTY IN VOLUME ESTIMATES.	68
3.5	ABSOLUTE AND RELATIVE CONTRIBUTION OF SOURCES OF ERROR FOR THE TOTAL UNCERTAINTY IN BIOMASS ESTIMATES.. . . .	69
3.6	CONFIDENCE INTERVALS FOR ALL a) VOLUME AND b) BIOMASS MODELS.. . . .	70
3.7	CONTRIBUTION OF FIRST AND SECOND DERIVATIVES FOR THE UN- CERTAINTY DUE TO THE MODEL.	71
3.8	UNCERTAINTY PER TREE IN VOLUME ESTIMATES.	71
3.9	UNCERTAINTY PER TREE IN BIOMASS ESTIMATES.. . . .	72
3.10	RESIDUALS OF a)VOLUME AND b)BIOMASS EQUATIONS.	73
3.11	UNCERTAINTY UNDERESTIMATED ACCORDING TO GOODNESS OF FIT TO VOLUME AND BIOMASS	74
3.12	UNDERESTIMATED VARIANCE WHEN IGNORING THE MODEL UNCER- TAINTY, IN DIFFERENT VARIANCE CLASSES OF PLOTS FOR VOLUME AND BIOMASS.	74
3.13	CONTRIBUTION OF MODEL AND SAMPLING TO THE TOTAL UNCER- TAINTY IN VOLUME AND BIOMASS ESTIMATES.. . . .	75
3.14	CONFIDENCE INTERVALS.	76

4.1	CONTRIBUTION OF EACH DERIVATIVE FOR THE MODEL-RELATED UNCERTAINTY FOR VOLUME AND BIOMASS ESTIMATES.	95
5.1	ABSOLUTE AND RELATIVE CONTRIBUTION OF THE TOTAL UNCER- TAINTY DUE TO THE MODEL- AND SAMPLING-RELATED UNCER- TAINTY IN VOLUME ESTIMATES FOR EACH SCENARIO.	111
5.2	ABSOLUTE AND RELATIVE CONTRIBUTION OF THE TOTAL UNCER- TAINTY DUE TO THE MODEL- AND SAMPLING-RELATED UNCER- TAINTY IN BIOMASS ESTIMATES FOR EACH SCENARIO.	111
5.3	CONFIDENCE INTERVALS FOR ALL SCENARIOS.	113
5.4	ECONOMICAL DIFFERENT AMONG THE SCENARIOS.	114

LIST OF TABLES

3.1	PLOT DATASET SUMMARY..	57
3.2	SUBPLOT DATASET SUMMARY.	58
3.3	VOLUME EQUATIONS AND GOODNESS-OF-FIT FOR VOLUME BY SAM- PLE SIZE GROUP.	63
3.4	BIOMASS EQUATIONS AND GOODNESS-OF-FIT FOR VOLUME BY SAM- PLE SIZE GROUP.	65
3.5	GOODNESS OF FIT FOR VOLUME AND BIOMASS MODELS, FOR SIMU- LATED DATA.	72
4.1	GOODNESS OF FIT FOR THE VOLUME AND BIOMASS MODELS.. . . .	93
4.2	SAMPLING AND MODEL CONTRIBUTION TO THE TOTAL UNCER- TAINTY WITH THE CORRECTION IN THE ANALYTICAL APPROACH AND BOOTSTRAP ESTIMATOR.	94
5.1	SUMMARY DESCRIPTIONS IN EACH SCENARIO.	108

CONTENTS

1	GENERAL INTRODUCTION	19
1.1	BACKGROUND	19
1.2	REASONS FOR DEVELOPING THE THESIS	20
1.3	SCIENTIFIC QUESTIONS.	21
1.4	OBJECTIVES.	22
1.4.1	Main objective	22
1.4.2	Specific objectives	22
1.5	THESIS STRUCTURE	22
	REFERENCES	25
2	CHAPTER 1: UNCERTAINTY ASSESSMENT IN FOREST INVENTORY THROUGH HYBRID INFERENCE: A REVIEW	28
2.1	INTRODUCTION	28
2.2	DESCRIBING UNCERTAINTY IN THE CONTEXT OF FOREST INVENTORY	30
2.3	DISSECTING THE SOURCES OF UNCERTAINTY AND ITS PROPAGATION IN FOREST INVENTORY	33
2.4	METHODS FOR ANALYSING UNCERTAINTY.	36
2.4.1	Analytical Methods	37
2.4.2	Bootstrap estimator	39
2.5	UNCERTAINTY IN VOLUME AND BIOMASS ESTIMATES	41
2.5.1	Volume	41
2.5.2	Biomass	43
2.6	CONCLUSIONS	45
	REFERENCES	46
3	CHAPTER 2: UNCERTAINTY ASSESSMENT IN VOLUME AND BIO- MASS ESTIMATES: EFFECTS OF SAMPLE SIZE, MODEL QUALITY AND PLOT VARIABILITY ON TOTAL UNCERTAINTY	53
3.1	INTRODUCTION	53
3.2	MATERIAL AND METHODS.	56
3.2.1	Data	56

3.2.2	Volume and aboveground biomass modelling	58
3.2.3	Uncertainty in volume and aboveground biomass estimates	58
3.2.4	Effect of the sample size to fit the model in the total uncertainty	62
3.2.5	Effect of inter-plot variability and model goodness of fit in the total uncertainty .	62
3.3	RESULTS	63
3.3.1	Volume and aboveground biomass models	63
3.3.2	Uncertainty Assessment.	66
3.4	DISCUSSION.	76
3.5	CONCLUSION	80
	REFERENCES	80
4	CHAPTER 3: A CORRECTION FOR ASSESSING UNCERTAINTY IN NON-LINEAR MODELS WITH AN ANALYTICAL APPROACH.	85
4.1	INTRODUCTION	85
4.2	MATERIAL AND METHODS.	87
4.2.1	Statistical developments.	87
4.2.2	Second order Taylor series approximations.	90
4.2.3	Case study.	91
4.3	RESULTS	93
4.4	DISCUSSION.	95
4.5	CONCLUSION	97
	REFERENCES	97
5	CHAPTER 4: HOW DO DIFFERENT SOURCES OF UNCERTAINTIES AFFECT VOLUME AND BIOMASS ESTIMATES?.	101
5.1	INTRODUCTION	101
5.2	MATERIAL AND METHODS.	103
5.2.1	Data and Sampling design	103
5.2.2	Volume and biomass modelling.	104
5.2.3	Height-diameter model	105
5.2.4	Analysis of uncertainty in different scenarios	105
5.2.5	Uncertainty assessment	109
5.2.6	Economic analysis	110

5.3	RESULTS	110
5.4	DISCUSSION.	114
5.5	CONCLUSION	117
	REFERENCES	118
6	FINAL CONSIDERATIONS: ANSWERS TO THE SCIENTIFIC QUESTI- ONS.	123
	REFERENCES	127
	APÊNDICE A – VARIANCE COVARIANCE MATRIX AND PARAME- TERS STANDARD ERROR	129
A.1	VOLUME	129
A.2	ABOVEGROUND BIOMASS	130
	APÊNDICE B – UNCERTAINTY ASSOCIATED IN THE PREDICTIONS	131
B.0.1	Volume	131
B.0.2	Biomass above-ground	131

1 GENERAL INTRODUCTION

1.1 BACKGROUND

Forests are key drivers of the global carbon cycle, biodiversity maintenance, overall planetary climate regulation and therefore human well-being. In addition, forests play a key role in helping to mitigate human-caused climate change, which has been further debated in recent years. Several international agreements (UNFCCC, 1997) and the Convention on Biological Diversity prescribe measures related to forest management and require periodic reporting of information on forest characterization and quantification conditions and changes. For this reason, providing accurate forest information at every stage of obtaining results is extremely important (STÅHL et al., 2016; SAARELA et al., 2020).

In forest characterization, it is critical to assess variables such as volume, as this quantifies how much wood there is in the forest at the preservation and commercialization level. Understanding and predicting biomass is hence key for the assessment of the global carbon cycle (RÖDIG et al., 2019). Under this condition, the quantification of biomass and carbon from living trees has been included in national forest inventory projects worldwide (AFFLECK-DIÉGUEZ-ARANDA, 2016). Wood volume and biomass of trees are highly relevant variables for management and decision-making purposes. Despite the increasing use of remote sensing in forestry (DALLA CORTE et al., 2020), conventional forest inventory is still a major tool to quantify of these variables. In some cases, samples are collected from trees in a destructive way (CHAVE et al., 2014). These samples are used to build statistical models that are applied to link easy-to-measure variables (e.g. tree diameter) to the variable of interest (e.g. volume or biomass), which allows predicting the latter across all sampling plots.

Forest inventories should provide results with high accuracy and precision (PÖRTNER et al., 2022), which means demanding all efforts to remove bias from estimates with representative equations, improve data collection approaches, and include assessments of uncertainty in biomass and carbon inventory estimates (BERNSTEIN et al., 2008; PÖRTNER et al., 2022). Regardless of the accuracy of the results, sampling and modelling involve inherent uncertainties that cannot be eliminated. This is due to the fact that only part of the population is measured and possibly modelled, while the other part is estimated using statistical estimators. Therefore, sampling- and modelling-related uncertainty are the main sources of uncertainty in forest inventory estimates (MCROBERTS–WESTFALL, 2014; KANGAS, 1996; MELO et al., 2018).

These sources of uncertainty along with other factors affecting it need to be quantified and described jointly, which is typically done through hybrid variance estimators (FATTORINI, 2012; CORONA et al., 2014; FORTIN et al., 2018). This is to avoid underestimating the total uncertainty in volume and biomass estimates (MCROBERTS–WESTFALL, 2014; FORTIN et al., 2016) and preventing a loss of information and reliability in political and administrative

decision-making pertaining to forest management and climate change mitigation (KANGAS et al., 2018). Uncertainty assessment has been gaining prominence in recent years in the area of forest modelling (CUNIA, 1987; GERTNER, 1990; KANGAS, 1996; CHAVE et al., 2004), and in international environmental policies, since uncertainty assessment is an important component at the level of good practice in national inventory development, and chapter 3 of the IPCC report (BERNSTEIN et al., 2008) exclusively addresses the issue of uncertainty in greenhouse gas emissions inventory estimates.

The thesis focuses on the assessment of uncertainty associated with volume and biomass estimates for *Acacia mearnsii* forest stands in Brazil. Thus, I addressed the main concepts of uncertainty present in volume and biomass population-level estimates and studied analytical and bootstrap-based hybrid variance estimators to quantify the main sources of uncertainty, including the correction of an existing uncertainty estimator for non-linear models. I also explored how different factors affect uncertainty, such as the number of observations used to fit the models. Furthermore, I showed how error is propagated and discussed how the uncertainty propagation chain in volume and biomass estimates affect the management of forest stands. I also discussed the possible consequences on the impact on the economic return of forestry, which in turn determines the value that forest data will have.

1.2 REASONS FOR DEVELOPING THE THESIS

The main reason for developing this research in Brazil is the fact that the variance of population-level estimates in forest inventories considers just the sampling-related uncertainty and the uncertainty related to the model is systematically ignored. Although methods to jointly evaluate the uncertainty of parameter estimates from the models and sampling are not commonly used in Brazil, they are indispensable for improving inventory results in Brazilian forests, as obtaining high-quality estimates is extremely important for the country, due to its large areas of native and planted forests and for being a world leader in the supply of timber and non-timber forest products, as well as a reference as a provider of environmental services. Therefore, this thesis paves the way for a most needed debate in Brazil on how to assess the uncertainty associated with volume and biomass estimates.

In addition, I believe in the relevance of bringing and demonstrating methodologies already used in other countries to quantify uncertainty in Brazilian forests. The gain in efficiency in using methodologies that account for sampling and model uncertainties together has already been reported in inventories in other countries. In terms of volume estimates, it is possible to mention the National forest inventory - NFI in the USA (MCROBERTS–WESTFALL, 2014, 2015), Austria (BERGER et al., 2014), Finland and Sweden (KANGAS, 1996; SAARELA et al., 2015, 2017). In terms of biomass, some examples include the NFI in the USA (CUNIA, 1987; MCROBERTS–WESTFALL, 2016), Sweden (STÅHL et al., 2016), Panama (CHAVE et al., 2004), or French Guiana (MOLTO et al., 2013). Further, uncertainty assessments in forest

inventory combined with growth models have been undertaken by Fortin et al. (2016) for the NFI in Spain, and by Fortin et al. (2018) and Melo et al. (2018) in Canada. Applying and adapting these methods to the Brazilian forest reality will make a significant contribution.

Another reason to carry out this research was to seek resources and make available forest information with enhanced accuracy in the results of volume and biomass estimates. Thus, we should consider the different sources of uncertainty, because when they are not considered, for example, in the uncertainty related to the estimates of the volume and biomass model or the sampling uncertainty, the evaluations of the final estimates of the variables of interest can be mistaken. Uncertainty assessment fully influences the transparency of results, better decision making that allows a better evaluation of risks in management and investment. The uncertainty assessment is, therefore, an essential component to improve the quality and reveal the accuracy of the estimates reported in the forest inventory.

Furthermore, although the subject has already been addressed in works in other countries, there is still room for improvement in the reporting of uncertainty. Uncertainty is often reported without adequate discussion of the sources or recognition of bias in the estimates, and the calculations themselves are not always completely documented (YANAI et al., 2020). The literature, however, is still scarce for error propagation methodology associating different components, such as variability in the measurement of diameter and height variables, height estimates by biometric models; evaluation of how much each component influences the variance along the error propagation chain. Thus, this research will shed new light on this subject.

1.3 SCIENTIFIC QUESTIONS

Given the background of the thesis and the reasons for conducting it, the scientific questions were designed to be answered throughout the text:

- i) What are the concepts and definitions of uncertainty in volume and biomass inventory estimates?*
- ii) What is the state of the art in uncertainty studies in volume and biomass estimates?*
- iii) How do some factors related to the inventory data, (e.g. number of trees to fit the models, between-plot variance and model precision) affect the uncertainty?*
- iv) Based on these factors, what is the amount and share of uncertainty arising from the model and sampling?*
- v) How can variance hybrid estimators be applied, and adapted to quantify uncertainty more accurately for linear and non-linear models?*
- vi) How do the different sources of uncertainty (measurement errors and height model-related uncertainty) interact in the error propagation chain and influence the uncertainty due to the sampling and model?*
- vii) How does the model for predicting height affect uncertainty in volume and biomass estimates?*

viii) *How do measurement errors in auxiliary variables -dbh and h- affect the uncertainty in volume and biomass estimates?*

ix) *What is the effect of uncertainty on the management of *Acacia mearnsii* stands and what is the influence for the economic issues?*

1.4 OBJECTIVES

1.4.1 Main objective

To assess the uncertainty and the composition of variance sources in volume and biomass estimates in *Acacia mearnsii* stands in Brazil.

1.4.2 Specific objectives

1. To describe the concepts of uncertainty and indicate its different sources in volume and biomass estimates, as well as gather the most important references and the state of the art on the subject.
2. To assess the uncertainty in volume and biomass estimates obtained with linear and non-linear regression models, using the hybrid inference methods - analytical and Bootstrap approach.
3. To analyze the influence of the number of trees for the model fit, the quality of the model and the variability of the plots on the total uncertainty.
4. To correct an existing analytic hybrid estimator for non-linear models by considering the within- and between-plot correlation of the prediction errors.
5. To analyze the influence of predicting the heights used in volume and biomass models plus including measurement errors on the total uncertainty in volume and biomass estimates.
6. To distinguish the contribution of each source of uncertainty and how these interact by the error propagation chain of the total uncertainty of the estimates.
7. To assess the economic implications of not considering uncertainty in volume and biomass estimates.

1.5 THESIS STRUCTURE

To answer the scientific questions and achieve the objectives of this research, our work is divided and organized into four chapters:

Chapter 1: Uncertainty assessment in forest inventory through hybrid inference: a review.

In the first chapter I identified and described the different sources of uncertainty in population-level estimates; presented the main methods of hybrid inference to jointly calculate uncertainty; and provided some relevant case studies found in the literature for the main variables

of interest in forest inventory. I attempted to answer questions 1 and 2 in this chapter. Objective 1 was the focus of this chapter.

Chapter 2: Uncertainty assessment in volume and biomass estimates: Effects of sample size, model quality and plot variability on total uncertainty.

In this chapter I presented analytical approaches and Bootstrap simulations to evaluate uncertainty in volume and biomass estimates, predicted with statistical models presented in linear and non-linear fashion. Furthermore, I study how some factors such as the number of trees to adjust the model, the quality of the model and the variability among plots influence the total uncertainty of the estimates. In this chapter, we answer questions 2 and 3. The purpose of this chapter was to achieve objectives 2 and 3.

Chapter 3: A correction for assessing uncertainty in non-linear models with an analytical approach.

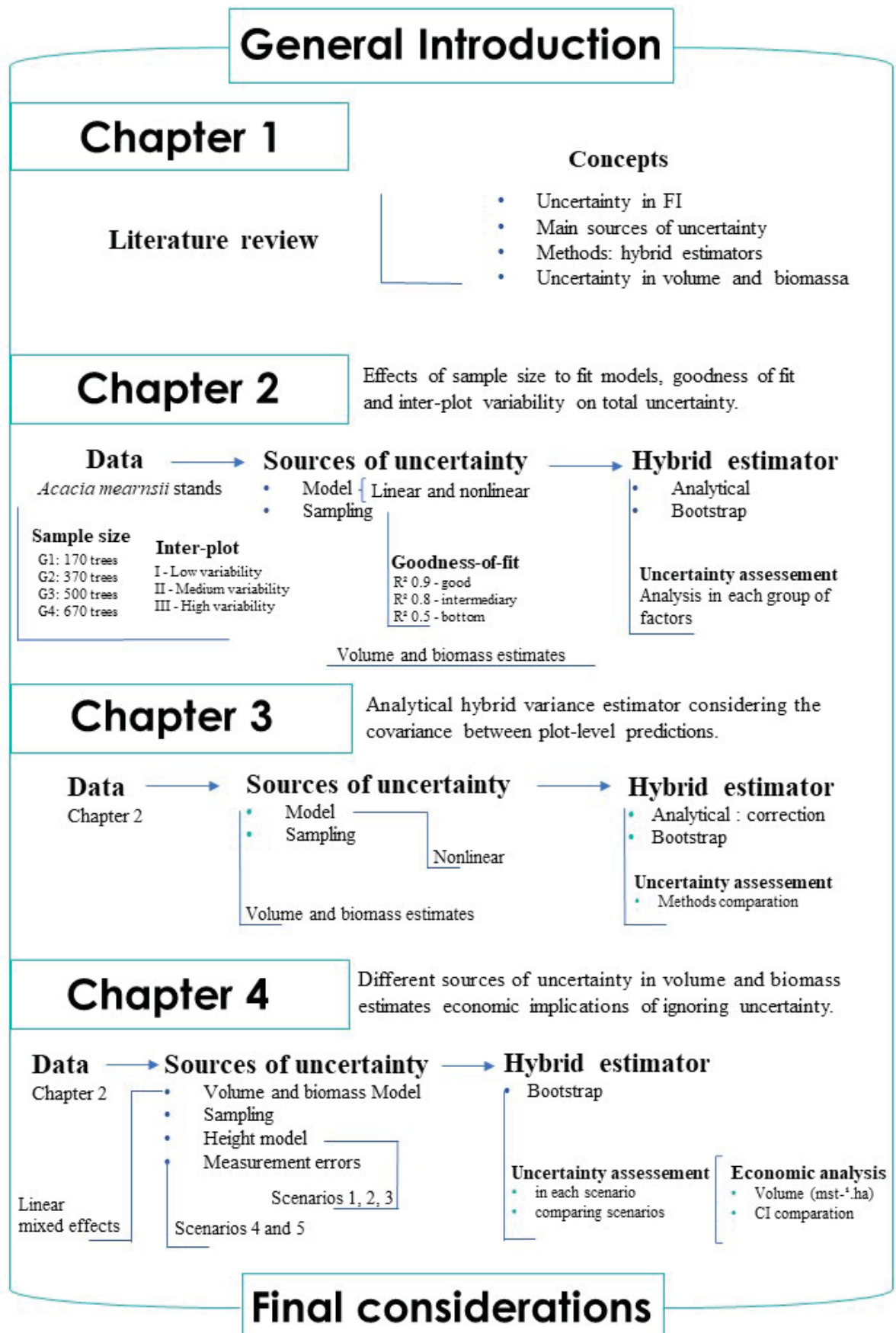
In chapter 3 I seek for more accurate uncertainty estimators. Thus, I corrected an analytical hybrid variance estimator for non-linear models by considering the correlation of prediction errors between trees and between plots. This corrected estimator was validated by comparing the results with a well-established unbiased bootstrap estimator. It addressed scientific question 5 and objective 4.

Chapter 4: How do different sources of uncertainties affect volume and biomass estimates?

In this chapter I quantified different sources of uncertainty in volume and biomass estimates, assessed height model-related uncertainty and measurement errors on volume and biomass estimates. I showed the paths of the error propagation chain and how these error sources interact with each other up to the final volume and biomass estimates obtained with a linear mixed model. In addition, I showed the economic implications of ignoring uncertainty in volume and biomass estimates. The chapter covered questions 6, 7, 8, and 9, and objectives 5, 6, and 7.

In figure 1.1 we present the structure of the thesis in more detail with some of the methodology in each chapter.

FIGURE 1.1 – SUMMARY OF THESIS STRUCTURE.



1.5 REFERENCES

- AFFLECK, D. L. e DIÉGUEZ-ARANDA, U. (2016). Additive nonlinear biomass equations: a likelihood-based approach. *Forest Science*, 62(2):129–140.
- BERGER, A., GSCHWANTNER, T., MCROBERTS, R. E. e SCHADAUER, K. (2014). Effects of measurement errors on individual tree stem volume estimates for the austrian national forest inventory. *Forest Science*, 60(1):14–24.
- BERNSTEIN, L., BOSCH, P., CANZIANI, O., CHEN, Z., CHRIST, R. e RIAHI, K. (2008). Ipcc, 2007: climate change 2007: synthesis report.
- CHAVE, J., CONDIT, R., AGUILAR, S., HERNANDEZ, A., LAO, S. e PEREZ, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1443):409–420.
- CHAVE, J., RÉJOU-MÉCHAIN, M., BÚRQUEZ, A., CHIDUMAYO, E., COLGAN, M. S., DELITTI, W. B., DUQUE, A., EID, T., FEARNSIDE, P. M., GOODMAN, R. C. e OTHERS (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global change biology*, 20(10):3177–3190.
- CORONA, P., FATTORINI, L., FRANCESCHI, S., SCRINZI, G. e TORRESAN, C. (2014). Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. *Canadian Journal of Forest Research*, 44(11):1303–1311.
- CUNIA, T. (1987). Error of forest inventory estimates: its main components.in: Wharton eh, cunia t (eds) estimating tree biomass regressions and their error.
- DALLA CORTE, A. P., SOUZA, D. V., REX, F. E., SANQUETTA, C. R., MOHAN, M., SILVA, C. A., ZAMBRANO, A. M. A., PRATA, G., DE ALMEIDA, D. R. A., TRAUTENMÜLLER, J. W. e OTHERS (2020). Forest inventory with high-density uav-lidar: Machine learning approaches for predicting individual tree attributes. *Computers and Electronics in Agriculture*, 179:105815.
- FATTORINI, L. (2012). Design-based or model-based inference? the role of hybrid approaches in environmental surveys. *Studies in Honor of Claudio Scala*, páginas 173–214.
- FORTIN, M., MANSO, R. e SCHNEIDER, R. (2018). Parametric bootstrap estimators for hybrid inference in forest inventories. *Forestry: An International Journal of Forest Research*, 91(3):354–365.
- FORTIN, M., ROBERT, N. e MANSO, R. (2016). Uncertainty assessment of large-scale forest growth predictions based on a transition-matrix model in catalonia. *Annals of Forest Science*, 73(4):871–883.

- GERTNER, G. Z. (1990). The sensitivity of measurement error in stand volume estimation. *Canadian Journal of Forest Research*, 20(6):800–804.
- KANGAS, A. (1996). On the bias and variance in tree volume predictions due to model and measurement errors. *Scandinavian Journal of Forest Research*, 11(1-4):281–290.
- KANGAS, A., KORHONEN, K. T., PACKALEN, T. e VAUHKONEN, J. (2018). Sources and types of uncertainties in the information on forest-related ecosystem services. *Forest Ecology and Management*, 427:7–16.
- MCROBERTS, R. E. e WESTFALL, J. A. (2014). Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. *Forest Science*, 60(1):34–42.
- MCROBERTS, R. E. e WESTFALL, J. A. (2015). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, 73(3):625–633.
- MCROBERTS, R. E. e WESTFALL, J. A. (2016). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, 73:625–633.
- MELO, L. C., SCHNEIDER, R. e FORTIN, M. (2018). Estimating model-and sampling-related uncertainty in large-area growth predictions. *Ecological Modelling*, 390:62–69.
- MOLTO, Q., ROSSI, V. e BLANC, L. (2013). Error propagation in biomass estimation in tropical forests. *Methods in Ecology and Evolution*, 4(2):175–183.
- PÖRTNER, H.-O., ROBERTS, D. C., ADAMS, H., ADLER, C., ALDUNCE, P., ALI, E., BEGUM, R. A., BETTS, R., KERR, R. B., BIESBROEK, R. e OTHERS (2022). *Climate change 2022: Impacts, adaptation and vulnerability*. IPCC Geneva, Switzerland:.
- RÖDIG, E., KNAPP, N., FISCHER, R., BOHN, F. J., DUBAYAH, R., TANG, H. e HUTH, A. (2019). From small-scale forest structure to amazon-wide carbon estimates. *Nature communications*, 10(1):5088.
- SAARELA, S., ANDERSEN, H.-E., GRAFSTRÖM, A., SCHNELL, S., GOBAKKEN, T., NÆSSET, E., NELSON, R. F., MCROBERTS, R. E., GREGOIRE, T. G. e STÅHL, G. (2017). A new prediction-based variance estimator for two-stage model-assisted surveys of forest resources. *Remote Sensing of Environment*, 192:1–11.
- SAARELA, S., SCHNELL, S., GRAFSTRÖM, A., TUOMINEN, S., NORDKVIST, K., HYYPPÄ, J., KANGAS, A. e STÅHL, G. (2015). Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. *Canadian Journal of Forest Research*, 45(11):1524–1534.

- SAARELA, S., WÄSTLUND, A., HOLMSTRÖM, E., MENSAH, A. A., HOLM, S., NILSSON, M., FRIDMAN, J. e STÅHL, G. (2020). Mapping aboveground biomass and its prediction uncertainty using lidar and field data, accounting for tree-level allometric and lidar model errors. *Forest Ecosystems*, 7(1):1–17.
- STÅHL, G., SAARELA, S., SCHNELL, S., HOLM, S., BREIDENBACH, J., HEALEY, S. P., PATTERSON, P. L., MAGNUSSEN, S., NÆSSET, E., MCROBERTS, R. E. e OTHERS (2016). Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. *Forest Ecosystems*, 3(1):1–11.
- UNFCCC (1997). United nations framework convention on climate change. *Kyoto Protocol*, Kyoto, 19(8):1–21.
- YANAI, R. D., WAYSON, C., LEE, D., ESPEJO, A., CAMPBELL, J. L., GREEN, M. B., ZUKSWERT, J. M., YOFFE, S., AUKEMA, J., LISTER, A. e OTHERS (2020). Improving uncertainty in forest carbon accounting for redd+ mitigation efforts. *Environmental Research Letters*, 15(12):124002.

2 CHAPTER 1: UNCERTAINTY ASSESSMENT IN FOREST INVENTORY THROUGH HYBRID INFERENCE: A REVIEW

ABSTRACT

Forest inventory is the main tool to assess the current state and development of forests. Population-level estimates from forest inventories are random variables whose uncertainty is mostly due to the sampling design and to the modelling of certain variables of interest. Very often, the latter is ignored when uncertainty is reported, which leads to an underestimation of the total uncertainty and loss of accuracy in estimates of important variables such as volume and biomass. To jointly assess those two sources of uncertainty, a special type of estimators, the hybrid estimators, have been developed using several analytical approaches and resampling methods. Although these estimators are now well established, the information available is scattered, highly technical and difficult to access to non-experts. This, and the general lack of awareness on the topic, justifies this review where we aim to bring together the relevant literature on uncertainty in classic forest inventories, focusing on the main variables of interest (volume and biomass). We describe the error propagation chain in the population-level estimates and the most used methodologies to address these problems. Transparent information on the nature of uncertainty involved in classic forest inventory, and its quantification, would increase the trust on scientific studies among researchers and decision makers.

Keywords: modelling; variance; wood quantification.

2.1 INTRODUCTION

Quantifying forest stocks is essential to determine the available timber resources, which are the basis of forest planning at national and regional levels. Measuring all trees in a forest would be unaffordable and, as a result, different modalities of forest inventory are instead conducted to achieve those aims (BONETES, 2003). Forest inventory is the main tool for that quantification as it allows assessing the current state and development of forests at those scales (PÉLLICO NETTO, 1997; VIDAL et al., 2016). Forest inventories are carried out by public and private organizations around the world to collect information, which is subsequently used by forest managers and policy makers to inform decisions and set strategic goals, not only regarding the quantity and quality of wood, but also the stability, ecology and health of forests, and the provision of ecosystem services (PÉLLICO NETTO, 1997; KANGAS et al., 2018).

Historically, small-area forest inventories started with unstructured collection of information of forest stands, and strongly developed with the establishment of a wide array of sampling techniques, focussing at first on timber volume and forest stand structure (MOHREN et al., 2012). Large-area inventories, such as the National Forest Inventories (NFIs), were later developed in most countries with aims beyond simple management, e.g. national strategy (MCROBERTS–WESTFALL, 2016; BREIDENBACH et al., 2014). The normal practice is to conduct these inventories through field surveys, in which detailed and precise assessments of

target variables related to trees, other vegetation, site conditions, and land use are undertaken (MCROBERTS–WESTFALL, 2016). By the end of the 20th century these measurements were combined with observations acquired through remote sensing techniques, which have played a major role in increasing the precision of the estimates (EHLERS et al., 2018). One example is the use of remote sensed data as auxiliary information in NFIs (STÅHL et al., 2014; NELSON et al., 2012; CORONA et al., 2014; SAARELA et al., 2015, 2017, 2020).

Beyond volume and biomass, global interest in forests is increasing due to the multiple services provided by forest ecosystems (SAARELA et al., 2015). Several international agreements such as the United Nations Framework Convention on Climate Change (UNFCCC) (UNFCCC, 1997) and the Convention on Biological Diversity prescribe measures related to the management of forests and require recurrent reporting of information about forest state and change. The need for forest inventories keeps therefore increasing, very especially due to the key role of forests in global carbon cycles and climate change mitigation (STÅHL et al., 2014; SAARELA et al., 2020). The financial incentives linked to forest carbon in these programs have led countries and forest landowners to closely track their forest carbon (VORSTER et al., 2020).

Forest inventories provide estimates through two different types of attributes: direct measurements of easily obtainable variables, either manually (e.g. diameter at breast height (dbh) (JEVŠENAK–SKUDNIK, 2021) or remotely (e.g. tree height); and processed variables, often modelled through regression models that use the abovementioned variables as predictors, which would otherwise be expensive and time-consuming to obtain, e.g. volume or biomass. The latter are highly relevant for management and decision-making purposes (FORTIN–DEBLOIS, 2010; HEROLD et al., 2019; BEHLING et al., 2018), and towards this aim population-level estimates are usually obtained through statistical estimators (PÉLLICO NETTO, 1997; STÅHL et al., 2016; VIDAL et al., 2016). To estimate the population-level parameter of interest e.g. mean volume or biomass, the models are applied to individual tree measurements, whose predictions are grouped within inventory plots. These values are then entered in the appropriate estimator, e.g. estimator of the mean (HENRY et al., 2015; KANGAS, 1996; MCROBERTS–WESTFALL, 2015).

Population-level estimates incur an uncertainty that is mostly associated to the sampling and the modelling of the relevant variables since sampling and model parameters have unknown errors, all we can quantify are uncertainties (JEVŠENAK–SKUDNIK, 2021; BREIDENBACH et al., 2014). This uncertainty about the estimates can be seen as a loss in the quality of the information and could lead to suboptimal or poorly informed decisions if not considered by planners (STÅHL et al., 2016; HAARA et al., 2019). In other words, forest managers may have a misleading perception of accuracy if uncertainty is ignored (KANGAS, 2006; KANGAS et al., 2018; FORTIN et al., 2018). In fact, estimates that come with no robust assessment of uncertainty are simply considered useless for decision making (LAROCQUE et al., 2008). Quantifying and understanding uncertainty is therefore crucial to improve the quality of the estimates reported in forest inventory (KANGAS et al., 2018; STÅHL et al., 2016).

Sampling and model uncertainty are unavoidable but can be estimated through statistical estimators, and reduced as much as possible through efficient sampling strategies and statistical models (PICARD et al., 2015; HENRY et al., 2015). Combining sampling and model uncertainty, which in regression models are due to the parameters and the residual, to calculate the total uncertainty is somewhat challenging, usually resulting in complex mathematical developments (STÅHL et al., 2016; KANGAS et al., 2018). Estimators that consider these sources of uncertainty in their formulation are called hybrid variance estimators. In other words hybrid estimators are based on both the probabilistic design and the model (FORTIN et al., 2018). In this respect, a significant research effort has been devoted over the last 30 years to develop hybrid estimators that jointly consider the main sources of uncertainty in forest inventories (CUNIA, 1987; GERTNER, 1990; CHAVE et al., 2004; FORTIN et al., 2016). Despite this body of literature, the uncertainty of population-level estimates in forest inventories is often reported without including the uncertainty due to the models. This naturally leads to an underestimation of the variance of the estimates (PICARD et al., 2015; FORTIN et al., 2016), with potential consequences for decision making.

As a consequence of the above, there is a need to present the main concepts on uncertainty in forest inventories and the most important methods to address this uncertainty together under a single paper that can serve as a reference, hence this literature review. We aimed to identify and describe the different sources of uncertainty in population-level estimates; to present the main methods of hybrid inference to jointly calculate uncertainty; and to provide some relevant case studies found in the literature for the main variables of interest in forest inventory.

2.2 DESCRIBING UNCERTAINTY IN THE CONTEXT OF FOREST INVENTORY

Uncertainty arises from imperfect or unknown information. It is however important to notice that uncertainty is not a flaw. In fact, some sources of uncertainty are intrinsic to the measurement process and can never be completely removed since no physical quantity can be measured with total certainty (TAYLOR, 1997). This is referred to as irreducible uncertainty or natural variability (KIRCHNER et al., 2021). Uncertainty carries important value in analysis and results, as it influences monitoring projects, management, and policy decisions. In addition, estimates of uncertainty are needed to guide investments in research and monitoring (PARÉ et al., 2016). The consequence of not taking uncertainty into account in forest management decisions can lead to loss of accuracy of results and to suboptimal management (KANGAS et al., 2018).

It was recommended that uncertainty should always be assessed and reported (LAROCQUE et al., 2008; PARÉ et al., 2016). Uncertainty assessments are mandatory for greenhouse gas emissions reports according to the good practice guidelines developed by the Intergovernmental Panel on Climate Change (BERNSTEIN et al., 2008). Chapter 3 of the IPCC report (2006), exclusively addresses the topic of uncertainty in the estimates of these informants, presenting the concepts of uncertainty, quantification techniques, and methods to combine as sources of uncertainty resulting in the estimates. This further underscores that quantifying uncertainty

is essential to establishing the significance of findings and making predictions with known confidence.

In forest inventory, the uncertainty type can in a way be associated with the fact that only part of the population can realistically be measured. In other words, it is the unsampled part of the population that generates the uncertainty, and this is regardless of the precision of the measurements themselves (TAYLOR, 1997). This type of uncertainty has also been termed as a random uncertainty (PÉLLICO NETTO, 1997). This notion of uncertainty is what most authors in the forestry field refer to (CUNIA, 1987; GERTNER, 1990; KANGAS, 1996; MCROBERTS–WESTFALL, 2014; FORTIN et al., 2016).

In this respect, the collection of information in either small-area or large-area forest inventories is entirely based on sampling theory procedures, in which the choice of sample units to be inventoried is probabilistically made (PÉLLICO NETTO, 1997) through different sampling designs (random, systematic, stratified, cluster. . .) (KANGAS, 2006). Design-base inference is then made upon the measured sample to estimate different population parameters (e.g. mean, total, ratio) (STÅHL et al., 2016). These estimates are however random variables that depend on sample selection, i.e. estimators produce different values depending on which elements of the population are included in the sample (STÅHL et al., 2016). As a result, the estimates are subject to a sampling error (PÉLLICO NETTO, 1997).

The sampling designs mentioned above can be applied to different scales such as countries, regions, forest enterprises or even small-scale projects. NFIs for example provide estimates of population parameters for national and regional scales (BREIDENBACH et al., 2014), where sample plots are deployed across the country typically following a systematic scheme. The uncertainty of the estimates of the population parameters from most inventories is usually evaluated using standard design-based inference principles, thus assuming that observed or derived variables are error-free, and therefore only considering the random uncertainty due to sampling (STÅHL et al., 2014).

One frequent example of derived variable is tree volume, which is key in forest management and planning. Accurately measuring tree volume has basically been impossible without destructive sampling (FORTIN et al., 2018) and will probably remain so until modern laser scanning techniques and related software are perfected for a wide range of forest types (CABO et al., 2018). This is also the case for tree biomass (TRAUTENMÜLLER et al., 2021), which provides support for climate and environmental policies (HEROLD et al., 2019). In turn, other variables that are even more difficult to assess, such as carbon sequestration, depend on tree volume or total tree biomass.

These variables of interest need then to be obtained somehow from tree-level measurements and a usual approach is regression models (GERTNER, 1990; STÅHL et al., 2014). This process introduces uncertainty in the form of model-related variability. An accessible auxiliary variable that is strongly correlated with the variable of interest, such as dbh, is measured in all trees within the sample units. The variable of interest is then predicted from the auxiliary variable

through an appropriate model, which is fitted to independent data that consist of measurements of the auxiliary variable(s) and the variables of interest taken on the same trees (CHAVE et al., 2004; FORTIN et al., 2018; TRAUTENMÜLLER et al., 2021). These predicted quantities also come with an error, whose magnitude depends on the stochastic elements of the statistical model (FORTIN et al., 2018; JEVŠENAK–SKUDNIK, 2021).

In modern inventories, especially in NFIs, traditional ground measurements are usually augmented with aerial LiDAR to increase the precision of the estimates (NELSON et al., 2012). Regression models also play an important role in this case: the different metrics that are derived from the resulting point clouds (e.g. percentiles of the distribution of returns' heights) can be used as predictors of the variables of interest at the plot level (YANAI et al., 2020). Similar to the allometric models described above, a proportion of the total uncertainty in the estimation of the population parameters will be due to the LiDAR-related models, and needs to be considered (GOBAKKEN et al., 2012; NELSON et al., 2012).

It is important to note that, although the typical accuracy and precision metrics used to evaluate model predictions are related to the uncertainty due to the model residuals (e.g. root mean square error, coefficient of determination), such as they contribute little to our understanding of the uncertainty due to the model in population-level parameter estimates. In other words, the uncertainty of the model predictions should be evaluated not only in absolute terms, but also in the context of the model application, i.e., the estimate of the population mean (GREGOIRE et al., 2011; MCROBERTS–WESTFALL, 2016).

In summary, the estimates of the population parameters need to be given with an estimate of its variance, which is essential in decision making and risk assessment (KANGAS et al., 2018). This determines the precision of the information, and that can be used to build a confidence interval at a given confidence level. If correctly calculated, a stakeholder can then expect that, for that confidence level (e.g. 95%), the confidence intervals built from different random samples will contain the true value of the population parameter e.g. 95% of the times (CARDOSO et al., 2018; STÅHL et al., 2016). The narrower the confidence intervals, the more reliable the reported point estimates (SAARELA et al., 2017). The most widely used population parameter in practice is the population mean by area (PÉLLICO NETTO, 1997; MCROBERTS–WESTFALL, 2014), although the total (FORTIN et al., 2018) and the ratio (STÅHL et al., 2016) have also been used in some studies. Whereas the population mean can straightforwardly be estimated through well-known estimators (HORVITZ–THOMPSON, 1952), the variance of this estimate requires special attention: as mentioned in the introduction, uncertainty estimates based on standard sampling theory, i.e. design-based inference typically used in forest inventory, do not account for model errors (STÅHL et al., 2014, 2016).

Central to the notion of a composite variance (i.e. sampling and modelling) is the concept of uncertainty propagation. In a system where the outputs of one phase become the inputs of the next phase, the uncertainty incurred at one of these stages is usually carried forward to, and have an impact in the next phases (TEMESGEN et al., 2015). Forest inventories are usually

such systems. The combination of the different sources of error generates an uncertainty model and forms the so-called error propagation chains (PICARD et al., 2015) that will be developed in depth in the next section.

There are other sources of uncertainty in forest inventory, such as the uncertainty arising from systematic or epistemic errors, that have been given some consideration in the literature (GERTNER, 1990; BERGER et al., 2014). Epistemic errors is usually caused by the limited precision and accuracy of the measurement instrument and originate from human action, from negligence in the marking of sample units or in the use of measurement instruments to mistakes in data processing (PÉLLICO NETTO, 1997). These errors are in addition to the uncertainty propagation chain. This concept is however beyond the scope of this paper and will not be developed further.

2.3 DISSECTING THE SOURCES OF UNCERTAINTY AND ITS PROPAGATION IN FOREST INVENTORY

The factors that affect the uncertainty due to sampling are well studied (CUNIA, 1987; CHAVE et al., 2004; PICARD et al., 2015; MCROBERTS–WESTFALL, 2015), and can be listed as:

- The probabilistic method used to select the samples (i.e. the design)
- Plot characteristics: Size, shape and number
- Type of estimator used

The variance between sample units of the variable of interest (e.g. tree volume or biomass), which depends on the landscape-level heterogeneity (e.g. geographical distribution of trees). Choosing an appropriate design is key to be able to select a representative field sample (GRAFSTRÖM–RINGVALL, 2013). When cover information is available at the planning stage of a forest inventory, an efficient sampling strategy is to ensure that the sample is well distributed across the area covered by the auxiliary variables. Well-distributed samples improve not only sampling, but also model fit, if data are collected for this latter purpose (GRAFSTRÖM–RINGVALL, 2013). If the remote sensing information is available before the field sampling is performed, then it opens up the possibility of including this information in the sampling design to select more representative samples.

Regarding plot characteristics, size is a key factor in determining the efficiency of plot design. As plot size increases, within-plot variation increases and between-plot variation decreases, resulting in a lower estimate of variation across all plots (SCOTT, 1998). The question of shape is however more a matter of field efficiency than mapping. When it comes to number of plots, sample size determines the precision as well as the unit cost of the inventory, since this directly affects the representativeness of the data (PICARD et al., 2015; SAARELA et al., 2015). The larger the sample size, the larger the inclusion probability and the smaller the uncertainty due to the design.

The uncertainty due to the model predictions is also well known (CUNIA, 1987; KANGAS, 1996; MCROBERTS–WESTFALL, 2015; PICARD et al., 2015; KIRCHNER et al., 2021), and is attributable to two main factors that correspond to two different phases of the modelling exercise (KIRCHNER et al., 2021):

- Uncertainty in the values of independent variables due to the phase of data collection and data input (such as diameter, height and/or wood density) - Prediction error of the models (analytical phase). This source of error consists of two components (PICARD et al., 2015):

- Uncertainty in the estimates of the models parameters (uncertainty in the expected value).
- Residual variability around the expected value of the models.

Focusing on this analytical phase, Cunia (1987) described in detail the aspects upon which the error of the predictions depends: the sampling design used to collect the data aimed at model fitting, the sample size, and the variability inherent to the variable of interest across trees. However, the magnitude of the variance of the model error will remain approximately the same regardless of the sample size of the sample plots. Like in the case of the uncertainty due to the design, the larger the sample used to fit the model the smaller the uncertainty due to the model (PICARD et al., 2015; SAARELA et al., 2015). This specifically down to the model parameters being estimated more accurately as a result of the sample being more evenly distributed across the population (VAN BREUGEL et al., 2011; SAARELA et al., 2015).

The residual variance is more related to the natural variability of the response and can only be reduced by adding new significant predictors, as also pointed out in McRoberts-Westfall (2014). In other words, uncertainty of a parameter can usually be reduced by collecting more information about that parameter. Natural variability is a characteristic of a parameter that cannot be reduced by collecting more information.

To the above categories related to modelling, Picard et al. (2015) added the error due to the choice of their biomass model. This uncertainty stems from the fact that we cannot entirely be sure that a given model is optimal to make predictions using an independent dataset. When more than one model is available, this uncertainty can be evaluated as the variance of a well-established estimator that combines the outcomes from the different models (Bayesian model averaging, BMA). Picard et al. (2015) found that model choice was the larger share of uncertainty in their study in Central African rainforests. Along these lines, Van Breugel et al. (2011) compared the error due to model choice, the error due to uncertainty related to the model parameters and the sampling error for predicting landscape-level forest biomass in Panama. The model choice was a significant source of uncertainty as a result of the fact that the predictions from the different models greatly varied for the same data points.

The contrary, Saarela et al. (2015) showed that model choice had a minor influence to moderate effect on uncertainty when stock volume increment for Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* (L.) Karst.) and birch (*Betula pubescens* Ehrh.), which are dominant species. The difference with Picard et al. (2015) and Van Breugel et al. (2011) is probably related to the high species diversity of these latter examples, for which different biomass equations

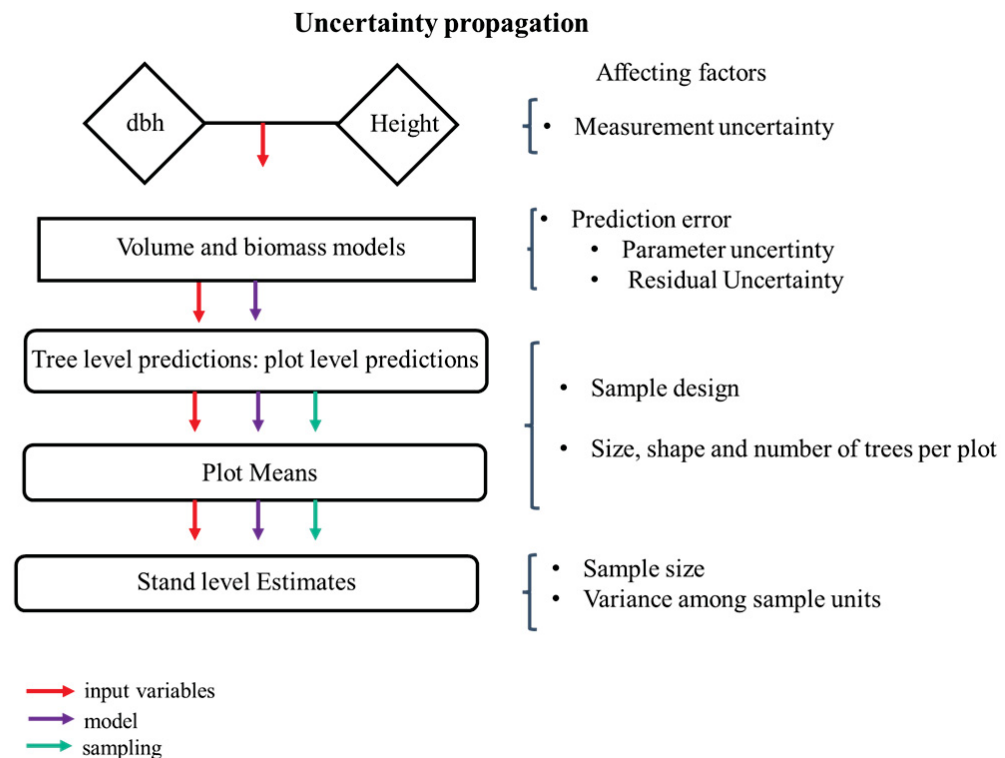
yielded outcomes that differed greatly. In spite of the relevance of this source of uncertainty, most of the times only one model is available, and thus we will not discuss this issue any deeper in the present paper.

A proper identification of the stages where the different sources of uncertainty occur is essential as it would allow diagnosing the impact of each source in the total error (PICARD et al., 2015; KIRCHNER et al., 2021). One way to do this is to track the pathways of error propagation from data collection to the estimate of the variables of interest. Figure 2.1 provides insight into how uncertainty is propagated in population-level estimation procedure of volume and biomass. We use here the example of the estimator of the mean, but a similar flux could be developed for the estimator of the total or any ratio estimator.

Uncertainty begins to be propagated from the very first step, when the explanatory variables are input in the model. This is due, in the first place, to the aforementioned notion that no physical quantity can be measured with total certainty (e.g. all instruments have a precision). Secondly, some variables are not directly measured but predicted. For example, many volume and biomass models require tree height as a predictor (SPURR, 1952; SCHUMACHER–HALL, 1933). Height measurements are however not always available, and it is common practice to predict the heights of the trees in the sample units through height-diameter models (GARBER et al., 2009). In these cases, the use of height may improve the precision of the predictions (SULLIVAN et al., 2018), but it also introduces uncertainty, which propagates through the volume and biomass model and has an impact on the parameter estimates and the residuals (GARBER et al., 2009; FORTIN–DEBLOIS, 2010; KEARSLEY et al., 2013).

Subsequently, this model is used to predict, with an error, the volume or biomass of each tree measured in the sample plots thus propagating the uncertainty from the input variables and the model predictions into the plot-level volume or biomass assessment. When population-level estimates are later computed, the uncertainty of that assessment propagates through the system and adds to the random uncertainty due to the sampling. In this context, it is important to note that the uncertainty due to the residual variance of the models is typically considered as negligible (STÅHL et al., 2016). This being said, Fortin et al. (2018) demonstrated that ignoring this component does induce bias in the uncertainty of the whole system, at least in the case of the estimator of the total.

FIGURE 2.1 – FLUX DIAGRAM ERROR PROPAGATION IN VOLUME AND BIOMASS ESTIMATES IN FOREST INVENTORIES.



SOURCE: The author.

The arrows represent the sense of uncertainty propagation, and the colours refer to the different sources of uncertainty (red, input variables; purple, model; green, sampling).

2.4 METHODS FOR ANALYSING UNCERTAINTY

As mentioned previously, the true variance in forest inventories where models are involved would tend to be underestimated if design-based estimators are used, as they do not take into account the variance related to model predictions (MCROBERTS–WESTFALL, 2014, 2015; FORTIN et al., 2018). In this section, well-established methods are reviewed that do consider the design and the model in their mathematical formulation to quantify their respective contributions to the total uncertainty. We specifically focus on a kind of estimator known as hybrid estimators (FATTORINI, 2012; CORONA et al., 2014; FORTIN et al., 2018). We do not explore here any method to deal with the uncertainty due to epistemic (e.g. human) errors, as these could in principle be minimised. Similarly, we will not touch on the matter of model choice as a source of uncertainty, given that this can only be addressed if several models are available, which is not normally the case.

Hybrid inferential methods are sophisticated and can be used for the aims of statistical inference by considering both the uncertainty of the model parameters and the uncertainty due to the sampling (CONDÉS–MCROBERTS, 2017). The main features of hybrid inference are: (1) a

sample selected under probability principles that contains information (e.g. dbh and height) that will be used as auxiliary in the modelling step, (2) a model that uses this auxiliary information to predict the response of interest in the actual trees in sampling units, (3) an estimator of the population mean, total or ratio that upscale the predictions in the sample units, and (4) both model and sampling inferential methods to estimate the variance of the hybrid estimator of the population mean, total or ratio (STÅHL et al., 2014; MCROBERTS–WESTFALL, 2014; FORTIN et al., 2016). The methodologies presented in this paper can be applied to any type of forest model: simple or complex, site-based or area-based.

In this paper we consider the two most important hybrid estimators presented in the literature, there are two main classes of methods that develop hybrid estimators: analytical approaches and Bootstrap simulation (BERGER et al., 2014). In both cases, it is easier to treat the different sources of uncertainty as independent than to account for the relationships among them (YANAI et al., 2020).

In this case, it is important to note that the term "analytic" refers to the fact that these estimators are treated using algebra, unlike computational methods such as the bootstrap variance estimator (FORTIN et al., 2018). The analytical methods are based on linear algebra and therefore Taylor series are typically used to approximate non-linear functions. In the case of non-linear models, the performance of the estimator depends on how good those approximations are (ARRAS, 1998). Additionally, the model under analysis needs to be differentiable. Although allometric functions usually are, this requirement makes it difficult to use the methodology in more complex situations, such as those where growth or multi-phase models are involved (KANGAS, 1996).

The Bootstrap approach relies on stochastic simulation (LAROCQUE et al., 2008). This method requires assumptions on the joint distribution of errors, and it is computationally demanding (BERGER et al., 2014). The main advantage of this approach is that it can handle overly complex systems, the models do not need to be differentiable and the mathematical developments behind are relatively simple (KWAK–INGALL, 2007; FORTIN et al., 2018). The term "bootstrapping" has also been used to refer to the sampling of theoretical distributions (FORTIN et al., 2018).

2.4.1 Analytical Methods

The perhaps first effort to develop a hybrid estimator in the forestry field is due to Cunia (1987). His analytical method for linear models consists in substituting the model expectations and their errors for the observed values in the design-based estimators. After development, it is possible to come up with a synthetic matrix equation that unbiasedly represents the total variance of the estimator of the mean. In this respect the total variance (using Cunia's terminology), can be expressed as:

$$s_{ww} = [b]' [s_{zz}] [b] + [z]' [s_{bb}] [z] \quad (2.1)$$

Where: the sampling stage denoted by $z_1, z_2, \dots, z_p S_{z_i z_j} i, j = 1, 2, \dots, k$, respectively, and model and the estimates of their variances and covariances are denoted by $b_1, b_2, \dots, b_m S_{b_i b_j} i, j = 1, 2, \dots, m$ respectively $z_1, z_2, \dots, z_m$ are calculated for the data of the first stage of sampling, i.e. the median covariates and variances of the inventory phase plots, using the independent variable for this calculation.

Cunia focused on linear models and we are not aware of any work that generalises Cunia's approach to non-linear models, which is the most common type among the allometric functions used in forest inventory.

Analytical method known as error propagation law (ARRAS, 1998) or delta method (KANGAS, 1996) refers to the procedure where the propagation of the uncertainty due to the parameter estimates is directly treated in algebraic form (note that the sampling error would still be calculated according to the standard methods). Contrarily to the previous approach, this method is suitable for non-linear models as long as they can be approximated to linear functions in the parameter estimates, the residual error and, possibly, the random effects using Taylor series (ABDAR et al., 2021). This is possible when the fixed effects are expressed as a single equation or as a system of simultaneous equations (FORTIN et al., 2018). More specifically, the contribution to the total variance of the uncertainty due to the parameter estimates is calculated through the well-known expression.

$$\text{Var}_{(m)} = Z' \Omega Z \quad (2.2)$$

where Z is the Jacobian matrix of the model with respect to the parameters (i.e. this would be equivalent to the design matrix in a linear context), that results from the first order Taylor expansion; and Ω is the variance-covariance matrix of the parameter estimates.

The contribution of the residual variance of the model is often regarded as negligible as compared with the variance due to the parameters (BREIDENBACH et al., 2014; STÅHL et al., 2016; FORTIN et al., 2016; FU et al., 2017; FORTIN et al., 2018), and especially in relation to the sampling variance (MCROBERTS–WESTFALL, 2016). Caution needs to be observed in the case of complex models, where residual component may be important, at least with estimator of the total (FORTIN et al., 2018).

The uncertainty due to the design is then simply calculated with a standard design-based estimator (HORVITZ–THOMPSON, 1952). The composition of the total variance ($\text{Var}(t)$) using the two sources of uncertainty is assumed to be additive (CUNIA, 1987; GERTNER, 1990) and can therefore be expressed as:

$$\text{Var}_{(t)} = \text{Var}_{(m)} + \text{Var}_{(d)} \quad (2.3)$$

where: $\text{Var}_{(m)}$ is the variance due to the model; $\text{Var}_{(d)}$ is the variance due to the sampling.

Analytical hybrid estimators have mainly been developed with linear and non-linear models for volume or biomass estimation (CUNIA, 1987; GERTNER, 1990; KANGAS, 1996; CHAVE et al., 2004; BERGER et al., 2014; MAGALHÃES et al., 2015). Cunia (1987) worked with a linear equation in biomass estimation, with data from the US national forest inventory. Chave et al. (2004) evaluated the uncertainty in a forest inventory context by propagating the errors of linear model for biomass in tropical forests in central Panama. Magalhães et al. (2015) estimated above and below-ground biomass, and carbon stocks of Mécru forests also using a linear equation and quantified the contribution of those errors to the total variance through an analytical approach. GERTNER (1990) pioneered the use of the error propagation method to assess the total variance with non-linear models in a case study involving *Pinus resinosa*, in the US. In addition, the author extended the application of this method to evaluate measurement errors in stand volume estimates.

Analytical hybrid estimators have recently been extended to heteroscedastic models (SAARELA et al., 2015) and (generalized) mixed-effects models (FORTIN et al., 2016) albeit at the cost of complicated statistical developments. In fact, these analytical derivations or approximations quickly become intractable when the model reaches a higher level of complexity and then other methods are required (FORTIN et al., 2018). An example of 'complex' models would be those composed of several sub-models, such as growth models (FORTIN et al., 2016; MELO et al., 2018), where the predictions from some of them are the predictors of others. Another example of complex models are the simultaneous biomass estimators, in which two or more variables can be adjusted at the same time (BEHLING et al., 2018; TRAUTENMÜLLER et al., 2021; ?).

2.4.2 Bootstrap estimator

The bootstrap approach has successfully been used in forestry as an alternative to analytical methods for uncertainty assessment and error propagation (GERTNER–DZIALOWY, 1984; KANGAS, 1996; BERGER et al., 2014). This method requires that the joint distribution of errors from the model is known (MAHADEVAN–SARKAR, 2009). The method also requires more computer resources than the analytical method (KANGAS, 1996). Random sampling can be based on any distribution, theoretical or empirical, but the normal distribution is a popular option used to obtain the distribution of the means (MAHADEVAN–SARKAR, 2009; RICHARDSON et al., 2010). The results of the repeated simulations are then analysed through the frequency distributions, the means or the standard deviations of the variables of interest (VAN OIJEN, 2017).

The use of bootstrap simulations in the context of hybrid estimators of the mean is as follows. First, a random realisation of the joint probability distribution of the model parameters is extracted under the assumption of normality and considering the estimated variance-covariance matrix of the parameter estimates (MAHADEVAN–SARKAR, 2009; MCROBERTS–WESTFALL, 2014, 2016; FORTIN et al., 2016). These realised parameter values are then

substituted for the estimated parameters in the model to obtain a realised prediction of, for example, volume or biomass (FORTIN et al., 2018). This process is repeated many times (i.e. B simulations), resulting in a probability distribution of the per unit area estimates of volume and biomass (RICHARDSON et al., 2010; VAN OIJEN, 2017).

If the variable of interest is volume or biomass, those realised output variables are specifically used to obtain, for each realisation, an estimate of the mean volume or biomass $\bar{Y}_b$, the mean for all realisations $\bar{Y}_{BS}$ (BS stands for bootstrap in Fortin et al. (2016), and its variance (through a standard design-based estimator). These quantities can be combined as:

$$\widehat{Var} [\bar{Y}] = \frac{(Yb - YBS)}{B} + \frac{\sum \widehat{Var}_d [\bar{Y}_b]}{B} \quad (2.4)$$

which is the bootstrap hybrid estimator of the total variance.

Similar developments would apply in the case of the estimator of the total or in ratio estimators. If the residual error is to be considered, further and laborious developments would be needed. An example for the estimator is given in Fortin et al. (2018).

As bootstrap analysis is based on relatively simple assumptions and does not require the development of complex algorithms, it constitutes a great advantage when applied to complex models, such as growth and yield models (FORTIN et al., 2016; MELO et al., 2018).

The bootstrap method requires the simulation exercise to be run many times, as the statistical validity of the results increases with the sampling intensity (HANESS et al., 1991; GUAN et al., 1997; SMITH-HEATH, 2001). Fortin et al. (2016) ran their model 10.000 times when performing a bootstrap analysis on a growth model to assess the uncertainty of large-scale predictions induced by the residual model error, the errors of the parameter estimates, and the sampling in Catalonia (Spain). The authors reported that number of simulations was sufficient to quantify the proportion of the variance due to the model and due to the design. McRoberts-Westfall (2014) also computed 10,000 simulations in a study of the effect of uncertainty in volume predictions of individual trees. The authors reported that the standard errors already started to stabilise around the 5.000th simulation.

In order to take into account modern methods of data acquisition, such as LiDAR flights, in hybrid inference, some authors have adapted the existing estimators. A pioneering example is Staahl et al. (2014), who developed a model-based (sensu hybrid) inference framework where the predictors of the used models were derived from LiDAR data. Staahl et al. (2014) in fact use existing analytical estimators, but these are then adapted to consider the cluster sampling that arises from dividing the continuous LiDAR information into smaller pieces that suits the model scale. Thus, the authors concluded that this approach should be very useful in inventory programs where regression models are used to predict the quantities of interest at the level of individual sampling units from LiDAR surveys. This approach was tested by Gobakken et al. (2012) and Nelson et al. (2012), who stated that the method has proven to be a viable way to estimate biomass and quantify the reliability of these estimates for estimating biomass in

Hedmark County, then was further applied by Staahl et al. (2012) to quantify the model-related variability of national forest biomass estimates in Sweden and Finland.

Another recent development in hybrid inference is due to Fu et al. (2017), who presented an adapted method that combines Taylor series principles with a bootstrap simulation procedure for quantifying the corresponding uncertainty arising from both sampling and model predictions in regional-scale aboveground biomass estimates. The research results showed that the proposed method generated good, more reliable estimates of both biomass and uncertainty, but also effectively and separately quantifies the uncertainties associated with different sources of error.

2.5 UNCERTAINTY IN VOLUME AND BIOMASS ESTIMATES

Many studies have been conducted to obtain population-level estimates of volume and biomass using hybrid variance estimators in the context of forest inventory (CUNIA, 1987; GERTNER, 1990; CHAVE et al., 2004; BERGER et al., 2014; CORONA et al., 2014). We review in this section some of these case studies and describe what is the impact on volume and biomass estimates, due to the uncertainty from the models and the sampling design.

2.5.1 Volume

Timber is the main forest product, and its availability is therefore essential for the supply chain and a key objective in forest management. The volume of standing timber is, consequently, a crucial attribute to assess the productive potential of forests (CORONA et al., 2014). In this respect, volume needs to be estimated with the greatest possible precision and provided to forest stakeholders together with an evaluation of the uncertainty associated to those estimates. Although there is some progress in volume mensuration from laser point clouds (HACKENBERG et al., 2015), measuring stem volume accurately has almost been impossible so far in forest inventories without felling trees (FORTIN-DEBLOIS, 2010). As a result, assessing volume still depends on modelling, which increases the total uncertainty of the estimates.

As explained before, the uncertainty due to the sampling can be reduced by increasing the sample size. The uncertainty due to the model depends on factors on which the forest planner has less control, e.g. the model may already exist. It is then important to understand what the contribution of this source of uncertainty to the total variance is. The covariance structure of the parameter estimates of the model is often missing in published forestry models, which makes its evaluation unfeasible in some cases (STÅHL et al., 2014). Given the importance of uncertainty in the parameter estimates and the growing interest in uncertainty, future research on biomass and volume modelling should be reported with that information (BREIDENBACH et al., 2014).

One early piece of work where the share of uncertainty due to the different sources is that of Gertner (1990), who analysed the measurement errors, the error linked to the parameter estimates of their non-linear volume model, and the error due to the sampling in the estimation of the mean volume. The author concluded that the variance due to the model parameters

amounted to a 6.76% of the total variance for *Pinus resinosa* in Minnesota, EUA. Kangas (1996) compared Taylor series expansion, bootstrap simulations and recursive modelling methods to assess uncertainty in volume predictions. The data used in the study were from the Finnish National Forest Inventory (NFI) on Scots pine (*Pinus sylvestris* L.) sample units. The percentage of the error due to the model was 3.53% for the total error. Berger et al. (2014) used bootstrap simulations and an analytical method in the context of the Austrian NFI. The authors found that the error due to the model was 5.6% of the total error for all species.

McRoberts-Westfall (2015) observed that the contribution of the variance due to the model was 10% of the total variance when using dbh as the sole predictor in their volume model. This percentage increased up to a 16% when the authors additionally used tree height as a predictor. Either way, the authors stated that this relatively low percentage needs to be met with caution, as it could be higher when using models fitted to small sample sizes and simply poorly fitted with data from the US NFI.

Fortin et al. (2018) tested a corrected variance estimator to estimate the volumes of different log types (pulpwood, sawlog, low-grade sawlog, veneer, low-grade veneer) in individual standing trees of *Acer saccharum* Marsh. and *Betula alleghaniensis* Britton. The contribution of uncertainty due to sampling was more than 80% of the total variance in all grades except low-grade sawlog. This large sampling variation was due to the scarcity of white birch in the sample. The differences between the results found in the mentioned works are probably due to the scale of the studies, the sample size and number of species that the authors considered in each study. This latter factor could be seen as a proxy of the between-plot variability, which is trivially key in the calculation of the variance due to the design.

Gertner (1990) used 46 plots for calibration of just one single species. Fortin et al. (2018) used 393 plots and two different species. McRobert-Westfall (2015) used 2,178 plots from the US NFI, and 38 species. Berger et al. (2014) used data from the Austrian NFI and assumed quality control measures with over 17,000 trees were felled. Irrespective of these differences, the sampling-related uncertainty was always much larger than model-related uncertainty in all cases, even with the smallest sample size, a notion we have tried to convey throughout the paper.

It is observed that the high value of the number of trees for adjustment decreases the percentage of model error. In some situations, such as when the quality of the model is good and there is a high number of samples for adjustment, it is possible to gather from the above literature that the uncertainty due to the volume model is relatively small compared to the sampling-related uncertainty (MCROBERTS–WESTFALL, 2014, 2015). However, the decision of omitting the former in uncertainty assessments should not be taken lightly: as mentioned before, practitioners normally have less control on the models.

On this note, Fortin et al. (2018) found that while the variance due to the sampling decreased with increasing sample size, the volume model-related variance remained stable. Also in absolute terms, Saarela et al. (2015) highlighted that the uncertainty in parameter decreases with an increase of the size of the sample used to fit their volume model while the sampling

uncertainty did not vary. This supports, again, the idea that the model-related uncertainty deserves attention. From the point of view of rigor, ignoring the uncertainty due to model predictions is no longer justifiable given the existing body of literature around this subject (HAARA et al., 2019; MCROBERTS–WESTFALL, 2015). From the point of view of decision making, many sub-optimal or poorly informed decisions could be avoided if uncertainty as whole were properly accurately assessed (HAARA et al., 2019).

2.5.2 Biomass

There is a need to increase our understanding on the dynamic processes occurring in forest ecosystems throughout a production cycle with a view to meet the global targets on decarbonization (PÉLLICO NETTO, 1997). Information about biomass stocks is key in this respect with the overarching objective of implementing appropriate management strategies in forests aimed at energy production (BEHLING et al., 2018). Estimating forest biomass is important for quantifying the role of forests as carbon sources or sinks and for supporting sustainable forest management (MOHREN et al., 2012).

In this context, having access to accurate biomass estimates are central to support harvesting operations (WAYSON et al., 2015). Like in the case of volume, it is important to assess the precision of carbon and biomass estimates. This requires evaluating the uncertainty of each source of error in forest inventories, which can in turn inform ways of increasing that precision (PICARD et al., 2015).

Direct measurements of biomass typically involve destructive methods that are unaffordable for operational purposes. This is the reason why most models are developed to relatively small sampling (TEMESGEN et al., 2015; DONG et al., 2018). Therefore, the variance due to the model could be significant in the assessment of the uncertainty in the population-level biomass estimates (TEMESGEN et al., 2015; MOLTO et al., 2013). On top of that, different tree characteristics that are important in the calculation of biomass and carbon, such as crown architecture and wood density entail a naturally high variability. There is also an appreciable variability among tree components (stem, twigs, leaves, roots and others) and among trees in most populations. Biomass allocation also depends on tree size and age, and the site, among other factors. Furthermore, the relationship between biomass and the usual predictors in biomass equations (i.e. dbh and height) also varies between and within species, and with age (TEMESGEN et al., 2015). All this unavoidably leads to an increase in the uncertainty due to biomass models.

For the same reasons as with volume, in many studies it has been sought to quantify the proportion of the uncertainty in biomass estimates that is due to the model. Cunia (1987) found that percentage to be as high as 34.86% of the total variance in mixed forests of the New York State Forest area. In another study, Djomo et al. (2010) assessed the uncertainty in carbon stock estimates and estimated that the allometric equations contributed with 30% of the total uncertainty in stock estimates over different vegetation types in Cameroon, Africa.

Chave et al. (2004) found an error greater than 20% due to the model in aboveground biomass estimates in tropical rainforests of Central Panama. The authors used then another version of the model that considered wood specific gravity, which reduced that percentage to 13%. They also warned about the error of biomass predictions in large trees, especially if their diameter exceeds the range for which the use of the allometric equation is valid, as they tend to contribute to the variance due to the model heavily and disproportionately. To take this into account, a large sample size and a wider range for the diameters, and significantly reduced the error due to the model, on average by 10%.

Picard et al. (2015) fitted seven biomass models, which they used to assess the variance due to the model choice and the model itself across a range of plot sizes (from 0.04 to 1 ha). The contribution to error due to the model increased for the larger plots. On one-hectare plots, and when all models were a priori equally likely, the error due to model choice was the largest source of error, in forests of the Democratic Republic of Congo. Magalhães et al. (2015) worked with estimates of above and below-ground biomass and carbon stocks of Mecrusse forests in Mozambique. The authors used a linear biomass model whose variance contributed in approximately 10% of the uncertainty in the estimate of total biomass. Molto et al. (2013), on a similar study in French Guiana, concluded that the most important source of the error was the biomass sub-model.

In studies based on countries from the northern hemisphere, we can see differences in the percentage of uncertainty due to the model as compared with the works cited above, which were based on tropical areas. Staahl et al. (2014) used NFI data from Finland and Sweden in a study on biomass to estimate greenhouse gas emissions. The proportion of model error to total mean square error ranged from about 5% (northern Finland) to about 15% (southern Finland). For Sweden, the value is intermediate, i.e., about 10%. The authors further noted that the contribution of model error to the estimates was very small for the stem component, while for the needle and dead branch components the contribution of model error is quite large, in relative terms. The authors further state that the rather small model error contributions in the study are due to the large number of sample trees used in fitting the biomass models in Finland and Sweden; with fewer sample trees, the model error contributions could have been substantial.

The larger errors from the biomass models introduced above relative to those from volume models is due to the fact that biomass varies greatly among species, among biomass components and also with dbh which leads to the conclusion that no biomass model can account for the tree-level biomass variability as precisely as a volume model does with volume variability (VORSTER et al., 2020). In addition to this, most biomass studies have been conducted in native forests (e.g. CUNIA (1987); PICARD et al. (2015); CHAVE et al. (2004)), which raises the question of the natural variability of the forest. Given the previous examples, it seems that ignoring the uncertainty due to the model in population-level biomass estimates can lead to a more severe underestimation of the total uncertainty than in the case of volume, which would in

consequence amplify the detrimental effects listed in the previous section regarding decision making.

The quality of the available data in large area such as NFIs is an important factor for the accuracy of the results in biomass estimates. Biomass data collection in the NFIs has long been established in many countries in the northern hemisphere, e.g. USA, Canada, Finland, Norway. The countries already have a large number of surveys, especially carbon sequestration. Many countries have not yet established repeatable forest inventories, e.g. Brazil, which is a large forest country and is in its first national-level survey. The uncertainty associated with biomass estimates is of central relevance for the subsequent use of inventory results in international climate change policy-making and reporting.

2.6 CONCLUSIONS

This review paper summarises the basic definitions, concepts and methods in the uncertainty associated to population-level volume and biomass estimates in forest inventories. We paid special attention to the uncertainty due to the model, as it is poorly understood and most often neglected. In addition, we presented the pathways of uncertainty propagation, showing that errors propagate in long chains from information collection to the estimates per hectare. Furthermore, we found that uncertainty is often reported without an adequate discussion on its origin, and the calculations themselves are not always fully documented. Information on the location, type and nature of uncertainty fosters transparency and trust in scientific analysis among researchers and decision-makers (YANAI et al., 2020; KIRCHNER et al., 2021). Describing this entire chain helps understanding the effect of the multiple sources of uncertainty on the estimates to identify how to better direct our efforts to improve the accuracy and precision of these estimates.

The existing literature demonstrates the efficiency and potential of the different methods to identify sources of uncertainty. Clearly, there is a need for better understanding and implementation of uncertainty quantification in forest estimates. However, few studies fully explore the relationship between model uncertainty and uncertainty due to the sampling, and most papers seek to investigate the errors of final inventory estimates without further consideration for model uncertainty and also for more rigorous uncertainty analysis. In this context, there is room for substantial improvement in uncertainty assessment and practical applications. One example could be to take into account and quantify the uncertainty linked to height predictions that are typically used in volume and biomass models as perfectly deterministic covariates.

In most surveys, the model uncertainty is smaller than the sampling uncertainty. We have however seen that ignoring this source of error may lead to poorly informed decisions in forest management, with economic implications, particularly in biomass estimates, where the associated variance is even greater than in volume due to the nature of that variable. In

addition, the variance due to the sampling can always be reduced by increasing the sampling effort, whereas the variance due to the model is more difficult to manage.

Related to the above, the benefits of improving the sampling techniques to minimize the uncertainty has been seen in some studies. However, this issue was not part of the main objective of our paper, and this was not the goal, but we would like to highlight some aspects for the reduction of uncertainty: A good tactic to optimize sampling is to have a preliminary view of the population, which would allow sampling to be transparent, robust and simple since with a first monitoring would result in an optimization of the sampling strategy will increase the representativeness of the sample in relation to the population. Another option to reduce the sampling- related variance is to use a stratified design instead of a simple random design and the use of remote sensing can be of help to support stratification, map structural traits of plant canopies and allow inference of traits and, in many cases, species ranges. In addition, RS techniques can increase the number of the samples at a low cost. There are more areas of research yet to contribute to the reduction of uncertainty in forest estimates and future work should make efforts on this issue.

Given those facts, we conclude that the analysis of uncertainty is essential to implement management strategies and policy interventions and should be included as part of the estimates of any forest inventory (YANAI et al., 2020). In full agreement with Kangas et al. (2018), assessing and acknowledging uncertainty in population estimates is the fastest and cheapest way to improve the reliability in decision making.

2.6 REFERENCES

- ABDAR, M., POURPANAH, F., HUSSAIN, S., REZAZADEGAN, D., LIU, L., GHAVAMZADEH, M., FIEGUTH, P., CAO, X., KHOSRAVI, A., ACHARYA, U. R. e OTHERS (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297.
- ARRAS, K. O. (1998). An introduction to error propagation: derivation, meaning and examples of equation $cy = fx \cdot cx \cdot fxt$. Relatório técnico, ETH Zurich.
- BEHLING, A., PÉLLICO NETTO, S., SANQUETTA, C. R., CORTE, A. P. D., AFFLECK, D. L., RODRIGUES, A. L. e BEHLING, M. (2018). Critical analyses when modeling tree biomass to ensure additivity of its components. *Anais da Academia Brasileira de Ciências*, 90:1759–1774.
- BERGER, A., GSCHWANTNER, T., MCROBERTS, R. E. e SCHADAUER, K. (2014). Effects of measurement errors on individual tree stem volume estimates for the austrian national forest inventory. *Forest Science*, 60(1):14–24.
- BERNSTEIN, L., BOSCH, P., CANZIANI, O., CHEN, Z., CHRIST, R. e RIAHI, K. (2008). Ipcc, 2007: climate change 2007: synthesis report.

- BONETES, L. (2003). *Tamanho de parcelas e intensidade amostral para estimar o estoque e índices fitossociológicos em uma floresta ombrófila mista*. Tese de doutorado, Dissertação de Mestrado. Universidade Federal do Paraná. Curitiba.
- BREIDENBACH, J., ANTÓN-FERNÁNDEZ, C., PETERSSON, H., MCROBERTS, R. E. e ASTRUP, R. (2014). Quantifying the model-related variability of biomass stock and change estimates in the norwegian national forest inventory. *Forest Science*, 60(1):25–33.
- CABO, C., ORDÓÑEZ, C., LÓPEZ-SÁNCHEZ, C. A. e ARMESTO, J. (2018). Automatic dendrometry: Tree detection, tree height and diameter estimation using terrestrial laser scanning. *International journal of applied earth observation and geoinformation*, 69:164–174.
- CARDOSO, D., DA SILVA, F., ROSOT, M., GARRASTAZU, M., DE OLIVEIRA, Y., DE MATOS, P. e DE FREITAS, J. (2018). Procedimentos para melhoria da qualidade dos resultados de inventários florestais: aspectos do processamento de dados.
- CHAVE, J., CONDIT, R., AGUILAR, S., HERNANDEZ, A., LAO, S. e PEREZ, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1443):409–420.
- CONDÉS, S. e MCROBERTS, R. E. (2017). Updating national forest inventory estimates of growing stock volume using hybrid inference. *Forest Ecology and Management*, 400:48–57.
- CORONA, P., FATTORINI, L., FRANCESCHI, S., SCRINZI, G. e TORRESAN, C. (2014). Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. *Canadian Journal of Forest Research*, 44(11):1303–1311.
- CUNIA, T. (1987). Error of forest inventory estimates: its main components.in: Wharton eh, cunia t (eds) estimating tree biomass regressions and their error.
- DONG, L., ZHANG, L. e LI, F. (2018). Additive biomass equations based on different dendrometric variables for two dominant species (*larix gmellini* rupr. and *betula platyphylla* suk.) in natural forests in the eastern daxing'an mountains, northeast china. *Forests*, 9(5):261.
- EHLERS, S., SAARELA, S., LINDGREN, N., LINDBERG, E., NYSTRÖM, M., PERSSON, H. J., OLSSON, H. e STÅHL, G. (2018). Assessing error correlations in remote sensing-based estimates of forest attributes for improved composite estimation. *Remote Sensing*, 10(5):667.
- FATTORINI, L. (2012). Design-based or model-based inference? the role of hybrid approaches in environmental surveys. *Studies in Honor of Claudio Scala*, páginas 173–214.
- FORTIN, M. e DEBLOIS, J. (2010). A statistical estimator to propagate height prediction errors into a general volume model. *Canadian journal of forest research*, 40(10):1930–1939.

- FORTIN, M., MANSO, R. e SCHNEIDER, R. (2018). Parametric bootstrap estimators for hybrid inference in forest inventories. *Forestry: An International Journal of Forest Research*, 91(3):354–365.
- FORTIN, M., ROBERT, N. e MANSO, R. (2016). Uncertainty assessment of large-scale forest growth predictions based on a transition-matrix model in catalonia. *Annals of Forest Science*, 73(4):871–883.
- FU, Y., LEI, Y., ZENG, W., HAO, R., ZHANG, G., ZHONG, Q. e XU, M. (2017). Uncertainty assessment in aboveground biomass estimation at the regional scale using a new method considering both sampling error and model error. *Canadian Journal of Forest Research*, 47(8):1095–1103.
- GARBER, S. M., TEMESGEN, H., MONLEON, V. J. e HANN, D. W. (2009). Effects of height imputation strategies on stand volume estimation. *Canadian journal of forest research*, 39(3):681–690.
- GERTNER, G. Z. (1990). The sensitivity of measurement error in stand volume estimation. *Canadian Journal of Forest Research*, 20(6):800–804.
- GERTNER, G. Z. e DZIALOWY, P. J. (1984). Effects of measurement errors on an individual tree-based growth projection system. *Canadian Journal of Forest Research*, 14(3):311–316.
- GOBAKKEN, T., NÆSSET, E., NELSON, R., BOLLANDSÅS, O. M., GREGOIRE, T. G., STÅHL, G., HOLM, S., ØRKA, H. O. e ASTRUP, R. (2012). Estimating biomass in hedmark county, norway using national forest inventory field plots and airborne laser scanning. *Remote Sensing of Environment*, 123:443–456.
- GRAFSTRÖM, A. e RINGVALL, A. H. (2013). Improving forest field inventories by using remote sensing data in novel sampling designs. *Canadian Journal of Forest Research*, 43(11):1015–1022.
- GREGOIRE, T. G., STÅHL, G., NÆSSET, E., GOBAKKEN, T., NELSON, R. e HOLM, S. (2011). Model-assisted estimation of biomass in a lidar sample survey in hedmark county, norway. *Canadian Journal of Forest Research*, 41(1):83–95.
- GUAN, B. T., GERTNER, G. Z. e PARYSOW, P. (1997). A framework for uncertainty assessment of mechanistic forest growth models: a neural network example. *Ecological modelling*, 98(1):47–58.
- HAARA, A., KANGAS, A. e TUOMINEN, S. (2019). Economic losses caused by tree species proportions and site type errors in forest management planning.

- HACKENBERG, J., SPIECKER, H., CALDERS, K., DISNEY, M. e RAUMONEN, P. (2015). Simpletree—an efficient open source tool to build tree models from tls clouds. *Forests*, 6(11):4245–4294.
- HANESS, S. J., ROBERTS, L. A., WARWICK, J. J. e CALE, W. G. (1991). Testing the utility of first order uncertainty analysis. *Ecological modelling*, 58(1-4):1–23.
- HENRY, M., CIFUENTES JARA, M., RÉJOU-MÉCHAIN, M., PIOTTO, D., MICHEL FUENTES, J. M., WAYSON, C., ALICE GUIER, F., CASTAÑEDA LOMBIS, H., CASTELLANOS LÓPEZ, E., CUENCA LARA, R. e OTHERS (2015). Recommendations for the use of tree models to estimate national forest biomass and assess their uncertainty. *Annals of Forest Science*, 72(6):769–777.
- HEROLD, M., CARTER, S., AVITABILE, V., ESPEJO, A. B., JONCKHEERE, I., LUCAS, R., MCROBERTS, R. E., NÆSSET, E., NIGHTINGALE, J., PETERSEN, R. e OTHERS (2019). The role and need for space-based forest biomass-related measurements in environmental management and policy. *Surveys in Geophysics*, 40(4):757–778.
- HORVITZ, D. G. e THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685.
- JEVŠENAK, J. e SKUDNIK, M. (2021). A random forest model for basal area increment predictions from national forest inventory data. *Forest Ecology and Management*, 479:118601.
- KANGAS, A. (1996). On the bias and variance in tree volume predictions due to model and measurement errors. *Scandinavian Journal of Forest Research*, 11(1-4):281–290.
- KANGAS, A. (2006). Design-based sampling and inference. *Forest inventory: methodology and applications*, páginas 13–38.
- KANGAS, A., KORHONEN, K. T., PACKALEN, T. e VAUHKONEN, J. (2018). Sources and types of uncertainties in the information on forest-related ecosystem services. *Forest Ecology and Management*, 427:7–16.
- KEARSLEY, E., DE HAULLEVILLE, T., HUFKENS, K., KIDIMBU, A., TOIRAMBE, B., BAERT, G., HUYGENS, D., KEBEDE, Y., DEFOURNY, P., BOGAERT, J. e OTHERS (2013). Conventional tree height–diameter relationships significantly overestimate aboveground carbon stocks in the central congo basin. *Nature communications*, 4(1):2269.
- KIRCHNER, M., MITTER, H., SCHNEIDER, U. A., SOMMER, M., FALKNER, K. e SCHMID, E. (2021). Uncertainty concepts for integrated modeling-review and application for identifying uncertainties and uncertainty propagation pathways. *Environmental Modelling & Software*, 135:104905.

- KWAK, Y. H. e INGALL, L. (2007). Exploring monte carlo simulation applications for project management. *Risk management*, 9:44–57.
- LAROCQUE, G. R., BHATTI, J. S., BOUTIN, R. e CHERTOV, O. (2008). Uncertainty analysis in carbon cycle models of forest ecosystems: research needs and development of a theoretical framework to estimate error propagation. *Ecological Modelling*, 219(3-4):400–412.
- MAGALHÃES, T. M., SEIFERT, T. e OTHERS (2015). Estimation of tree biomass, carbon stocks, and error propagation in mecrusse woodlands. *Open Journal of Forestry*, 5(04):471.
- MAHADEVAN, S. e SARKAR, S. (2009). Uncertainty analysis methods. *US Department of Energy, Washington, DC, USA*.
- MCROBERTS, R. E. e WESTFALL, J. A. (2014). Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. *Forest Science*, 60(1):34–42.
- MCROBERTS, R. E. e WESTFALL, J. A. (2015). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, 73(3):625–633.
- MCROBERTS, R. E. e WESTFALL, J. A. (2016). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, 73:625–633.
- MELO, L. C., SCHNEIDER, R. e FORTIN, M. (2018). Estimating model-and sampling-related uncertainty in large-area growth predictions. *Ecological Modelling*, 390:62–69.
- MOHREN, G., HASENAUER, H., KÖHL, M. e NABUURS, G. (2012). Forest inventories for carbon change assessments. *Current Opinion in Environmental Sustainability*, 4(6):686–695.
- MOLTO, Q., ROSSI, V. e BLANC, L. (2013). Error propagation in biomass estimation in tropical forests. *Methods in Ecology and Evolution*, 4(2):175–183.
- NELSON, R., GOBAKKEN, T., NÆSSET, E., GREGOIRE, T., STÅHL, G., HOLM, S. e FLEWELLING, J. (2012). Lidar sampling—using an airborne profiler to estimate forest biomass in hedmark county, norway. *Remote Sensing of Environment*, 123:563–578.
- PARÉ, D., GERTNER, G., BERNIER, P. e YANAI, R. (2016). Quantifying uncertainty in forest measurements and models: approaches and applications. *Canadian Journal of Forest Research*, 46(3):v–v.
- PICARD, N., BOYEMBA BOSELA, F. e ROSSI, V. (2015). Reducing the error in biomass estimates strongly depends on model selection. *Annals of forest Science*, 72(6):811–823.
- PÉLLICO NETTO, S.; BRENA, D. (1997). Inventário florestal. 1:316.

- RICHARDSON, A. D., WILLIAMS, M., HOLLINGER, D. Y., MOORE, D. J., DAIL, D. B., DAVIDSON, E. A., SCOTT, N. A., EVANS, R. S., HUGHES, H., LEE, J. T. e OTHERS (2010). Estimating parameters of a forest ecosystem model with measurements of stocks and fluxes as joint constraints. *Oecologia*, 164:25–40.
- SAARELA, S., ANDERSEN, H.-E., GRAFSTRÖM, A., SCHNELL, S., GOBAKKEN, T., NÆSSET, E., NELSON, R. F., MCROBERTS, R. E., GREGOIRE, T. G. e STÅHL, G. (2017). A new prediction-based variance estimator for two-stage model-assisted surveys of forest resources. *Remote Sensing of Environment*, 192:1–11.
- SAARELA, S., SCHNELL, S., GRAFSTRÖM, A., TUOMINEN, S., NORDKVIST, K., HYYPPÄ, J., KANGAS, A. e STÅHL, G. (2015). Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. *Canadian Journal of Forest Research*, 45(11):1524–1534.
- SAARELA, S., WÄSTLUND, A., HOLMSTRÖM, E., MENSAH, A. A., HOLM, S., NILSSON, M., FRIDMAN, J. e STÅHL, G. (2020). Mapping aboveground biomass and its prediction uncertainty using lidar and field data, accounting for tree-level allometric and lidar model errors. *Forest Ecosystems*, 7(1):1–17.
- SCHUMACHER, F. X. e HALL, F. S. (1933). Logarithmic expression of timber-tree volume. *J Agric Res*, 47:719–734.
- SCOTT, C. T. (1998). Sampling methods for estimating change in forest resources. *Ecological Applications*, 8(2):228–233.
- SMITH, J. E. e HEATH, L. S. (2001). Identifying influences on model uncertainty: an application using a forest carbon budget model. *Environmental Management*, 27:253–267.
- SPURR, S. H. (1952). Forest inventory.
- STÅHL, G., HEIKKINEN, J., PETERSSON, H., REPOLA, J. e HOLM, S. (2014). Sample-based estimation of greenhouse gas emissions from forests—a new approach to account for both sampling and model errors. *Forest Science*, 60(1):3–13.
- STÅHL, G., SAARELA, S., SCHNELL, S., HOLM, S., BREIDENBACH, J., HEALEY, S. P., PATTERSON, P. L., MAGNUSSEN, S., NÆSSET, E., MCROBERTS, R. E. e OTHERS (2016). Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. *Forest Ecosystems*, 3(1):1–11.
- SULLIVAN, M. J., LEWIS, S. L., HUBAU, W., QIE, L., BAKER, T. R., BANIN, L. F., CHAVE, J., CUNI-SANCHEZ, A., FELDPAUSCH, T. R., LOPEZ-GONZALEZ, G. e OTHERS (2018). Field methods for sampling tree height for tropical forest biomass estimation. *Methods in Ecology and Evolution*, 9(5):1179–1189.

- TAYLOR, J. (1997). *Introduction to error analysis, the study of uncertainties in physical measurements*.
- TEMESGEN, H., AFFLECK, D., POUDEL, K., GRAY, A. e SESSIONS, J. (2015). A review of the challenges and opportunities in estimating above ground forest biomass using tree-level models. *Scandinavian Journal of Forest Research*, 30(4):326–335.
- TRAUTENMÜLLER, J. W., NETTO, S. P., BALBINOT, R., WATZLAWICK, L. F., DALLA CORTE, A. P., SANQUETTA, C. R. e BEHLING, A. (2021). Regression estimators for aboveground biomass and its constituent parts of trees in native southern brazilian forests. *Ecological Indicators*, 130:108025.
- UNFCCC (1997). United nations framework convention on climate change. *Kyoto Protocol*, Kyoto, 19(8):1–21.
- VAN BREUGEL, M., RANSIJN, J., CRAVEN, D., BONGERS, F. e HALL, J. S. (2011). Estimating carbon stock in secondary forests: decisions and uncertainties associated with allometric biomass models. *Forest ecology and management*, 262(8):1648–1657.
- VAN OIJEN, M. (2017). Bayesian methods for quantifying and reducing uncertainty and error in forest models. *Current Forestry Reports*, 3:269–280.
- VIDAL, C., ALBERDI, I., REDMOND, J., VESTMAN, M., LANZ, A. e SCHADAUER, K. (2016). The role of european national forest inventories for international forestry reporting. *Annals of Forest Science*, 73(4):793–806.
- VORSTER, A. G., EVANGELISTA, P. H., STOVALL, A. E. e EX, S. (2020). Variability and uncertainty in forest biomass estimates from the tree to landscape scale: The role of allometric equations. *Carbon Balance and Management*, 15:1–20.
- WAYSON, C. A., JOHNSON, K. D., COLE, J. A., OLGUÍN, M. I., CARRILLO, O. I. e BIRDSEY, R. A. (2015). Estimating uncertainty of allometric biomass equations with incomplete fit error information using a pseudo-data approach: methods. *Annals of Forest Science*, 72(6):825–834.
- YANAI, R. D., WAYSON, C., LEE, D., ESPEJO, A., CAMPBELL, J. L., GREEN, M. B., ZUKSWERT, J. M., YOFFE, S., AUKEMA, J., LISTER, A. e OTHERS (2020). Improving uncertainty in forest carbon accounting for redd+ mitigation efforts. *Environmental Research Letters*, 15(12):124002.

3 CHAPTER 2: UNCERTAINTY ASSESSMENT IN VOLUME AND BIOMASS ESTIMATES: EFFECTS OF SAMPLE SIZE, MODEL QUALITY AND PLOT VARIABILITY ON TOTAL UNCERTAINTY

ABSTRACT

The importance of assessing precision in forest estimates from forest inventory has been highlighted in recent years, especially with international policies aimed at meeting conservation and climate change demands. To verify forest estimates' reliability, uncertainty analysis is crucial. We aimed (i) to assess the uncertainty in volume and aboveground biomass estimates in stands of *Acacia mearnsii* from southern Brazil using linear and non-linear regression models, and (ii) to evaluate the effect on the total uncertainty (a) of using different sample sizes to fit the models, (b) of varying levels of model goodness-of-fit, and (c) of contrasting within-plot variability at the inventory phase. Data were collected at two levels: plot level (inventory purposes) and subplot level (model fitting purposes). We fitted three models widely used in forest biometrics through the weighted ordinary least squares: Spurr in linear and nonlinear fashion and Schumacher-Hall. We used two methods to analyze the uncertainty: bootstrap estimator and an analytical approach. High-order Taylor series were employed for non-linear models. To test the influence of the number of trees on uncertainty we generated 4 groups increasing the number of trees gradually. Furthermore, we simulated a data set to fit two equations with low quality and separated our plots by variability class. According to our results, smaller sample sizes could lead to slightly more model-related uncertainty in the estimates. This also influences the contribution of first and second order in the volume uncertainty, where the higher order series will mainly depend on the sources that affect the degree of uncertainty in the model parameters. The contribution of sampling uncertainty to the total variance remains approximately the same regardless of the sample size for model fitting. Even though the sampling uncertainty represents the largest portion of the total uncertainty for linear models in all sample size groups and non-linear when the sample size increases and the variance related to the model is low and we also notice a large contribution of sampling related uncertainty when we separate the highest variability plots and the equation presented goodness of fit. When we do not consider model-related uncertainty leads to underestimates of the total uncertainty.

Keywords: Wood quantification; modelling; volume, aboveground biomass; uncertainty assessment; bootstrap estimator; high-order taylor series; model error; error propagation; sample size.

3.1 INTRODUCTION

Reaching international goals related to climate change, forest conservation and management, and sustainable development requires reliable, accurate, and credible monitoring of forest volume and biomass stocks and changes in sequestered carbon (HEROLD et al., 2019). Several international agreements such as the United Nations Framework Convention on Climate Change (UNFCCC) (UNFCCC 1997) prescribe measures related to the management of forests and requires regular reporting about the state of, and change in forests. The mechanism for the reduction in the emissions from deforestation and forest degradation (REDD+) has led to an

even stronger focus on forest information (SAARELA et al., 2015). Forest inventory thus plays a central role in this strategy (STÅHL et al., 2014; SAARELA et al., 2020).

Despite this key role, the variables of interest may be difficult to access in some situations. This is the case of volume and biomass, which are key in forest management and planning, and provide support for climate and environmental policies. Although modern sensors are changing the way mensuration is done, accurately measuring tree volume and biomass has basically been impossible without destructive sampling (FORTIN et al., 2018; TRAUTENMÜLLER et al., 2021; HEROLD et al., 2019). In turn, other variables that are even more difficult to assess, such as carbon sequestration, depend on tree volume or total tree biomass. Those variables need then to be obtained somehow from tree-level measurements and a well-established approach is regression models (GERTNER, 1990; STÅHL et al., 2014). These models allow making predictions on these variables from other accessible auxiliary variables, such as DBH, which is typically measured in all trees within the sample units.

Population-level estimates incur an uncertainty that is mostly associated to the sampling design and the modelling of the relevant variables. The uncertainty related to the sampling is due to the fact that only some trees are measured, not the total population. The uncertainty linked to the model is related to the sample used to fit the model and the goodness of fit of the model itself, which in turn depends on the stochastic elements of the statistical model, such as parameter and residual uncertainty, respectively (FORTIN et al., 2018; JEVŠENAK–SKUDNIK, 2021). Uncertainty is associated with loss of information, accuracy, and reliability of estimates (FU et al., 2017), and has therefore an impact on monitoring designs, management, and policy decisions. Uncertainty estimates are needed to guide research investments (PARÉ et al., 2016) and they are mandatory in greenhouse gas emission reported according to the good practice guidance developed by the Intergovernmental Panel on Climate Change (IPCC) (BERNSTEIN et al., 2008).

Additionally, uncertainty assessment helps identify the sources (sampling, model) that contribute most to the overall uncertainty, which can be useful to improve future inventories (JONAS et al., 2010). The fraction of uncertainty attributed to the variance of the probability-based estimator of the population mean using the plot-level predictions is usually correctly estimated. On the contrary, the fraction attributable to the variance of the allometric model estimator is often ignored. These two sources of uncertainty can however be considered simultaneously through a specific type of variance estimators, termed hybrid estimators (FATTORINI, 2012; CORONA et al., 2014; FORTIN et al., 2018). In this respect, a significant research effort has been devoted over the last years to develop hybrid estimators that jointly consider the main sources of uncertainty in forest inventories (CUNIA, 1987; GERTNER, 1990; CHAVE et al., 2004; BERGER et al., 2014; FATTORINI, 2012; CORONA et al., 2014; STÅHL et al., 2014; FORTIN et al., 2018). In the literature, there are two main classes of methods that develop hybrid estimators: analytical approaches (CUNIA, 1987; GERTNER, 1990; CHAVE et al., 2004;

BERGER et al., 2014) and bootstrap simulation (BERGER et al., 2014; FORTIN et al., 2016; MCROBERTS–WESTFALL, 2015; FORTIN et al., 2018).

The bootstrap approach relies on stochastic simulation (LAROCQUE et al., 2008). This method requires assumptions on the joint distribution of errors, and it is computationally demanding (BERGER et al., 2014; MCROBERTS–WESTFALL, 2015; FORTIN et al., 2018). Analytical methods are based on linear algebra and therefore Taylor series are typically used to approximate non-linear functions. Uncertainty assessment using the analytical method has shown good results for linear biomass equations (CUNIA, 1987; CHAVE et al., 2004; MAGALHÃES et al., 2015), linear volume equations (BERGER et al., 2014; MCROBERTS–WESTFALL, 2014), and non-linear volume equations (GERTNER, 1990). In the case of non-linear models, the performance of the estimator depends on how good those approximations are (ARRAS, 1998). Some authors have in fact found it difficult to accurately predict the uncertainty of a nonlinear model using a first-order Taylor expansion (FORTIN et al., 2018). In this respect, adding higher-order terms of the Taylor expansion has been reported as a possible solution (GERTNER, 1991; SAMBAKHE et al., 2014; KANGAS, 1996). This brings up, the first scientific questions of this chapter:

i) Which order in the Taylor expansion is most appropriate for assessing uncertainty in non-linear volume and biomass models? ii) How much uncertainty due to the model parameters does each order explain?

The factors that affect the uncertainty in volume and biomass estimates are well studied, e.g. the sample size where the data are collected for model fitting purposes and for inventory purposes is known to have an effect on the uncertainty of both processes (PICARD et al., 2015; SAARELA et al., 2015; FORTIN, 2020). The variability between sample units, which depends on landscape-level heterogeneity, e.g. geographical distribution of trees (CUNIA, 1987), and the quality of the model used to predict the variable of interests (PICARD et al., 2015), are other factors that have an effect in the total uncertainty. However, it is still unclear what these effects, and their contribution to the total uncertainty are in different forest types. This raises the following scientific questions:

iii) What is the effect of the sample size used to fit the model on the Taylor approximation and on the uncertainty related to the model? iv) What is the amount of uncertainty due to the model and sampling when using different sample size to fit the models? v) What is the effect of model quality and plots variability on model- and sampling-related uncertainty?

We aimed to assess the uncertainty in volume and biomass above ground population-level estimates in *Acacia mearnsii* stands in a hybrid inference framework. The specific aims were: to assess the effects of sample size to fit volume and biomass models on sampling and model-related uncertainty; to evaluate the effects of model quality and inter-plot variability on total uncertainty.

To achieve our goals and answer the scientific questions we fitted linear and nonlinear models widely used in forest science. To quantify uncertainty in nonlinear models we used the first and second model derivatives by Taylor approximation and quantified how much each

derivative contributes to the total uncertainty due to the model. For comparison with hybrid variance estimators, we used a bootstrap estimator.

3.2 MATERIAL AND METHODS

3.2.1 Data

The study was carried out in *Acacia mearnsii* stands located in Rio Grande do Sul, south of Brazil, across three major regions: Cristal, Encruzilhada do Sul and Piratini. In each region, four stands were sampled several times after planting to cover the entire rotation (10 years), with temporary plots, without measurements over time. Specifically, the measurements were conducted at ages 1.75, 2.75, 5.00 and 10.08 years in Cristal; 1.83, 3.08, 5.75 and 10.75 in Encruzilhada do Sul; and 1, 2.33, 5.25 and 9.83 in Piratini. Decimal values are relevant given the fast growth of the species and the very short rotation length.

3.2.1.1 Data collection

Data collection consisted of two phases: plot measurement (phase 1) and subplot measurement (phase 2).

Phase 1: Plot measurement

In each stand, four circular 22,56-m diameter plots (400 m²) were randomly set up, totaling 94 temporary plots. In these plots, the diameter at breast height (cm) of 6,927 trees were measured. Twenty percent of the trees in each plot had their heights measured (m). These variables were later used to predict the stem volume and the above ground biomass components of each tree through regression models. Diameter at breast height and total height were measured using a dendrometer tape and a hypsometer (Haglöf), respectively. The statistical characteristics of the sample trees are listed in Table 3.1.

TABLE 3.1 – PLOT DATASET SUMMARY.

Descriptive statistics	dbh (cm)	h (m)
Mean	14.78	17.41
Max	33.74	21.48
min	5.09	9.25
Var	23.53	8.97
Sd	4.85	2.99
Cv%	32.81	17.19
Plot		
Cv min Plot	24.06	11.8
Cv max Plot	39.8	20.61
Var min	10.99	4.78
Var max	37.06	12.44

Where: dbh is diameter at breast height, h is height.

Phase 2: Subplot Measurement

A 10-m diameter subplot (78,54 m²) was established in the center of 48 subplots where biomass from branches and crown, and total volume were measured on 670 trees. This subsampling scheme was conducted because it was not possible to measure all variables from all trees in the larger plot, due to the high cost. Towards that end, all trees in the subplots were felled and the following measurements taken: height, total crown height, crown length, diameter at breast height, and the diameter of the cross-sections at 5%, 15%, 25%, 35%, 45%, 55%, 65%, 75%, 85%, and 95% of the total height. The stem volume was scaled using the Hohenadl method.

Biomass measurements were taken for the stem (stem wood + bark) and crown (living branches and dead leaves, flowers and fruits). For each tree, these components were separated and weighed with a digital scale (Portable Electronic Scale) with an accuracy of 5 g to determine green biomass .

The samples, aimed at calculating stem and crown biomass, were collected and weighed *in situ* using a digital balance (Hoyle) with an accuracy of 1 g. From the stem, five 2-cm thick discs were extracted at the cross-sections corresponding to the 0%, 25%, 50%, 75%, and 95% of the total height. The samples were dried in an air circulation-assisted oven at 100°C and after a constant weight, the material was weighed on a digital scale with an accuracy of 1 g. The statistical characteristics of the sample trees in the subplots are listed in Table 2.

TABLE 3.2 – SUBPLOT DATASET SUMMARY.

Descriptive statistics	dbh (cm)	h (m)	v (m ³)	bt (kg)
Mean	9.50	12.17	0.06407	42.52
Max	23.55	21.90	0.4773	361.0660
min	1	5	0.0001	0.31
Var	19.28	24.45	0,00431	2.315,96
Sd	4.39	4.94	0,065647	48.12
Cv%	46.22	40.63	102,4614	113.18

Where: dbh is diameter at breast height, h is height, v is volume and bt is total biomass.

3.2.2 Volume and aboveground biomass modelling

We used the data from the phase 2 of the data collection to fitted models with three different functional forms for the purpose of predicting individual tree volume and biomass: SPURR (1952) in linear fashion (equation 1), SPURR (1952) in non-linear fashion (equation 2) and SCHUMACHER–HALL (1933), in non-linear fashion (equation 3):

$$\hat{y} = \beta_1 + \beta_2 d^2 h \quad (3.1)$$

$$\hat{y} = \beta_1 \cdot d^2 h^{\beta_2} \quad (3.2)$$

$$\hat{y} = \beta_1 \cdot d^{\beta_2} \cdot h^{\beta_3} \quad (3.3)$$

Where: $\hat{y}$ is the variable of interest - volume and biomass aboveground - β_1 , β_2 and β_3 are the model coefficients, d is diameter at breast height and h is height.

The coefficients $[\beta]$ were estimated by ordinary least squares. The White's test (WHITE, 1980) was applied to evaluate the assumption of homogeneity of the residual variance. When heteroscedasticity was observed, we model the residual variance to establish weights and fit the model by means of the weighted least squares method. The matrix of weights was calculated as in PARRESOL (2001) (BEHLING et al., 2018).

The rest of the basic regression assumptions were checked, and the models were evaluated using fit and precision statistics: standard error of estimate in percentage ($Syx\%$), coefficient of determination (R^2) and Akaike information criterion (AIC) presented in GUJARATI–PORTER (2011).

3.2.3 Uncertainty in volume and aboveground biomass estimates

We focus on the estimator of the population mean in this particular case. To assess the uncertainty in the estimates of the mean volume and above-ground biomass, we used hybrid esti-

mators. These are well-established estimators in which it is considered the sampling and model as sources of uncertainty in their mathematical formulation to quantify their respective contributions to the total uncertainty (CUNIA, 1987; GERTNER, 1990; MCROBERTS–WESTFALL, 2014; FORTIN et al., 2018). We considered in the present study the uncertainty from the estimates of the model coefficients and from the sampling design using hybrid estimators based on analytical procedures and Bootstrap estimator.

The uncertainty was expressed in terms of the variance, which is more convenient than other statistics, especially because it can be divided among several additive components that represent a source of error, provided that certain conditions are met (CUNIA, 1987).

3.2.3.1 Analytical methods

Linear models

The development of this section follows (CUNIA, 1987) analytical procedure for quantifying uncertainty. The estimates of mean volume stock and biomass in the population combine the statistics from modelling exercise (i.e. parameter estimates), with the statistics resulting from the first phase of data collection (i.e. number of trees and explanatory variables (dbh and h), and combinations of them, as it will be described later. To this end, the following conditions must be met:

(i) the statistics of the first stage of data collection are denoted by $z_1, z_2, \dots, z_i$ and the estimates of their variances and covariances ($\sigma_{zz}^2, i = 1, \dots, j$) can be calculated for the sample plot data set, where i is the trees in the plot k.

(ii) the statistics of the second phase are denoted by $b_1, b_2, \dots, b_m$ (the total parameters estimates) and the estimates of their variances and covariances (σ_{bb}^2) can be calculated for the subplot data set.

iii) the statistics for the first phase of data collection are independent of the statistics of the modelling phase.

iv) the estimator can be expressed as an explicit function of statistics Z and $\hat{\beta}$, i.e. $f(Z, \hat{\beta})$.

The total variance can be calculated as:

$$Var(\hat{Y}) = \sum_{i=1}^p \left(\frac{\partial f}{\partial Z_i} \right) \sigma_{zz}^2 + \sum_{i=1}^m \left(\frac{\partial f}{\partial \beta_i} \right) \sigma_{bb}^2 \quad (3.4)$$

The statistics $z_1, z_2, \dots, z_i$ are calculated from the data collected in phase 2. Specifically, and for our case, the elements in matrix $[z]$ are the sum of the number of trees per hectare and sum of explanatory variables for all the trees per plot. For any given plot, we have $[z] = \begin{bmatrix} \sum_k^n n \\ \sum_k^n d^2 h \end{bmatrix}$. Matrix $[Z]$ is then defined as the mean of the elements in the matrices $[z]$ across plots.

The estimator of the covariance matrix $[\sigma_{zz}]$ is defined as:

$$[\sigma_{zz}^2] = \begin{bmatrix} \sigma_{z_1}^2 & \sigma_{z_1 z_2} & \cdots & \sigma_{z_1 z_m} \\ \sigma_{z_1 z_2} & \sigma_{z_2}^2 & \cdots & \sigma_{z_2 z_m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{z_1 z_m} & \sigma_{z_2 z_m} & \cdots & \sigma_{z_m}^2 \end{bmatrix}$$

where: the diagonal elements are the variances of the independent variables of the z matrix, and the off-diagonal values are the covariances between these independent variables. Firstly we calculate these values by plots, and secondly, we calculate the mean among them.

The estimation of the vector $[\beta]$ is denoted by $[b] = \begin{bmatrix} b_1 & b_2 & \cdots & b_m \end{bmatrix}$ and the estimate of the covariance matrix $[\sigma_{bb}^2]$ is defined as:

$$[\sigma_{bb}^2] = \begin{bmatrix} \sigma_{b_1}^2 & \sigma_{b_1 b_2} & \cdots & \sigma_{b_1 b_m} \\ \sigma_{b_1 b_2} & \sigma_{b_2}^2 & \cdots & \sigma_{b_2 b_m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{b_1 b_m} & \sigma_{b_2 b_m} & \cdots & \sigma_{b_m}^2 \end{bmatrix}$$

Substituting the above in equation 4, the variance estimator for the population mean $Var(\hat{Y})$ in the case of linear models can be obtained as:

$$Var(\hat{Y}) = [b]' [\sigma_{zz}^2] [b] + [Z]' [\sigma_{bb}^2] [Z] \quad (3.5)$$

where the first term can be interpreted as the uncertainty component due to the sampling and the second term as the uncertainty component due to the model.

Non-linear Models

When regression models are non-linear, the estimator of volume and biomass and their variances can be so complex that it becomes extremely cumbersome to apply the above formula (CUNIA, 1987). Thus, for non-linear models, we use the law of error propagation to calculate the uncertainty due to the model parameters. In the non-linear framework, this implies linearising the functional form of the model through approximation to a, typically, first-order Taylor series expansion. However, depending on the degree of nonlinearity, the first-order term may not be a good enough approximation (GERTNER, 1991), and a higher-order Taylor series would be needed (GERTNER, 1991; KANGAS, 1996; SAMBAKHE et al., 2014), which is the approach we tested here. The first and second derivatives with respect to the parameters of the functional forms for our two non-linear models are:

Model	First derivative	Second Derivative
2	$f(\beta_1) = d^2 h^{\beta_2}$ $f(\beta_2) = \beta_1 \cdot d^2 h^{\beta_2} \cdot \log_{d^2 h}$	$f(\beta_1) = 0$ $f(\beta_2) = \beta_1 \cdot d^2 h^{\beta_2} \cdot \log_{d^2 h} \cdot \log_{d^2 h}$
3	$f(\beta_1) = d^{\beta_2} \cdot h^{\beta_3}$ $f(\beta_2) = \beta_1 \cdot d^{\beta_2} \cdot h^{\beta_3} \cdot \log_d$ $f(\beta_3) = \beta_1 \cdot d^{\beta_2} \cdot h^{\beta_3} \cdot \log_h$	$f(\beta_1) = 0$ $f(\beta_2) = \beta_1 \cdot d^{\beta_2} \cdot h^{\beta_3} \cdot \log_d \cdot \log_d$ $f(\beta_3) = \beta_1 \cdot d^{\beta_2} \cdot h^{\beta_3} \cdot \log_h \cdot \log_h$

If these derivatives are substituted in the law of error propagation, the uncertainty due to the model coefficients can then be approximated as:

$$Var_m = \sum_{i=1}^p \left(\frac{\partial f}{\partial \beta_i} \sigma_{bb} \right)^2 + \frac{1}{2} \left(\sum_{i=1}^p \left(\frac{\partial^2 f}{\partial \beta_i} \sigma_{bb} \right)^2 \right)^2 \quad (3.6)$$

which in matrix notation yields

$$Var_m = [J]' [\sigma_{bb}^2] [J] + \frac{1}{2} Tr([H] [\sigma_{bb}^2])^2 \quad (3.7)$$

where $[J]$ denotes the Jacobian matrix, i.e. matrix of the first derivatives with respect to $\hat{\beta}_i$, $[J]'$ is its transposed, $[H]$ is the Hessian matrix, i.e. matrix of the second partial derivatives with respect to $\hat{\beta}_i$, and $[\sigma_{bb}]$ is the covariance matrix of the parameter estimates.

The uncertainty due to the design was calculated using the estimator of the variance of the population mean estimator for simple random sampling (PÉLLICO NETTO, 1997).

$$Var_d = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}{n - 1} \quad (3.8)$$

Where: n is the total sample size and $\hat{y}_i$ is the observation for the i^{th} plot selected for the sample.

The total uncertainty in nonlinear models can be obtained by adding model and design uncertainty together, as follows:

$$Var(\hat{Y}) = Var_d + Var_m \quad (3.9)$$

3.2.3.2 Bootstrap estimator

Bootstrap estimator was used for comparison with analytical approach. This technique consists in computing a large number of realizations of some random variables in order to reproduce the variability of a process (FORTIN et al., 2018).

To assess the uncertainty with bootstrap estimator, sequential steps are followed:

(i) For a given realization, random deviations are generated to account for the model-related uncertainty. The random deviation value is added to the model parameters and the joint distribution of the parameters, which is given by $\beta N(\hat{\beta}, \Omega)$, with Ω the variance-covariance matrix of the coefficients

(ii) Values of volume and above ground biomass in the plots are obtained in $\bar{Y}_b$ with coefficients generated in (i).

(iii) We ran the simulations B realizations.

(iv) To distinguish the contribution of each source of uncertainty to the total uncertainty we can use the hybrid estimator, presented in the equation (10):

$$Var(\hat{Y}) = \frac{\sum (\hat{Y}_b - \hat{Y}_{BS})^2}{B} + \frac{\sum \widehat{Var_d} [\hat{Y}_b]}{B} \quad (3.10)$$

Where: for each realisation, an estimate of the mean volume or biomass $\hat{Y}_b$, the mean for all realizations $\hat{Y}_{BS}$, B is the number of realization, in our case 10.000 and $\widehat{Var_d} [\hat{Y}_b]$ represents the contribution of the sampling to the total variance.

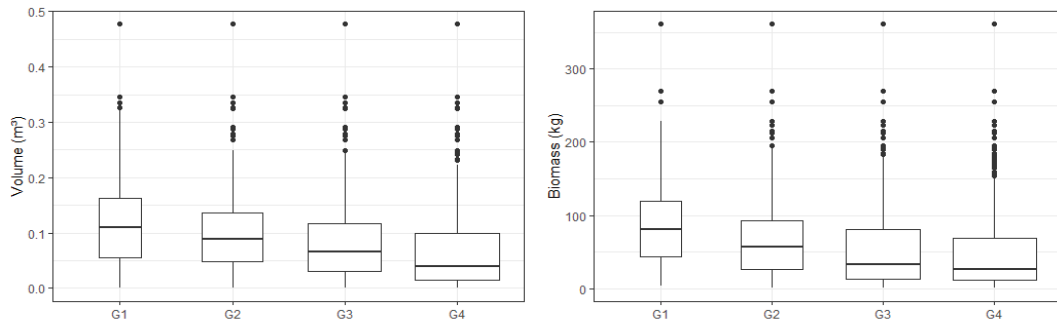
3.2.4 Effect of the sample size to fit the model in the total uncertainty

To verify the influence of number of trees in the uncertainty due to the model, we fitted the models with different number of trees. We separated four groups, to which the number of trees were grouped by productivity class:

G1 170 trees, G2 370 trees, G3 500 trees and G4 670 trees.

The three models were fitted for each set of trees. The total uncertainty of the estimates was evaluated in each model, with the analytical and bootstrap estimator. Data description for each group are presented in box-plot graphs, in figure 3.1.

FIGURE 3.1 – BOX-PLOT OF THE VOLUME AND BIOMASS DATASETS ACCORDING TO SAMPLE SIZE.



SOURCE: The author.

3.2.5 Effect of inter-plot variability and model goodness of fit in the total uncertainty

Considering the conception of the uncertainty due to the model is very small and almost negligible in some situations, such as when the quality of the model is good (MCROBERTS–WESTFALL, 2014), and assuming that, similarly, it may be the case that sampling uncertainty is close enough to zero in situations where the plots are very homogeneous and well distributed in the population so that it is reasonable to propose that one of these uncertainties could be ignored (CUNIA, 1987). We performed simulations to check when these conditions occur in stands.

We simulated two volume data sets and two biomass data sets, with values to fit Spurr's linear model, resulting in R^2 values of 0.8 and 0.5. We used a deviation value for the input variables to find these values. Thus, we can observe the influence of equation quality in the total uncertainty.

In addition, we separated our plots into variance classes, according to variance in the input variables of the model d^2h , to notice the behavior of the combination of the two error sources:

Class I: Low within-plot variance.

Class II: Medium to high within-plot variance.

Class III: High within-plot variance.

The entire processing was carried out in R.

3.3 RESULTS

3.3.1 Volume and aboveground biomass models

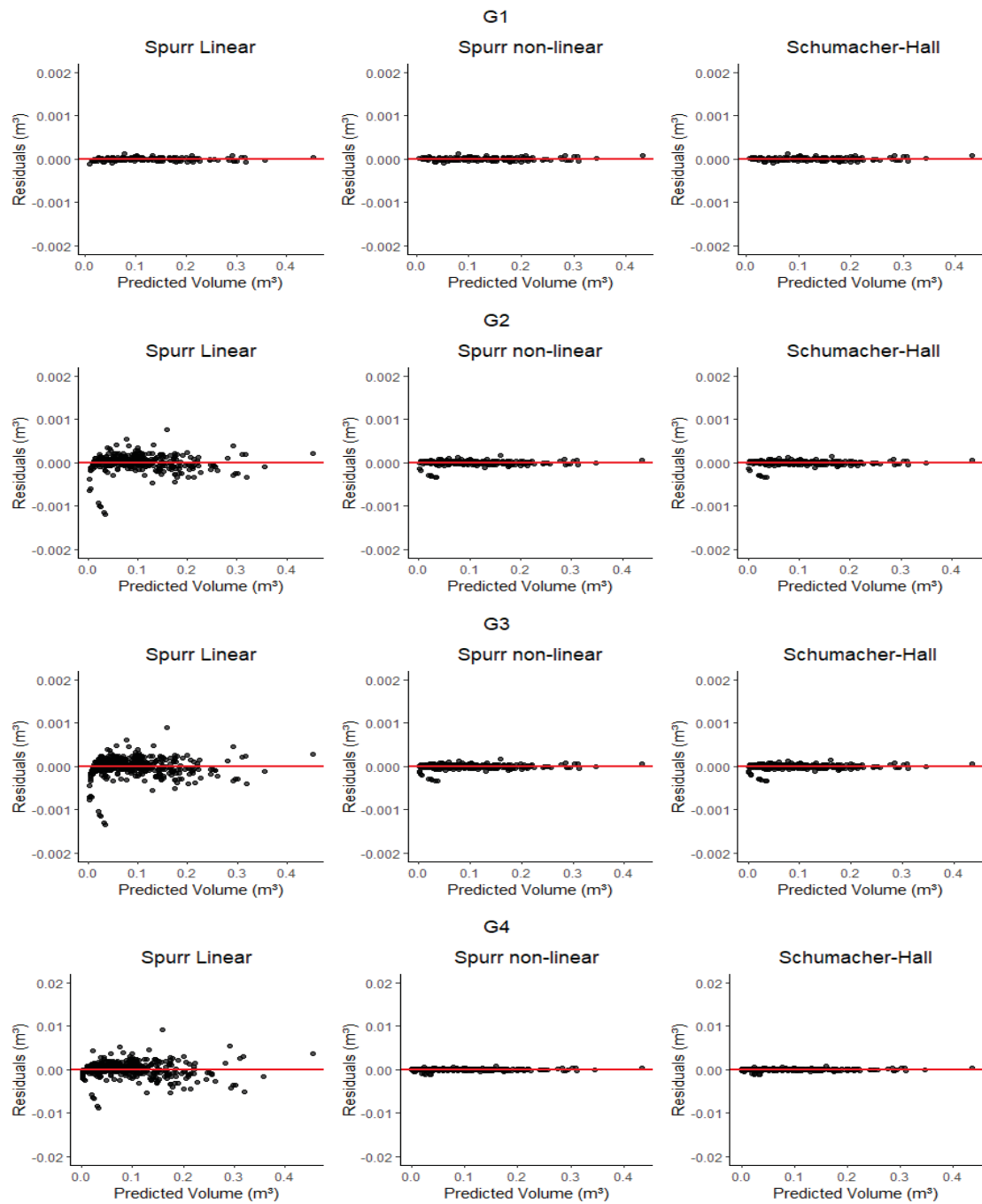
The goodness of fit of all three volume and biomass models within the four fitting groups were similar in terms of their AICs (TABLE 3.3 and 3.4). All models across the fitting groups exhibited very high coefficient of determination (over 0.98 for volume and over 0.96 for biomass). Syx% had a small change between G2 and G3. In G1 and G4 it had an expressive change, varying around 7% to 10%. The residuals from model 1 exhibited a slight trend in G2, G3 and G4. Otherwise, the graphs of residuals revealed no systematic lack of fit nor any heteroscedastic pattern (Figure 3.2 and 3.3).

TABLE 3.3 – VOLUME EQUATIONS AND GOODNESS-OF-FIT FOR VOLUME BY SAMPLE SIZE GROUP.

Model	Equation	$R^2_{adjust.}$	AIC	$S_{yx}\%$	CV%
G1					
1	$\hat{v} = 0.004278 + 0.00003724 \cdot d^2h$	0.989	-1152.495	7.88	64.35
2	$\hat{v} = 0.00005378 \cdot d^2h^{0.9586}$	0.984	-1157.515	7.91	63.46
3	$\hat{v} = 0.0000608617 \cdot d^{1.8731556} \cdot h^{0.957134}$	0.984	-1194.764	8.3	63.99
G2					
1	$\hat{v} = 0.002377 + 0.00003745 \cdot d^2h$	0.984	-2375.48	9.05	70.76
2	$\hat{v} = 0.00005378 \cdot d^2h^{0.9586}$	0.983	-2396.035	8.97	69.71
3	$\hat{v} = 0.00005718 \cdot d^{1.937} \cdot h^{0.9182}$	0.983	-2394.680	8.96	69.73
G3					
1	$\hat{v} = 0.0025476 + 0.0000374 \cdot d^2h$	0.986	-3569.267	9.51	81.19
2	$\hat{v} = 0.00005651 \cdot d^2h^{0.9525}$	0.986	-3625.320	9.46	79.98
3	$\hat{v} = 0.00005737 \cdot d^{1.912} \cdot h^{0.9404}$	0.986	-3623.442	9.46	79.98
G4					
1	$\hat{v} = 0.0019830 + 0.0000375 \cdot d^2h$	0.990	-4956.491	10.71	103.01
2	$\hat{v} = 0.0000567 \cdot d^2h^{0.9520817}$	0.989	-5227.849	10.58	101.94
3	$\hat{v} = 0.00005759 \cdot d^{1.913} \cdot h^{0.9380}$	0.989	-5226.125	10.59	101.94

Where: $R^2_{adjust.}$ is coefficient of determination, AIC is Aikake information criteria, $S_{yx}\%$ is standard deviation of estimation CV% is coefficient of variation.

FIGURE 3.2 – RESIDUALS RESULTS FOR VOLUME EQUATIONS BY SAMPLE SIZE GROUPS.



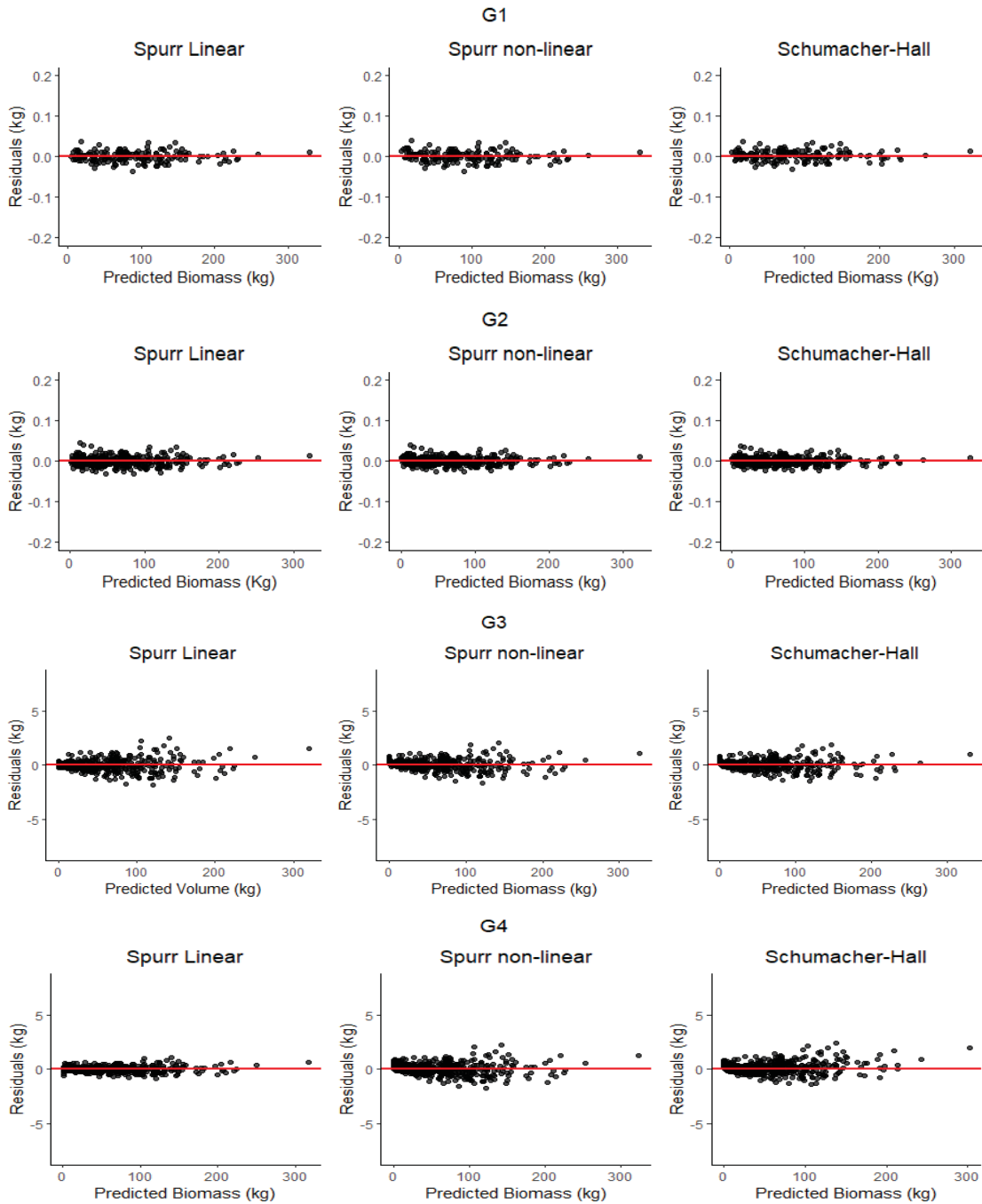
SOURCE: The author.

TABLE 3.4 – BIOMASS EQUATIONS AND GOODNESS-OF-FIT FOR VOLUME BY SAMPLE SIZE GROUP.

Model	Equation	$R^2_{adjust.}$	AIC	$S_{yx}\%$	CV%
G1					
1	$\hat{b} = 0.7697 + 0.0273 \cdot d^2h$	0.969	1267.124	14.04	66.06
2	$\hat{b} = 0.02796 \cdot d^2h^{0.99809}$	0.958	1282.17	14.01	66.72
3	$\hat{b} = 0.05279 \cdot d^{2.17934} \cdot h^{0.60561}$	0.960	1239.008	14.16	66.67
G2					
1	$\hat{b} = 0.20765 + 0.026 \cdot d^2h$	0.971	2581.494	15.51	74.39
2	$\hat{b} = 0.02607 \cdot d^2h^{1.00306}$	0.959	2577.019	15.48	74.82
3	$\hat{b} = 0.03841 \cdot d^{2.14656} \cdot h^{0.73634}$	0.961	2566.733	15.11	75.11
G3					
1	$\hat{b} = 0.52025 + 0.02655 \cdot d^2h$	0.977	3533.90	17.63	97.49
2	$\hat{b} = 0.0237 \cdot d^2h^{1.0144}$	0.969	3520.29	17.52	99.60
3	$\hat{b} = 0.03454 \cdot d^{2.17318} \cdot h^{0.74703}$	0.970	3503.65	17.17	99.45
G4					
1	$\hat{b} = 0.78810 + 0.02634 \cdot d^2h$	0.978	4272.67	18.26	102.73
2	$\hat{b} = 0.02498 \cdot d^2h^{1.00770}$	0.970	4364.44	18.17	105.21
3	$\hat{b} = 0.03078 \cdot d^{2.12057} \cdot h^{0.83503}$	0.971	4353.24	18.00	105.04

Where: $R^2_{adjust.}$ is coefficient of determination, AIC is Aikake information criteria, $S_{yx}\%$ is standard deviation of estimation CV% is coefficient of variation.

FIGURE 3.3 – RESIDUALS RESULTS FOR BIOMASS EQUATIONS BY SAMPLE SIZE GROUPS.



SOURCE: The author.

3.3.2 Uncertainty Assessment

3.3.2.1 Effect of the sample size to fit the model on the total uncertainty

We considered the total uncertainty by adding the model-related uncertainty together with the uncertainty due to inventory sampling. The impact of increasing the sample size to fit models can clearly be seen in figures 3.4 and 3.5 for both studied variables. The variance due

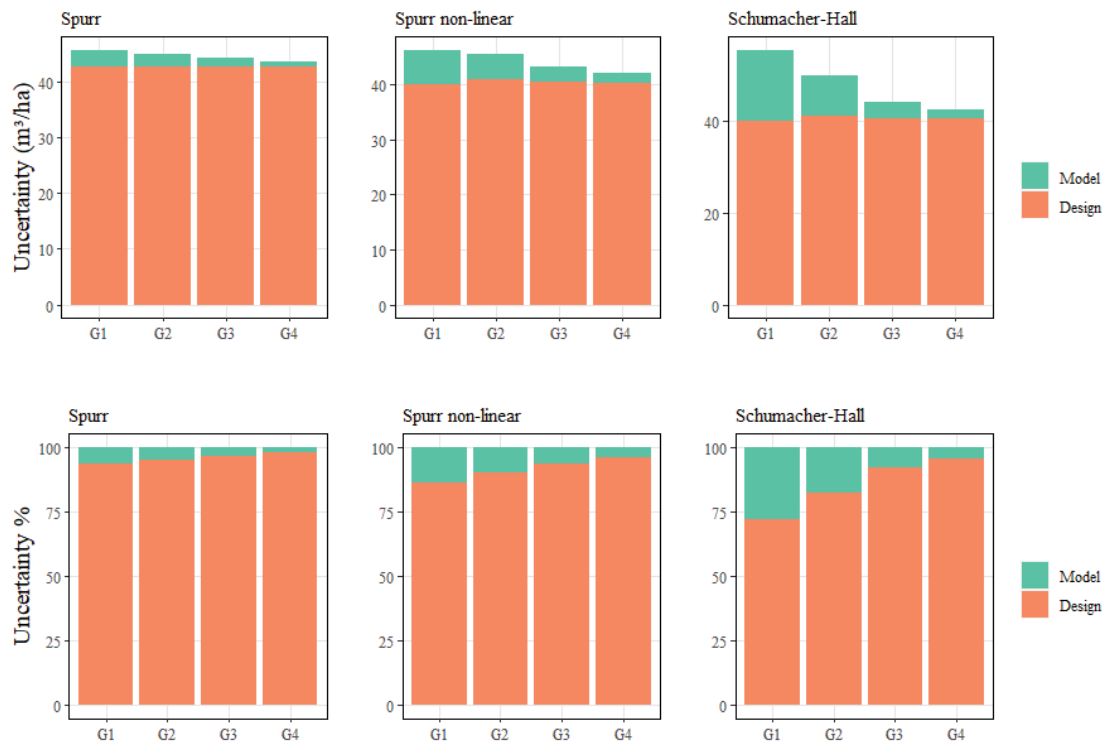
to the model was affected for sample size. Consequently, the relative contribution of the error source due to the model changed through the groups with different number of trees, where, by increasing the number of trees to fit the models, the model-related variance decreased, therefore the total uncertainty in forest inventory estimates decreased with increasing number of trees for fitting models. All models showed these results in all groups.

Error due to sampling was almost the same for all groups and models for volume and biomass. We noticed that the absolute error due to the design was slightly higher with model 1 and the associated method to assess uncertainty than with the other two models, even though it presented in fact a smaller error due to the model. This is true for both volume and biomass estimates. The results for this model exhibited the least variation across the groups in this respect ($\sim 42 \text{ m}^3 \cdot \text{ha}^{-1}$). Model 2 figures ranged from $39.97 \text{ m}^3 \cdot \text{ha}^{-1}$ to $40.92 \text{ m}^3 \cdot \text{ha}^{-1}$. Model 3 presented around 39.95 to $40.95 \text{ m}^3 \cdot \text{ha}^{-1}$.

In volume estimates, the proportional contribution of the sampling was the largest in all groups and accounted for more than 93% and 85% of the total uncertainty in models 1 and 2, respectively. In the case of model 3, this contribution was more than 70% in G1, increasing to 95% in G4. With respect to the biomass estimates, the contribution of the sampling uncertainty is notably smaller than in the volume estimates and there is a greater variation between groups. The contribution of the sampling was 29% in G1 and 77% in G4 for model 3. It was an increase to 60% in G2, 75% in G3, and 81.54% in G4. Error due to sampling contributed 81% in G1, and 94% in G4 in model 1, and model 2 the uncertainty changed from 56% to 90% of sampling contribution.

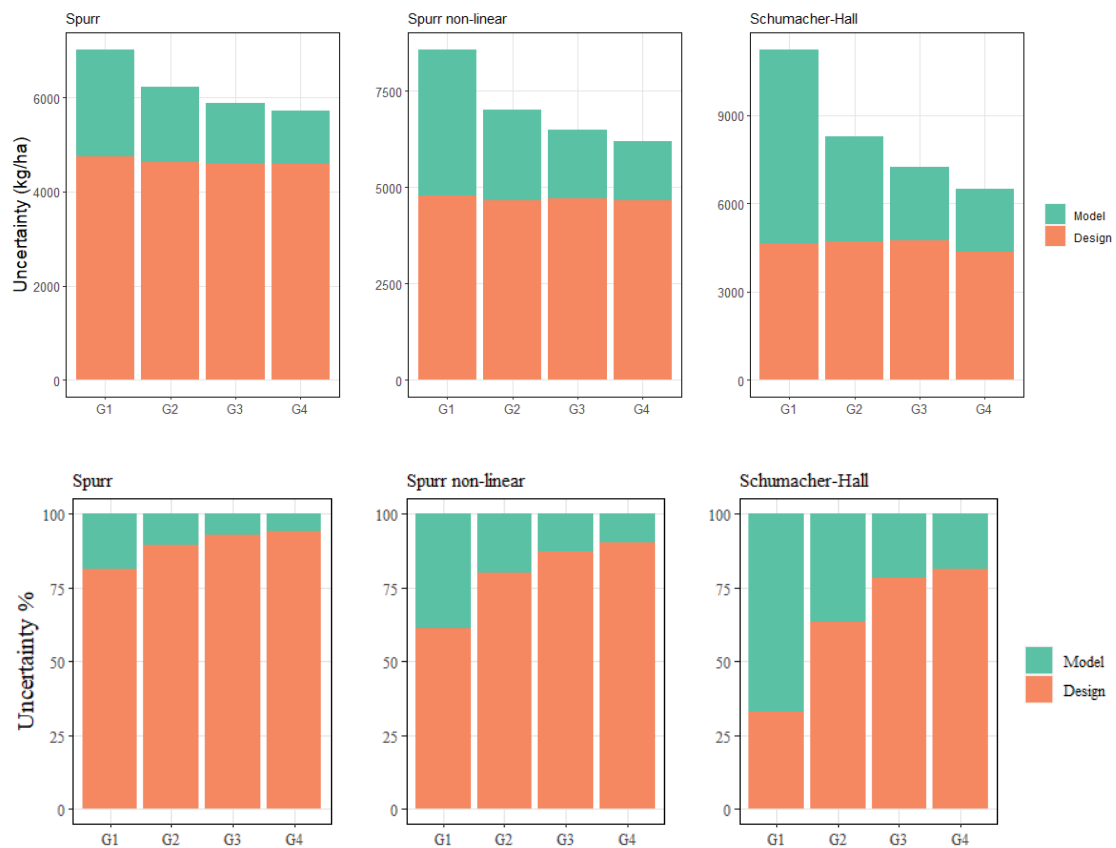
Both methodologies presented similar results in the quantification of the contribution for each source of uncertainty, in the total uncertainty of the estimates (Appendix 2).

FIGURE 3.4 – ABSOLUTE AND RELATIVE CONTRIBUTION OF SOURCES OF ERROR FOR THE TOTAL UNCERTAINTY IN VOLUME ESTIMATES.



SOURCE: The author.

FIGURE 3.5 – ABSOLUTE AND RELATIVE CONTRIBUTION OF SOURCES OF ERROR FOR THE TOTAL UNCERTAINTY IN BIOMASS ESTIMATES.

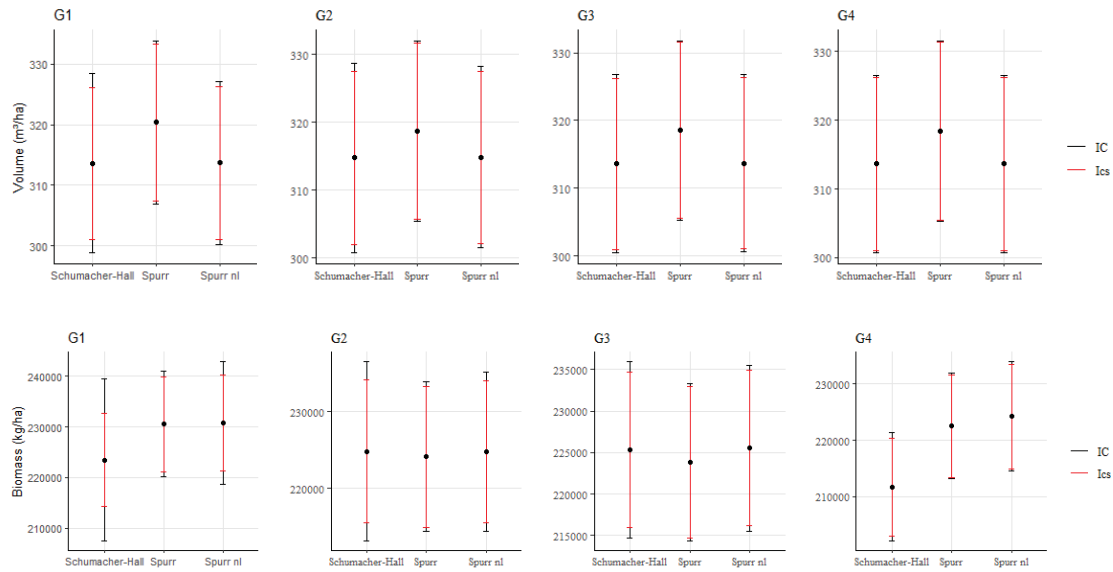


SOURCE: The author.

The confidence interval reflects the results presented earlier (FIGURE 3.6). The sample size to fit models affected the width of the confidence intervals, which decreased in G4. The coefficient intervals without considering the model are almost the same for all models in all groups since the error due to sampling is similar for the models. The confidence intervals are underestimated not considering the model-related uncertainty, as we can observe with red width.

The mean volume per hectare, estimated by model 2, was higher compared to the others in all groups (FIGURE 3.6). The mean biomass was lower with model 1 in G1 and G2.

FIGURE 3.6 – CONFIDENCE INTERVALS FOR ALL a) VOLUME AND b) BIOMASS MODELS.



SOURCE: The author.

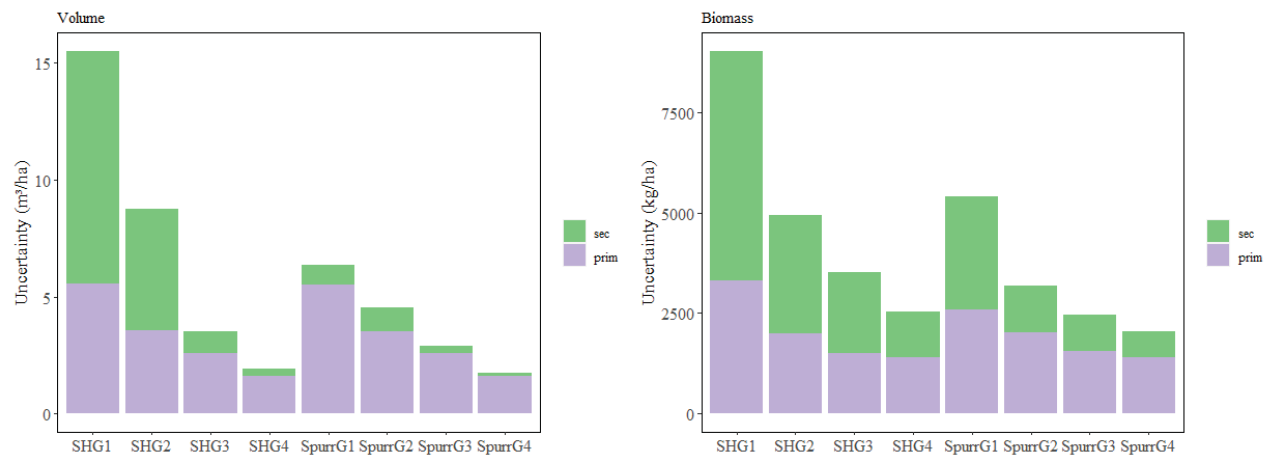
3.3.2.2 Effect of the sample size to fit the model on derivatives

We assessed how much model-related uncertainty was taken into account by the second-order expansion of the Taylor series, in addition to the first-order derivative, in our non-linear models. Our results showed that the higher the uncertainty generated by the model coefficients, the more the second derivative contributed to the total error, as we can observe in G1 for both models (FIGURE 3.7). When the model-related uncertainty is low, the largest contribution to the total uncertainty lies in the first derivative, as we can observe in G3 and G4.

Since model 2 showed less uncertainty, the contribution of the second derivative to the total uncertainty in model coefficients was low in all groups for both variables. Volume in model 3 showed a high contribution of the second derivative (highest contribution) in G1 and G2, and this contribution decreased in G3 and G4. For biomass, we observed a higher contribution of the second derivative in all groups with model 3, this being more than 50%.

The degree of non-linearity of the models had a great influence on the second derivative, which, consequently, contributed to the curvature of the uncertainty per tree (FIGURE 3.8). Models 1 and 2 showed less dispersion to uncertainty per tree, while model 3 showed more dispersion in all groups. In G1, this dispersion is more visible. We compared the results of total uncertainty, presented in figure 3.7, with the bootstrap estimator and the results proved to be the same as those given by bootstrap simulations.

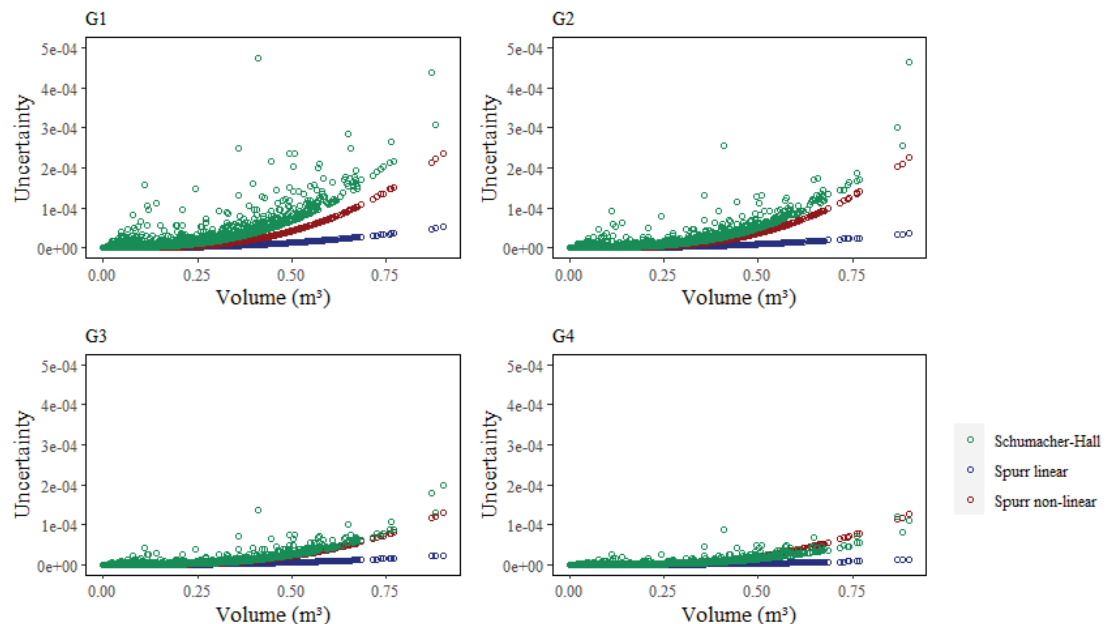
FIGURE 3.7 – CONTRIBUTION OF FIRST AND SECOND DERIVATIVES FOR THE UNCERTAINTY DUE TO THE MODEL.



SOURCE: The author.

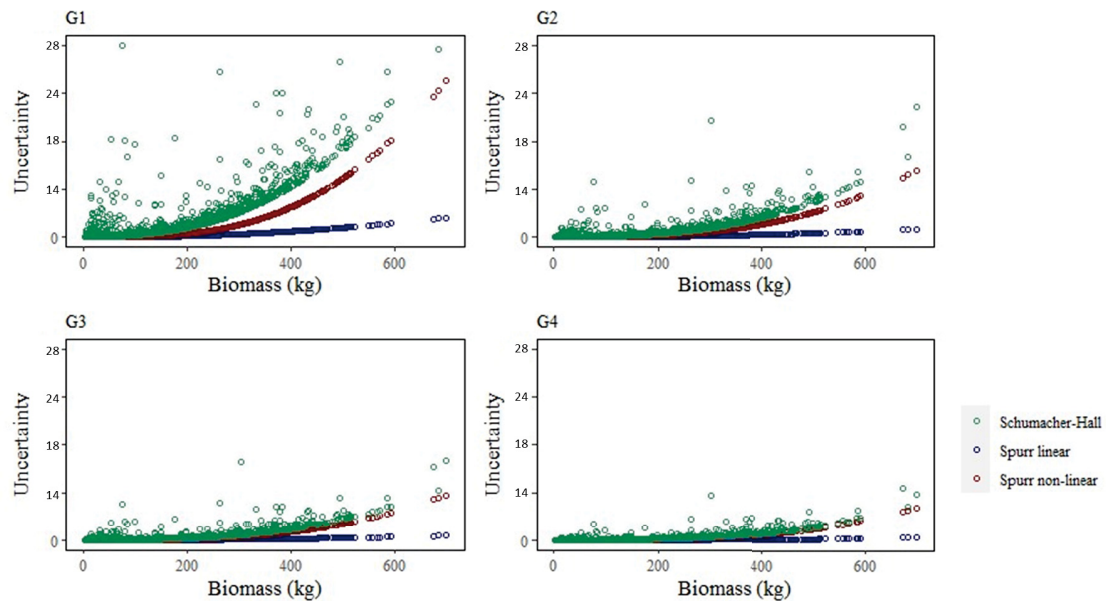
Where: SHG1 is Schumacher-Hal to group 1, SHG2 is Schumacher-Hal to group 2, SHG3 is Schumacher-Hal to group 3, SHG4 is Schumacher-Hal to group 4. SpurrG1 is Spurr model to group 1, SpurrG2 is Spurr model to group 2, SpurrG3 is Spurr model to group 3, SpurrG4. is Spurr model to group 4.

FIGURE 3.8 – UNCERTAINTY PER TREE IN VOLUME ESTIMATES.



SOURCE: The author.

FIGURE 3.9 – UNCERTAINTY PER TREE IN BIOMASS ESTIMATES.



SOURCE: The author.

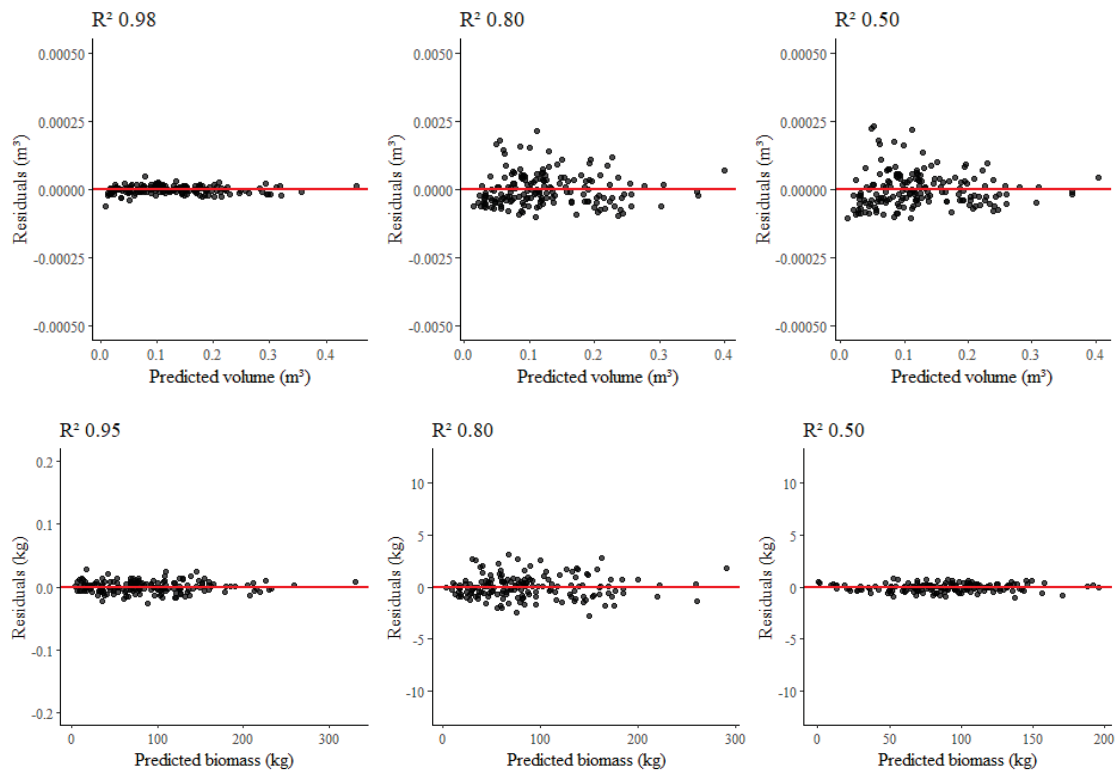
3.3.2.3 Effect of plot variability and goodness of fit of volume and biomass models in the total uncertainty

The results of goodness of fit are presented in table 3.5. We obtained one equation considered to be of good quality, with real data, and two other equations with simulated data, of which the R^2 is equal to 0.8 and the other equation of R^2 is equal to 0.5. After readjusting the $R^2 = 0.5$ equation with the weights, the volume weighted average estimate was almost the same as the obtained by the equation $R^2 = 0.8$.

TABLE 3.5 – GOODNESS OF FIT FOR VOLUME AND BIOMASS MODELS, FOR SIMULATED DATA.

Equation	$R^2_{adjust.}$	AIC	$S_{yx}\%$	White test
Volume				
$\hat{v} = 0.004278 + 0.00003724 \cdot d^2h$	0.989	-1152.49	7.89	2.83
$\hat{v} = 0.006352 + 0.00003875 \cdot d^2h$	0.819	-675.32	28.13	0.56
$\hat{v} = 0.007724 + 0.00003829 \cdot d^2h$	0.813	-663.75	28.08	6.62
Biomass				
$\hat{b} = 0.78332 + 0.02729 \cdot d^2h$	0.969	1265.87	120.31	0.68
$\hat{b} = -2.6084 + 0.02812 \cdot d^2h$	0.83	1560.57	115.18	1.27
$\hat{b} = -100.768 + 0.06684 \cdot d^2h$	0.57	1727.83	105.74	8.9

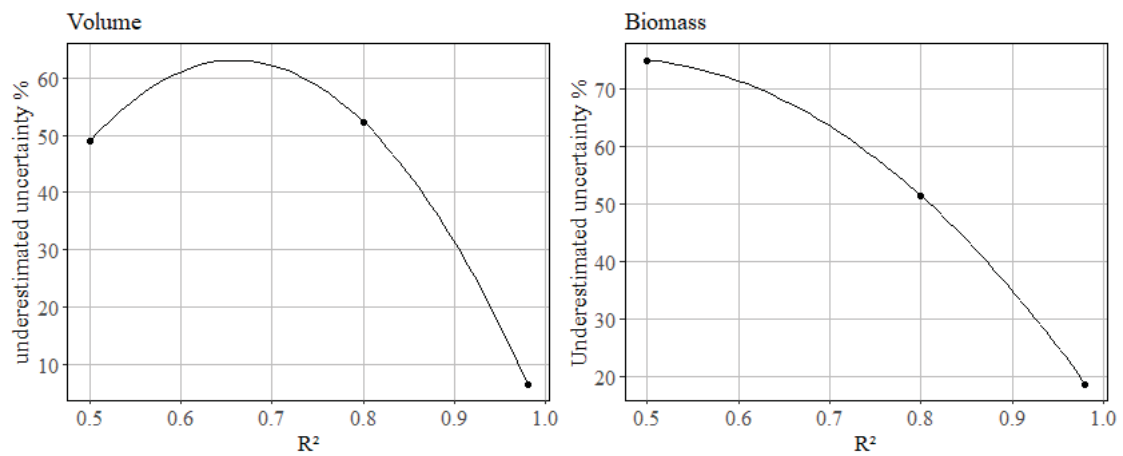
FIGURE 3.10 – RESIDUALS OF a)VOLUME AND b)BIOMASS EQUATIONS.



SOURCE: The author.

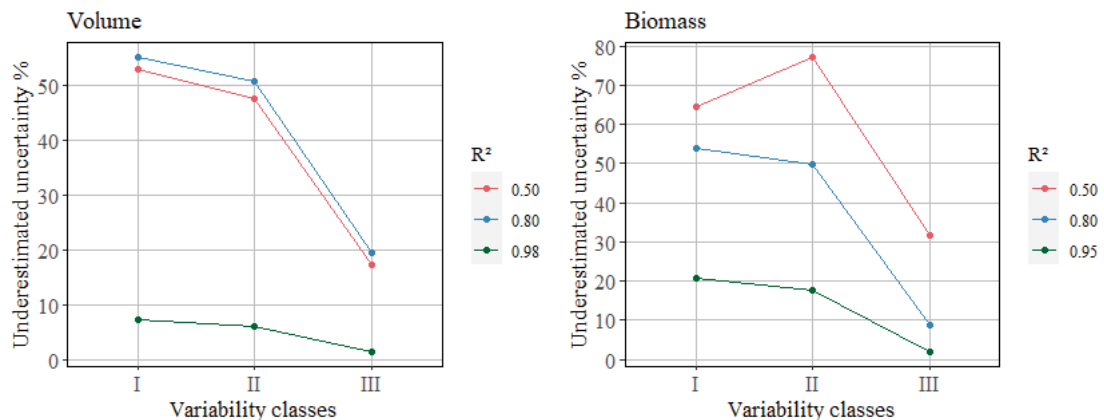
According to the models fitted to the simulated data 1, variance was underestimated by more than the equation with $R^2 = 0.9$ for both variables. To volume, the underestimated variance was 6.51% with the equation $R^2 = 0.98$, 54.33% to the $R^2 = 0.80$ and 70.79% to the equation $R^2 = 0.5$. For Biomass the underestimated variation was 17.97% for the $R^2 = 0.95$, $R^2 = 0.80$ was 50.76%, and $R^2 = 0.65$ was 75.08%.

FIGURE 3.11 – UNCERTAINTY UNDERESTIMATED ACCORDING TO GOODNESS OF FIT TO VOLUME AND BIOMASS



SOURCE: The author.

FIGURE 3.12 – UNDERESTIMATED VARIANCE WHEN IGNORING THE MODEL UNCERTAINTY, IN DIFFERENT VARIANCE CLASSES OF PLOTS FOR VOLUME AND BIOMASS.

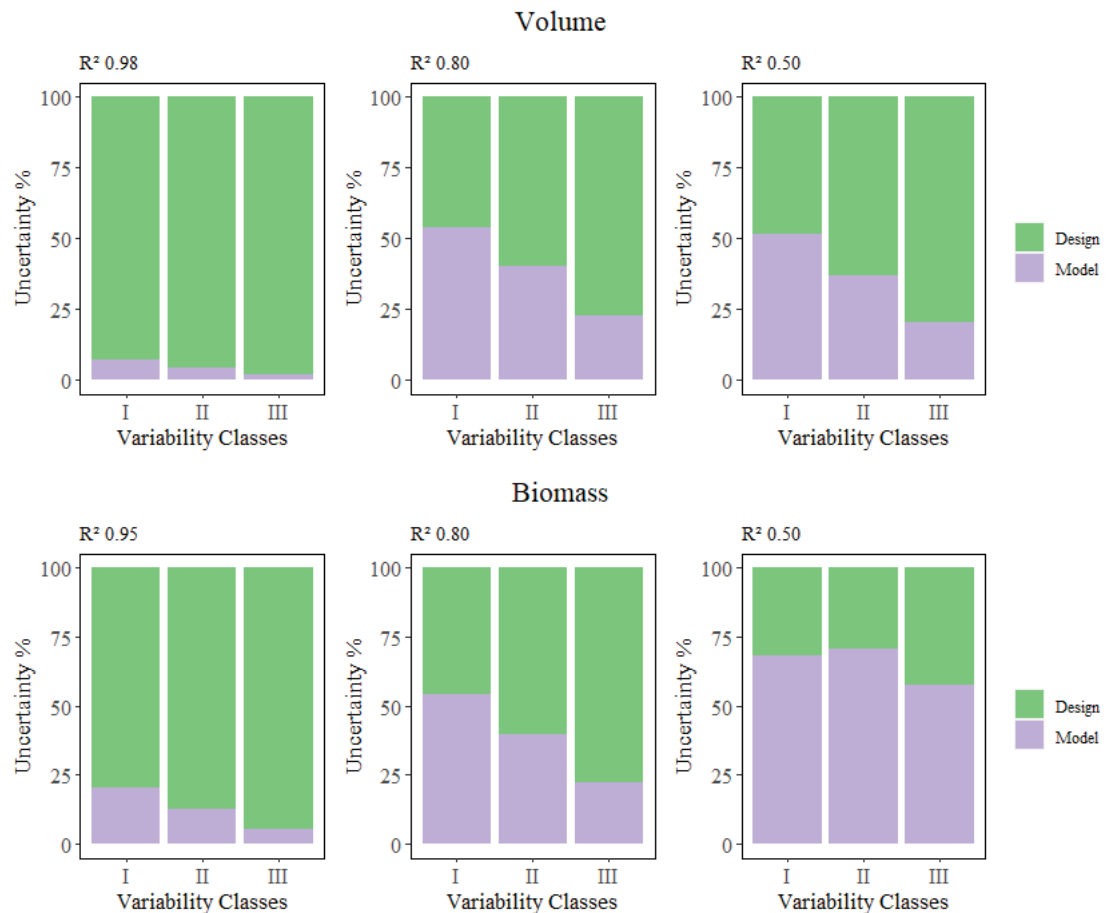


SOURCE: The author.

The relative uncertainty contribution to the total uncertainty in volume and biomass estimates is highlighted in figure 3.13. In plots with lower variability, the contribution of the model-related uncertainty to the total decreases and the sampling error has a greater contribution to the uncertainty due to the sampling. This contribution is greater than that of the model. When considering the model fitted to the original data ($R^2 = 0.98$), the contribution of the model-related variance is lower for the total variance. However, for the equation of 0.8, in plots with low variability, the contribution of the equation error was greater, and in plots with high variability, the contribution of the sampling was greater. Biomass equation 0.5 had a greater contribution to the total uncertainty for all plot variability classes. For biomass above-ground the variance due to the sampling was 81.26% and the variance of the second component of sampling error was

18.74%. This means that to disregard the propagation of error due to the equation of volume, the variance of 6.51% would be underestimated, and in the biomass equation 18.74%, reflecting on the quality of the value of the confidence interval for the mean. $R^2 < 80\%$ leads to the variance due to the model being larger than that of the design, even with highly variable plots.

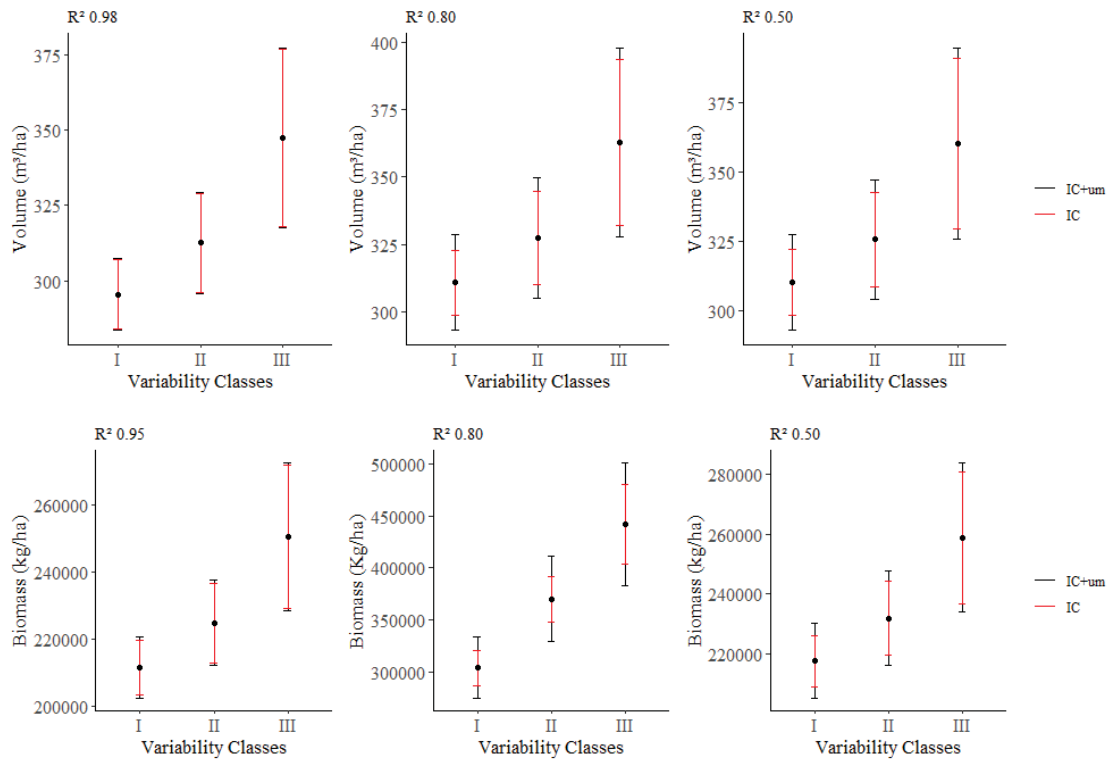
FIGURE 3.13 – CONTRIBUTION OF MODEL AND SAMPLING TO THE TOTAL UNCERTAINTY IN VOLUME AND BIOMASS ESTIMATES.



SOURCE: The author.

The confidence interval for the equation with good quality presents similar values between confidence intervals - CI of the simple random sampling, with the combined variance. For equations with R^2 of 0.8 and 0.65 where the contribution of the equation to the total variance is greater than the contribution of the sampling variance, the confidence interval with the combined variance is greater, as it treats the two variances together. As the data were simulated so that we found the quality of fit that we could express uncertainty, the data for equation 0.8 generated larger values for both volume and biomass mean.

FIGURE 3.14 – CONFIDENCE INTERVALS.



SOURCE: The author.

3.4 DISCUSSION

In this chapter, we quantified the effect of several factors on the variance of the volume and biomass estimates of the mean in the framework of hybrid inference, through the case study of *Acacia mearnsii* in South Brazil. These factors were sample size to fit the model, within-plot variability and goodness of fit and, to our knowledge, had not been jointly evaluated before with a single dataset, which contributes to the originality of the present study. Linear and non-linear allometric equations for individual trees were constructed specifically for this study so that the covariances of the model parameter could be estimated rigorously.

The sample size exerts a large influence on model-related uncertainty. Model-related uncertainty varies greatly by increasing the number of trees, especially for the nonlinear models, it decreases with an increase in the sample size. This is specifically due to the model parameters being estimated more accurately as a result of the sample being more evenly distributed across the population (SAARELA et al., 2015). As a consequence, the relative contribution of the model-related uncertainty changed through the group sample size. Another important consideration is that the sample size has a strong impact on the residual variance, and the residual variance is used to calculate the variance-covariance matrix of the parameter estimates. Hence the effect of the sample size on the uncertainty of the parameter estimates.

Furthermore, the correlation between the parameter estimates is also affected by sample size of the fitting dataset, which was more noticeable in the non-linear models. This is considered another reason for the variation in uncertainty of the parameters of the non-linear models at different sample sizes. It is common for parameters to be correlated with each other (XU–GERTNER, 2008), as this emerges as a result of unknown mechanisms between the parameters, as it could be the parameter itself that contributes to the correlation (XU–GERTNER, 2008). The correlation between parameters estimates is part of the uncertainty linked to these estimates via the variance-covariance matrix of the parameters and therefore may affect the magnitude of the uncertainty. The extent of such impacts, in general, depends on the shape, non-linearity and non-interactions with the response variable, the magnitude of the correlations and the types of distribution (GUJARATI–PORTER, 2011).

The contribution of the sampling uncertainty to the total variance remained approximately the same regardless of the sample size for model fitting. The decrease in uncertainty in one source does not interfere with the other source but only in the total uncertainty (FORTIN, 2020; PICARD et al., 2015; SAARELA et al., 2015). Our findings are in accordance with other surveys on this topic (MCROBERTS–WESTFALL, 2014; STÅHL et al., 2014; PICARD et al., 2015; SAARELA et al., 2015; FORTIN, 2020). For Fortin (2020), the size of the data sets was decisive in explaining the small model-related variation for the model estimates in France. The model fitting was performed with thousands of observations and the estimated variations of the estimated parameters were quite small (FORTIN, 2020). Stahl et al. (2014) reported that the rather small model error contributions were due to the large number of sample trees used in the model fitting and with fewer sample trees, the model error contributions could be substantial to biomass models in Finland and Sweden. McRobert-Westfall (2014) noted that using smaller data sets for model fitting resulted in greater covariance in the model parameters and, therefore, larger standard errors in the final estimates.

Even though sampling error was not affected by the change in the size of the fitting sample, it remained the main source of error. It appears that sampling uncertainty represents the biggest share in the total uncertainty for linear models at all sample size groups and nonlinear when sample size increases and the variance related to the model is low (over 90% for total uncertainty in volume and 80% for biomass). We also noticed a large contribution of sampling-related uncertainty when we separated the plots of higher variability and the equation presented a good fit. As a matter of fact, model-related uncertainty appeared to be the main source of uncertainty only for nonlinear models in biomass, when the sample size is low and in simulating equations presented a not-so-good quality. According to research by McRobert-Westfall (2014,2015) and Fortin et al. (2016), the uncertainty related to the sampling is typically larger than that from the model.

Model-related uncertainty can usually be reduced by collecting more information and raising the size of the dataset for model assembly. From a practical point of view, taking into account that collecting more information, on direct measurements of volume and biomass is a

cost and time-intensive process, and often more information is not available. The viable option is to choose a model that best fits the data set in the sense of a good fit and also reduced uncertainty. In our study, the linear model is an option for situations with few trees, the difference in model parameter uncertainty was low between sample size groups, about 4% between the class with fewer and more trees for volume. For biomass, the difference was larger but also expressed less uncertainty in the smaller sample sizes compared to the nonlinear models. It is important to describe these forms of variation to make efficient use of relatively small numbers of biomass (TEMESGEN et al., 2015) and volume measurements. Here we highlight the importance of not only analyzing the model for good fit, but also the uncertainty in the models, as there is an imbalance between the quality of the model fit and its propagated uncertainty. The quality of the fit statistic expresses how well the equation is able to explain the dependent variable, while the uncertainty expresses how much the model coefficients may be losing information and propagating error in the overall estimates.

The addition of the second derivative to the first one resulted in the exact expression of uncertainty once the second derivative is closely associated with the non-linearity of the response equations (SAMBAKHE et al., 2014; GERTNER, 1991). Our findings showed that the more non-linear a model is the worse its uncertainty is approximating to a first order Taylor series. Therefore, when we have high uncertainty in non-linear models, we estimated a dominant uncertainty in the second derivative. When the uncertainty decreases, the Hessian matrix that is involved in the second derivative is close to zero, and the contribution of the first order and the jacobian matrix is higher. This highlights the need to take into account that higher-order series will depend mainly on the non-linearity of the model and sources that affect the degree of uncertainty in the model parameters, as the results in our case study: the sample size to fit the model.

The first-order Taylor series has successfully been used by some authors for linear and nonlinear models (GERTNER, 1990; BERGER et al., 2014; MCROBERTS–WESTFALL, 2014). However, in linear models, the Hessian matrix is null (SAMBAKHE et al., 2014), and in these cases, the first-order uncertainty is trivially sufficient to characterize the model-related uncertainty. Thus, adding second order terms results is more accurate than using only the first order, in nonlinear models. The second order is especially beneficial for models with large uncertainties and strong non-linearities (VON PFINGSTEN et al., 2017).

The bootstrap estimator was used to validate the analytical method used for linear (CUNIA, 1987) and non-linear models (law of error propagation). The results were shown to be exactly the same in the case of the linear model and also with the non-linear models, if a second-order Taylor series had been used. Some papers highlight the idea of replacing some time-consuming and computational efforts of bootstrap by choosing the appropriate order of the Taylor series (GU et al., 2021). We agree with the statement made by Fortin et al. (2018) that for linear models or non-linear models that use a complex system, e.g., systems with more than one

equation, or at a lower hierarchical level, the analytical method can be uniquely used, with the concern of using a concise and accurate hybrid variance estimator.

Biomass presents high uncertainty compared to volume, due to the high natural variability across the landscape (TEMESGEN et al., 2015). In addition, the datasets to fit the model also present a wide range of sites varying in age and site quality, which contributes to variation within the dataset and consequently affects the model parameters. Similar uncertainty amount due to the biomass model was found by Stahl et al. (2014) who compared model and sampling variability in national forest inventories to find that the model-related variability accounted for 28%, 10% and 10% of the total variability from the Norwegian, Finnish, and Swedish, respectively. Cunia (1987) found that percentage to be as high as 34.86% of the total variance in mixed forests of the New York State Forest area. In another study, Djomo et al. (2010) assessed the uncertainty in carbon stock estimates and estimated that the allometric equations contributed 30% of the total uncertainty in stock estimates over different vegetation types in Cameroon, Africa. Chave et al. (2004) found an error greater than 20% due to the model in aboveground biomass estimates in the tropical rainforest of Central Panama.

Regarding the equations fitted with the simulated data, with low fit quality, they presented greater uncertainty in the coefficients. The quality of the fit, besides directly influencing the model-related uncertainty, propagates this uncertainty and increases the sampling-related uncertainty. Estimating tree values for plots with models that do not have a good fit greatly increases the sampling error. Therefore, the choice of model is very important for the overall estimates (PICARD et al., 2015).

Our research takes as a hypothetical situation two equations with inferior fit quality, which contribute more than 50% in the plots with lower variability for volume and more than 70% for biomass. Thus, the plots will have a lower contribution to the total uncertainty when models of low quality. We separated the plots into low, medium, and high variance and studied the behavior of the uncertainty. Even though our case study consisted in a planted forest, there was a noticeable effect of this within plot variability. It is then expected that, when there is very high variability within the plots, this sampling error will be so high that it may be negligible and could have a smaller contribution in the total variance. However, in real situations, it is always expected that a representative sample of the population will reduce this sampling uncertainty. Especially in inventories with small areas, as in the case of plantations, the plots need to be representative because they represent the truth of a population. A good tactic to optimize sampling with environmental heterogeneity is to use a stratified design rather than a simple random design, map structural traits of plant canopies and allow inference of traits and, in many cases, species ranges.

Confidence intervals without considering the uncertainty of volume and biomass models become underestimated. As shown in this study, this could be as a result of a reduced fitting sample, unnecessary model complexity (i.e. nonlinear models) or biased variance estimators (i.e. poor approximations to non linear functions). We can have narrow intervals that lead us to mistaken conclusions that it is a good estimate, while in fact the estimate leads to large

uncertainty (PICARD et al., 2015; SAARELA et al., 2017). We emphasize that ignoring model-related uncertainty in the total error in forestry inventory can lead to erroneous decisions in forest management, with economic implications. Uncertainty analysis is essential to implement management strategies and policy interventions and should be included as part of the estimates of any forest inventory (YANAI et al., 2020).

3.5 CONCLUSION

In this study, we were able to take the uncertainty due to sampling and model together in allometric models for *Acacia mearnsii* stands. We present the behavior and quantification of uncertainty in two variables of extreme importance for forest quantification: volume and biomass. Thus, based on our results we conclude:

Smaller sample sizes could lead to somewhat more model-related uncertainty in the estimates.

However, models respond differently to uncertainty in models with different numbers of trees, with linear models presenting less uncertainty in their parameters with fewer trees to fit the model.

The non-linearity of the model contribute to the presentation of uncertainty and the use of more statistically demanding methods, such as the use of second order.

Higher order Taylor series may exhibit many desirable properties, among which detection and characterization of non-linearity. It is important to investigate procedures to represent exact distributions of input data measurements without any discrimination and which could be very useful for non-linear analysis and comparison with expanded uncertainty.

The uncertainty related to the volume model is generally smaller than the sampling uncertainty. For classes with larger sample sizes the uncertainty of the biomass model is also smaller considered sampling. Even though, not considering the uncertainty of the model makes all the results of the total uncertainty work underestimated, especially in cases where this uncertainty was very large, such as with few numbers of trees and poor fit quality.

3.5 REFERENCES

- ARRAS, K. O. (1998). An introduction to error propagation: derivation, meaning and examples of equation $y = f(x)$. Relatório técnico, ETH Zurich.
- BEHLING, A., PÉLLICO NETTO, S., SANQUETTA, C. R., CORTE, A. P. D., AFFLECK, D. L., RODRIGUES, A. L. e BEHLING, M. (2018). Critical analyses when modeling tree biomass to ensure additivity of its components. *Anais da Academia Brasileira de Ciências*, 90:1759–1774.

- BERGER, A., GSCHWANTNER, T., MCROBERTS, R. E. e SCHADAUER, K. (2014). Effects of measurement errors on individual tree stem volume estimates for the austrian national forest inventory. *Forest Science*, 60(1):14–24.
- BERNSTEIN, L., BOSCH, P., CANZIANI, O., CHEN, Z., CHRIST, R. e RIAHI, K. (2008). Ipcc, 2007: climate change 2007: synthesis report.
- CHAVE, J., CONDIT, R., AGUILAR, S., HERNANDEZ, A., LAO, S. e PEREZ, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1443):409–420.
- CORONA, P., FATTORINI, L., FRANCESCHI, S., SCRINZI, G. e TORRESAN, C. (2014). Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. *Canadian Journal of Forest Research*, 44(11):1303–1311.
- CUNIA, T. (1987). Error of forest inventory estimates: its main components.in: Wharton eh, cunia t (eds) estimating tree biomass regressions and their error.
- FATTORINI, L. (2012). Design-based or model-based inference? the role of hybrid approaches in environmental surveys. *Studies in Honor of Claudio Scala*, páginas 173–214.
- FORTIN, M. (2020). Updating plots to improve the precision of small-area estimates: the example of the lorraine region, france. *Canadian Journal of Forest Research*, 50(7):648–658.
- FORTIN, M., MANSO, R. e SCHNEIDER, R. (2018). Parametric bootstrap estimators for hybrid inference in forest inventories. *Forestry: An International Journal of Forest Research*, 91(3):354–365.
- FORTIN, M., ROBERT, N. e MANSO, R. (2016). Uncertainty assessment of large-scale forest growth predictions based on a transition-matrix model in catalonia. *Annals of Forest Science*, 73(4):871–883.
- FU, Y., LEI, Y., ZENG, W., HAO, R., ZHANG, G., ZHONG, Q. e XU, M. (2017). Uncertainty assessment in aboveground biomass estimation at the regional scale using a new method considering both sampling error and model error. *Canadian Journal of Forest Research*, 47(8):1095–1103.
- GERTNER, G. (1991). Prediction bias and response surface curvature. *Forest Science*, 37(3):755–765.
- GERTNER, G. Z. (1990). The sensitivity of measurement error in stand volume estimation. *Canadian Journal of Forest Research*, 20(6):800–804.

- GU, M.-H., CHO, C., CHU, H.-Y., KANG, N.-W. e LEE, J.-G. (2021). Uncertainty propagation on a nonlinear measurement model based on Taylor expansion. *Measurement and Control*, 54(3-4):209–215.
- GUJARATI, D. N. e PORTER, D. C. (2011). *Essentials of econometrics*. SAGE Publications.
- HEROLD, M., CARTER, S., AVITABILE, V., ESPEJO, A. B., JONCKHEERE, I., LUCAS, R., MCROBERTS, R. E., NÆSSET, E., NIGHTINGALE, J., PETERSEN, R. e OTHERS (2019). The role and need for space-based forest biomass-related measurements in environmental management and policy. *Surveys in Geophysics*, 40(4):757–778.
- JEVŠENAK, J. e SKUDNIK, M. (2021). A random forest model for basal area increment predictions from national forest inventory data. *Forest Ecology and Management*, 479:118601.
- JONAS, M., MARLAND, G., WINIWARTER, W., WHITE, T., NAHORSKI, Z., BUN, R. e NILSSON, S. (2010). Benefits of dealing with uncertainty in greenhouse gas inventories: introduction. Em *Greenhouse gas inventories*, páginas 3–18. Springer.
- KANGAS, A. (1996). On the bias and variance in tree volume predictions due to model and measurement errors. *Scandinavian Journal of Forest Research*, 11(1-4):281–290.
- LAROCQUE, G. R., BHATTI, J. S., BOUTIN, R. e CHERTOV, O. (2008). Uncertainty analysis in carbon cycle models of forest ecosystems: research needs and development of a theoretical framework to estimate error propagation. *Ecological Modelling*, 219(3-4):400–412.
- MAGALHÃES, T. M., SEIFERT, T. e OTHERS (2015). Estimation of tree biomass, carbon stocks, and error propagation in maturated woodlands. *Open Journal of Forestry*, 5(04):471.
- MCROBERTS, R. E. e WESTFALL, J. A. (2014). Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. *Forest Science*, 60(1):34–42.
- MCROBERTS, R. E. e WESTFALL, J. A. (2015). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, 73(3):625–633.
- PARÉ, D., GERTNER, G., BERNIER, P. e YANAI, R. (2016). Quantifying uncertainty in forest measurements and models: approaches and applications. *Canadian Journal of Forest Research*, 46(3):v–v.
- PARRESOL, B. R. (2001). Additivity of nonlinear biomass equations. *Canadian Journal of Forest Research*, 31(5):865–878.
- PICARD, N., BOYEMBA BOSELA, F. e ROSSI, V. (2015). Reducing the error in biomass estimates strongly depends on model selection. *Annals of forest Science*, 72(6):811–823.

- PÉLLICO NETTO, S.; BRENA, D. (1997). Inventário florestal. 1:316.
- SAARELA, S., ANDERSEN, H.-E., GRAFSTRÖM, A., SCHNELL, S., GOBAKKEN, T., NÆSSET, E., NELSON, R. F., MCROBERTS, R. E., GREGOIRE, T. G. e STÅHL, G. (2017). A new prediction-based variance estimator for two-stage model-assisted surveys of forest resources. *Remote Sensing of Environment*, 192:1–11.
- SAARELA, S., SCHNELL, S., GRAFSTRÖM, A., TUOMINEN, S., NORDKVIST, K., HYYPPÄ, J., KANGAS, A. e STÅHL, G. (2015). Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. *Canadian Journal of Forest Research*, 45(11):1524–1534.
- SAARELA, S., WÄSTLUND, A., HOLMSTRÖM, E., MENSAH, A. A., HOLM, S., NILSSON, M., FRIDMAN, J. e STÅHL, G. (2020). Mapping aboveground biomass and its prediction uncertainty using lidar and field data, accounting for tree-level allometric and lidar model errors. *Forest Ecosystems*, 7(1):1–17.
- SAMBAKHE, D., FORTIN, M., RENAUD, J.-P., DELEUZE, C., DREYFUS, P. e PICARD, N. (2014). Prediction bias induced by plot size in forest growth models. *Forest Science*, 60(6):1050–1059.
- SCHUMACHER, F. X. e HALL, F. S. (1933). Logarithmic expression of timber-tree volume. *J Agric Res*, 47:719–734.
- SPURR, S. H. (1952). Forest inventory.
- STÅHL, G., HEIKKINEN, J., PETERSSON, H., REPOLA, J. e HOLM, S. (2014). Sample-based estimation of greenhouse gas emissions from forests—a new approach to account for both sampling and model errors. *Forest Science*, 60(1):3–13.
- TEMESGEN, H., AFFLECK, D., POUDEL, K., GRAY, A. e SESSIONS, J. (2015). A review of the challenges and opportunities in estimating above ground forest biomass using tree-level models. *Scandinavian Journal of Forest Research*, 30(4):326–335.
- TRAUTENMÜLLER, J. W., NETTO, S. P., BALBINOT, R., WATZLAWICK, L. F., DALLA CORTE, A. P., SANQUETTA, C. R. e BEHLING, A. (2021). Regression estimators for aboveground biomass and its constituent parts of trees in native southern brazilian forests. *Ecological Indicators*, 130:108025.
- VON PFINGSTEN, S., BROLL, D. O., VON DER ASSEN, N. e BARDOW, A. (2017). Second-order analytical uncertainty analysis in life cycle assessment. *Environmental Science & Technology*, 51(22):13199–13204.
- WHITE, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, páginas 817–838.

- XU, C. e GERTNER, G. Z. (2008). Uncertainty and sensitivity analysis for models with correlated parameters. *Reliability Engineering & System Safety*, 93(10):1563–1573.
- YANAI, R. D., WAYSON, C., LEE, D., ESPEJO, A., CAMPBELL, J. L., GREEN, M. B., ZUKSWERT, J. M., YOFFE, S., AUKEMA, J., LISTER, A. e OTHERS (2020). Improving uncertainty in forest carbon accounting for redd+ mitigation efforts. *Environmental Research Letters*, 15(12):124002.

4 CHAPTER 3: A CORRECTION FOR ASSESSING UNCERTAINTY IN NON-LINEAR MODELS WITH AN ANALYTICAL APPROACH

ABSTRACT

Accurate uncertainty estimates are essential for determining the reliability of forest information results. Therefore, uncertainty estimators need to be unbiased, impartial, accurate and provide trusted estimates of the uncertainty of forest estimates. We aimed to adapt an analytical hybrid variance estimator, considering the correlation of uncertainty between trees predicted with the same model, for non-linear models. We fitted two non-linear models for the variables volume and biomass for *Acacia mearnsii* stands. We used the second Taylor series order along with the first order to quantify the uncertainty of the models. We also used a correction to the Bootstrap estimator. The hypothesis was proved, and considering the correlation between the plot estimates is necessary to ensure that the analytical hybrid variance estimator is not underestimated. It was evident that when correlations are ignored in the uncertainty matrix, the values are low and incorrect. Our correction in the analytic hybrid estimator generated accurate results, we thus recommend the use of this analytic estimator for assessing uncertainty in non-linear models. The use of the second derivative was necessary to express the model-related uncertainty. Our framework does not need the correction for the bootstrap estimator (BS) to correct the underestimation of uncertainty due to sampling once it was almost the same as the value without using the correction.

Keywords: nonlinear models; sampling-related uncertainty; analytical estimator; volume estimator; biomass quantification.

4.1 INTRODUCTION

Forest-predicted variables such as volume, biomass, and carbon are essential for the characterization and quantification of forest resources and wood stocks (FORTIN et al., 2018; BEHLING et al., 2018; SUN et al., 2022). Information on these variables supports forest management and climate change mitigation initiatives. Some sources of uncertainty are intrinsic to the measurement and prediction process to obtain those variables due to only part of the population being measured in the sampling and estimated with an equation (CUNIA, 1987; MCROBERTS–WESTFALL, 2014; FORTIN et al., 2018; KIRCHNER et al., 2021). Accurate uncertainty estimates are therefore essential for determining the reliability on the uncertainty of forest estimates (FORTIN et al., 2016). Hence, uncertainty estimators need to be impartial and consider the main sources of errors associated with estimates (MELO et al., 2018). They also need to be accurate, so that they provide reliable estimates of the uncertainty of forest estimates (YANAI et al., 2020).

Quantifying sampling- and model-related uncertainty in the forest variables is essential since it can help us identify which issues need to be addressed in order to reduce the uncertainty of the predictions (MELO et al., 2018). However, combining sampling and model uncertainty (parameter and residual error) to calculate the total uncertainty is somewhat challenging, usually

resulting in complex mathematical developments (STÅHL et al., 2016; KANGAS et al., 2018). In this respect, a significant research effort has been devoted over the last 30 years to develop estimators that jointly consider the main sources of uncertainty in forest inventories (CUNIA, 1987; GERTNER, 1990; CHAVE et al., 2004; FORTIN et al., 2016; MCROBERTS–WESTFALL, 2015; FORTIN et al., 2018).

Uncertainty estimators that consider sampling and model uncertainty are known as hybrid inference (FATTORINI, 2012; CORONA et al., 2014). The hybrid estimators widely applied in forestry to analyze uncertainty are analytical estimators and bootstrap estimator, also called bootstrap estimators (CUNIA, 1987; GERTNER, 1990; MCROBERTS–WESTFALL, 2014; FORTIN et al., 2018). Many commonly used models to predict the volume and biomass of trees are non-linear (SPURR, 1952; SCHUMACHER–HALL, 1933). Fortin et al. (2018) suggest using hybrid analytical estimators to estimate the uncertainty of linear and non-linear models. An analytical approach for combining uncertainty sources is easier to implement and avoids the mistake of reporting the wrong property of the distribution of bootstrap estimator trials (YANAI et al., 2020).

Gertner (1990) was one of the first authors to apply an analytical approach to analyze the uncertainties associated with tree volume estimates. The author introduced a methodology for assessing the relative importance of measurement error on the estimated mean of stand volume and the variance of this estimate when a nonlinear regional volume function is used to predict the volume of individual trees from a simple random sample of plots. Gertner's research was, and still is, a major contribution to the uncertainty assessment for nonlinear models in forestry (MOORE, 1991; KANGAS, 1996). However, we detected a potential underestimation of the model-related uncertainty when we used this approach faithfully followed the steps therein in our own research. In fact, in a discussion paper published shortly afterwards Moore (1991) identified a potential problem in that the original approach did not consider, in the error budget, the covariance between plot-level predictions. A closer look showed that also the predictions between trees within the same plot were also assumed to be uncorrelated.

Gertner (1990) remains an elegant exercise in the field of analytical hybrid estimators, and also one that can be straightforwardly be implemented. Motivated then by the curiosity of whether the above assumptions hold, we built this paper under the hypothesis that between-tree and between-plot predictions are correlated, as all of them come from the same model, which is just a realisation of the true model for the variables of interest. Thus, our objective was to test those assumptions and to adapt Gertner's analytical hybrid variance estimator, considering the above-mentioned correlations in a non-linear modelling context.

To achieve our objective, we used the case study of *Acacia mearnsii* planted forests in South Brazil and we fitted non-linear models (SPURR, 1952; SCHUMACHER–HALL, 1933) for the variables volume and biomass. The original and corrected approaches were assessed and then compared with a bootstrap hybrid estimator (FORTIN et al., 2018).

In addition, we used a second-order Taylor expansion in order to linearise the functional form of the models. This necessary intermediate step is typically carried out with the first-order term of the series (GERTNER, 1990; BERGER et al., 2014). Some authors (GU et al., 2021; GERTNER, 1991) have however shown that this may be not enough in some cases with strong non-linearity in the models, and the second-order Taylor expansion has been proposed as a way to circumventing this issue (GERTNER, 1991; SAMBAKHE et al., 2014).

4.2 MATERIAL AND METHODS

4.2.1 Statistical developments

It seems relevant to firstly introduce the basis of Gertner's approach, as well as the terminology used, in some detail. Let, first of all, i and j be the indexes for the tree and the plot, respectively, under a simple random sampling scheme. Let us also assume that there exists a unbiased tree-level volume or biomass model such that $\hat{y}_{ij} = f(\mathbf{x}_{ij}, \hat{\boldsymbol{\beta}})$, where $\hat{y}_{ij}$ is the prediction of tree volume or biomass, $\mathbf{x}_{ij}$ is a vector of covariates, $\hat{\boldsymbol{\beta}}$ is a vector of parameter estimates and $f(\cdot)$ is some non-linear function.

The estimates of the variable of interest (i.e. volume or biomass) per area in a given plot can then be defined as

$$\hat{Y}_i = \frac{1}{A} \sum_j \hat{y}_{ij} \quad (4.1)$$

where A is the area of the plot. The estimate of the mean per area when considering all n plots is

$$\hat{\bar{Y}} = \frac{1}{n} \sum_j \hat{Y}_j \quad (4.2)$$

where Y is the estimated variable and n is the number of trees in plot j . There are two sources of uncertainty in the estimator of the mean in Eq. 4.2. One is the uncertainty due to the sampling, which stems from the estimator itself. The other one is the uncertainty due to the model predictions in Eq. 4.1. GERTNER (1990) introduced estimators of the variance related to both sources of uncertainty that we can denote as $Var(\hat{\bar{Y}})_d$ and $Var(\hat{\bar{Y}})_p$, with subscripts d and p standing for “design” and “model parameters”, respectively. The total variance would simply be the sum of both terms:

$$Var(\hat{\bar{Y}}) = Var(\hat{\bar{Y}})_d + Var(\hat{\bar{Y}})_p \quad (4.3)$$

The uncertainty due to the design was calculated using the variance estimator of the population mean estimator for simple random sampling (HORVITZ–THOMPSON, 1952; PÉLLICO NETTO, 1997).

$$Var(\bar{Y})_d = \frac{\sum_{j=1}^n (\hat{Y}_j - \bar{\hat{Y}})^2}{n-1} \quad (4.4)$$

The developments to obtain $Var(\hat{Y})_p$ are more complex and consist in several steps that are detailed in the following lines. It is also here where our corrections apply.

In the first place, Gertner (1990) approximated the variance of tree-level predictions of the variable of interest as:

$$Var(\hat{y}_{ij}) \approx [\mathbf{z}_{ij}]' [\mathbf{\Omega}] [\mathbf{z}_{ij}] \quad (4.5)$$

where: $[\mathbf{z}_{ij}]$ denotes the vector matrix of the first derivative of the model with respect to the parameter estimates such that $\mathbf{z}_{ij} = \left(\frac{\partial \hat{y}_{ij}}{\partial \hat{\beta}} \right)$, $[\mathbf{z}_{ij}]'$ is its transposed matrix, and $\mathbf{\Omega}$ is the variance-covariance matrix of the parameter estimates:

$$\mathbf{\Omega} = \begin{bmatrix} \sigma_{\beta_1}^2 & \cdots & \sigma_{\beta_1 \beta_n} \\ \vdots & \ddots & \vdots \\ \sigma_{\beta_1 \beta_n} & \cdots & \sigma_{\beta_n}^2 \end{bmatrix}$$

The next step was to calculate the variance of the joint predictions within a plot. In an elegant attempt to simplify the calculations, Gertner (1990) assumed that the elements in $\hat{Y}_j$ must be uncorrelated. Under this assumption, the following equality holds:

$$Var(\hat{Y}_j) = Var\left(\sum_i \hat{y}_{ij} / A\right) = \frac{1}{A^2} \sum_i Var(\hat{y}_{ij}) \quad (4.6)$$

We, on the contrary, believe that this assumption can hardly hold because all the elements in $\hat{Y}_j$ are generated from the same model $\hat{y}_{ij} = f(\mathbf{x}_{ij}, \hat{\beta})$, which may be unbiased globally but must almost certainly be biased locally.

The assumption of independence is in fact not needed if the covariance between the predictions is considered:

$$Var(\hat{Y}_j) = \frac{1}{A^2} \left(\sum_i Var(\hat{y}_{ij}) + \sum_i \sum_{i' \neq i} Cov(\hat{y}_{ij}, \hat{y}_{i'j}) \right) \quad (4.7)$$

The elements of Eq. 4.7 can be obtained by extending Eq. 4.5 for all trees in a plot:

$$Var(\hat{\mathbf{y}}_j) = [\mathbf{Z}_j]' [\mathbf{\Omega}] [\mathbf{Z}_j] \quad (4.8)$$

where $\hat{\mathbf{y}}_j$ is the vector with all predictions for a plot and $\mathbf{Z}_j$ is the matrix that includes all $\mathbf{z}_{ij}$ vectors. The resulting matrix, in extended form, is

$$Var(\hat{\mathbf{y}}_j) = \begin{bmatrix} Var(\hat{y}_{1j}) & Cov(\hat{y}_{1j}\hat{y}_{2j}) & \cdots & Cov(\hat{y}_{1j}\hat{y}_{nj}) \\ Cov(\hat{y}_{1j}\hat{y}_{2j}) & Var(\hat{y}_{2j}) & \cdots & Cov(\hat{y}_{2j}\hat{y}_{nj}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{y}_{1j}\hat{y}_{nj}) & Cov(\hat{y}_{2j}\hat{y}_{nj}) & \cdots & Var(\hat{y}_{nj}) \end{bmatrix}$$

The diagonal elements correspond to the variance of each tree-level prediction (first term in the sum in Eq. 4.7) and the off-diagonal elements are the covariance between predictions.

It only remains to obtain the variance of the mean estimate ($\hat{\bar{Y}}$). To do this, GERTNER (1990) resorted to the same simplification as above:

$$Var(\hat{\bar{Y}})_p = Var\left(\sum_j^n \hat{Y}_j / n\right) \approx \frac{1}{n^2} \sum_j^n Var(\hat{Y}_j) \quad (4.9)$$

Again, the predictions of the variable of interest between plots do not need to be uncorrelated. The complete expression for this estimator must include the covariance between the plot-level predictions:

$$Var(\hat{\bar{Y}})'_p = Var\left(\sum_j \hat{Y}_j / n\right) = \frac{1}{n^2} \left(\sum_j^n Var(\hat{Y}_j)' + \sum_j^n \sum_{j' \neq j}^n Cov(\hat{Y}_j \hat{Y}_{j'}) \right) \quad (4.10)$$

Note that we already considered our corrected term $Var(\hat{Y}_j)'$ in Eq. 4.10. The challenge is to calculate the covariance term. To illustrate how this can be derived, we use first the example of a simple linear model to then generalise it to the non-linear case.

Let $\hat{Y}_j$ and $\hat{Y}_{j'}$ be the predictions of the variable of interest for plots j and j' , respectively, and let us assume, for the sake of the example, that the individual tree-level model $\hat{y}_{ij} = \hat{\gamma}_0 + \hat{\gamma}_1 x_{ij}$ applies. $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are estimates of the true parameters γ_0 and γ_1 . The covariance between the plot-level predictions of plots j and j' will be defined as

$$Cov(\hat{Y}_j, \hat{Y}_{j'}) = \left(\sum_i^{m_j} \hat{\gamma}_0 + \hat{\gamma}_1 x_{ij} - \sum_i^{m_j} \gamma_0 + \gamma_1 x_{ij} \right) \left(\sum_i^{m_{j'}} \hat{\gamma}_0 + \hat{\gamma}_1 x_{ij'} - \sum_i^{m_{j'}} \gamma_0 + \gamma_1 x_{ij'} \right) \quad (4.11)$$

with m_j and $m_{j'}$ the number of trees in plots j and j' , respectively.

Rearranging, one gets

$$Cov(\hat{Y}_j, \hat{Y}_{j'}) = \left(m_j(\hat{\gamma}_0 - \gamma_0) + (\hat{\gamma}_1 - \gamma_1) \sum_i^{m_j} x_{ij} \right) \left(m_{j'}(\hat{\gamma}_0 - \gamma_0) + (\hat{\gamma}_1 - \gamma_1) \sum_i^{m_{j'}} x_{ij'} \right) \quad (4.12)$$

where the terms $(\hat{\gamma}_0 - \gamma_0)$ and $(\hat{\gamma}_1 - \gamma_1)$ are in fact the standard errors of the parameter estimates, which can be replaced by the more convenient terms ϵ_0 and ϵ_1 .

Further rearrangement leads to

$$Cov(\hat{Y}_j, \hat{Y}_{j'}) = m^2 \epsilon_0^2 + m_j \epsilon_0 \epsilon_1 \sum_i^{m_{j'}} x_{ij'} + m_{j'} \epsilon_0 \epsilon_1 \sum_i^{m_j} x_{ij} + \epsilon_1^2 \sum_i^{m_j} x_{ij} \sum_i^{m_{j'}} x_{ij'} \quad (4.13)$$

It is easy to see that ϵ_0^2 , ϵ_1^2 , and $\epsilon_0 \epsilon_1$ are the variances and the covariance of the parameter estimates, which can be used to fully define the variance-covariance matrix Ψ . If the equation above is set in matrix form, we obtain

$$Cov(\hat{Y}_j, \hat{Y}_{j'}) = \sum_i^{m_j} \mathbf{x}_{ij} \Psi \sum_i^{m_{j'}} \mathbf{x}_{ij'}^T \quad (4.14)$$

where $\mathbf{x}_{ij}$ and $\mathbf{x}_{ij'}$ are the generic vectors of covariates for each tree (in this specific case $\mathbf{x}_{ij} = [1, x_{ij}]$ and $\mathbf{x}_{ij'} = [1, x_{ij'}]$).

In a non-linear context, these vectors are typically approximated with the matrix of the first derivatives of the model with respect to the parameter estimates. Coming back to our general rationale, these Jacobian matrices are those already defined in Eq. 4.5. The covariance between plot-level predictions for any pair of plots in the general case would then be

$$Cov(\hat{Y}_j, \hat{Y}_{j'}) \approx \sum_i^{m_j} \mathbf{z}_{ij} \Omega \sum_i^{m_{j'}} \mathbf{z}_{ij'}^T \quad (4.15)$$

which can be substituted in Eq. 4.10 to yield our corrected estimator of the variance of the mean related to the uncertainty of the parameter estimates:

$$Var(\hat{Y})'_p = Var(\sum \hat{Y}_j / n) \approx \frac{1}{n^2} \left(\sum_j^n Var(\hat{Y}_j)' + \sum_j^n \sum_{j' \neq j}^n \left(\sum_i \mathbf{z}_{ij} \Omega \sum_i \mathbf{z}_{ij'}^T \right) \right) \quad (4.16)$$

4.2.2 Second order Taylor series approximations

The approximations in Eqs. 4.5 and 4.15 in the previous section involve the matrix of first derivatives of the model with respect to the parameter estimates. This is done in a non-linear framework as a way of linearising the functional form of the model through, typically, a first-order Taylor expansion. Although this is normally considered to be a good approximation, some authors have reported that a second-order Taylor expansion may be necessary in some cases (GU et al., 2021; GERTNER, 1991). As an additional development, we took on board this second-degree expansion in our corrected estimator (Eq. 4.16) by extending Eq. 4.5.

If the Jacobian matrix $\mathbf{z}_{ij}$ is renamed as $\mathbf{J}_{ij}$ and we define the Hessian matrix including the second partial derivatives of the model with respect to the parameters as $\mathbf{H}_{ij}$, the extended tree-level variance of the predictions will be

$$Var(\hat{y}_{ij}) \approx [\mathbf{J}_{ij}]' [\mathbf{\Omega}] [\mathbf{J}_{ij}] + \frac{1}{2} Tr([\mathbf{H}_{ij}] [\mathbf{\Omega}])^2 \quad (4.17)$$

4.2.3 Case study

In order to test the correction introduced in the previous section, we used the case study of a forest inventory conducted in stands of *Acacia mearnsii* in Rio Grande do Sul state (South Brazil).

4.2.3.1 Data and models used

The studied forests were divided in three regions: Cristal, Encruzilhada do Sul and Piratini. In each region four stands were sampled in an age sequence after planting to cover the entire rotation (10 years). In Cristal the stands were sampled at 1.75, 2.75, 5.00 and 10.08 years, in Encruzilhada do Sul at: 1.83, 3.08, 5.75 and 10.75 years and in Piratini at 1, 2.33, 5.25 and 9.83 years. Data collection consisted of two phases: (i) plot measurement, which we called phase 1, and (ii) subplot measurement, phase 2.

Phase 1: Measurement of the variables in the plots.

In each stand four circular plots were randomly installed, with a diameter of 22.56 m (400 m²), totaling 48 temporary plots. In these plots 3,462 trees were measured in the diameter at breast height and total height, which were used to quantify the biomass stocks of the components and the stem volume, by adjusted regression models. Diameter at breast height was measured using a dendrometer tape and total height with a hypsometer (Haglöf).

Phase 2: Measurement of variables in the subplots.

A 10 m diameter subplot (78.54 m²) was established in the center of each plot, totaling 48 subplots. This was done because it was not possible to measure all variables of all trees in the larger plot, due to the high cost. Twig and crown biomass, crown length, total volume, and crown volume were measured on 670 trees.

All trees were felled and evaluated for their characteristics: diameter at breast height was measured using a dendrometer tape and total crown height and length with a tape measure. Crown volume was calculated using the Huber method and crown scale was obtained using the Hohenadl method. The stem volume (with bark) was obtained using the Huber method and its scale was made using the Hohenadl method. Measurements were taken with a dendrometric tape along the stem at positions of 5%, 15%, 25%, 35%, 45%, 55%, 65%, 75%, 85%, and 95% of the total height.

Biomass measurements were taken for the stem (stem wood + bark) and crown (living branches and dead leaves, flowers and fruits). For each tree, these components were separated and weighed to determine green biomass with a digital scale (Portable Electronic Scale) with an accuracy of 5 g.

To determine the dry biomass of the stem and crown, samples were collected and immediately weighed using a digital balance (Hoyle) with an accuracy of 1 g. From the stem, 5 disks (2 cm thick) were collected at the positions 0%, 25%, 50%, 75%, and 95% along the total height. The samples were dried in an oven with air circulation at 100°C and after a constant weight the material was weighed on a digital scale with an accuracy of 1 g.

We fitted SPURR (1952) and SCHUMACHER–HALL (1933) models, in non-linear fashion, with the database from de phase 2 (670 trees), to predict volume and biomass:

$$\hat{y} = \beta_1 d^2 h^{\beta_2} \quad (4.18)$$

$$\hat{y} = \beta_1 d^{\beta_2} h^{\beta_3} \quad (4.19)$$

where d and h are diameter at breast height (1.30 m off the ground) and tree height, respectively. The β terms are parameters that were estimated by ordinary least squares. The White's test (1980) was applied to evaluate the hypothesis of homogeneity of residuals of the model and, when heteroscedasticity was observed, we chose to model the residual variance to establish weights and readjust the equation by means of the weighted least squares method, constructing the weight matrix according to PARRESOL (2001).

4.2.3.2 Calculations

We firstly followed the approach in GERTNER (1990) step by step using our data and models. We then applied the corrections developed in the previous sections. This allowed us to come up with estimates of the mean volume and above ground biomass in our case study, and the design-related and model-related variance of this estimator, through both methods and to compare the results. The first and second partial derivatives used to calculate the second-order Taylor expansion are shown below

Model	First derivative	Second Derivative
Spurr	$f(\beta_1) = d^2 h^{\beta_2}$	$f(\beta_1) = 0$
	$f(\beta_2) = \beta_1 \cdot d^2 h^{\beta_2} \cdot \log_{d^2 h}$	$f(\beta_2) = \beta_1 \cdot d^2 h^{\beta_2} \cdot \log_{d^2 h} \cdot \log_{d^2 h}$
Schumacher-Hall	$f(\beta_1) = d^{\beta_2} \cdot h^{\beta_3}$	$f(\beta_1) = 0$
	$f(\beta_2) = \beta_1 \cdot d^{\beta_2} \cdot h^{\beta_3} \cdot \log_d$	$f(\beta_2) = \beta_1 \cdot d^{\beta_2} \cdot h^{\beta_3} \cdot \log_d \cdot \log_d$
	$f(\beta_3) = \beta_1 \cdot d^{\beta_2} \cdot h^{\beta_3} \cdot \log_h$	$f(\beta_3) = \beta_1 \cdot d^{\beta_2} \cdot h^{\beta_3} \cdot \log_h \cdot \log_h$

Because we do not have proof that our corrections completely correct the bias in the model-related uncertainty that we believe exists in Gertner (1990), we applied a well-established bootstrap estimator that produces unbiased estimates of the total variance. This method was introduced by Fortin et al. (2018) for the population total and further developed for the population

mean by Melo et al. (2019). We followed the latter here and compared the results with those of our own correction. The steps are listed below:

1. For a given realization, random deviations are generated from the matrix $\mathbf{\Omega}$ to account for the model-related uncertainty. The random deviation value is added to the model coefficients.
2. Values of volume and above-ground biomass in the plots are obtained $\hat{Y}_b$ with coefficients generated in the previous step.
3. We ran the simulations for B realizations.
4. The results are fed into equation 4.20

$$Var(\hat{Y}) = \frac{\sum (\hat{Y}_b - \hat{Y}_{BS})^2}{B} + 2Var(\hat{Y}_{\bar{y}_j}) - \frac{\sum \widehat{Var_d} [\hat{Y}_b]}{B} \quad (4.20)$$

Where subindex b is assigned to the realization, and $\hat{Y}_{BS}$ is the estimator of the mean in the bootstrap method (i.e. over all realizations), and B is the total number of realizations, in our case 10.000. The second term in equation 4.20 can be obtained as

$$Var(\hat{Y}_{\bar{y}_j}) = \frac{\sum_{j=1}^n (\hat{y}_j - \hat{Y}_{BS})^2}{n(n-1)} \quad (4.21)$$

where $\hat{y}_j = \sum_{b=1}^B \frac{\hat{y}_{j,b}}{B}$.

4.3 RESULTS

The models presented good results regarding the quality of fit (TABLE 4.1). The statistics were similar for the models and for each variable.

TABLE 4.1 – GOODNESS OF FIT FOR THE VOLUME AND BIOMASS MODELS.

Model	Equation	$R^2_{adjust.}$	AIC	$S_{yx}\%$	CV%
Volume					
Spurr	$\hat{v} = 0.0000567 \cdot d^2 h^{0.9520817}$	0.989	-5227.849	10.58	101.94
Schumacher-Hall	$\hat{v} = 0.00005759 \cdot d^{1.913} \cdot h^{0.9380}$	0.989	-5226.125	10.59	101.94
Biomass					
Spurr	$\hat{v} = 0.02498 \cdot d^2 h^{1.00770}$	0.970	4364.44	18.17	105.21
Schumacher-Hall	$\hat{v} = 0.03078 \cdot d^{2.12057} \cdot h^{0.83503}$	0.971	4353.24	18.00	105.04

When the covariance between the predictions was ignored, as in the original approach, the uncertainty due to the model was many orders of magnitude lower than when our correction was applied. The use of the bootstrap estimator correction yielded very similar results to our

corrected analytical approach, which confirms the fact that our developments are sound (Table 4.2). The uncertainty due to the design was very similar regardless of the method used, as expected.

The different models produced similar results for the total uncertainty. In relative terms, for the volume, the contribution to the total uncertainty due to the model and sampling in the Spurr model was 4.15% and 95.85% respectively. For biomass, this contribution was 9.82% for the model and 90.18% for sampling. For the Schumacher-Hall model, this contribution was 4.56% and 95.44% for the model and sampling for volume. For biomass, it was 18.94% for the model and 81.06% for sampling 4.2.

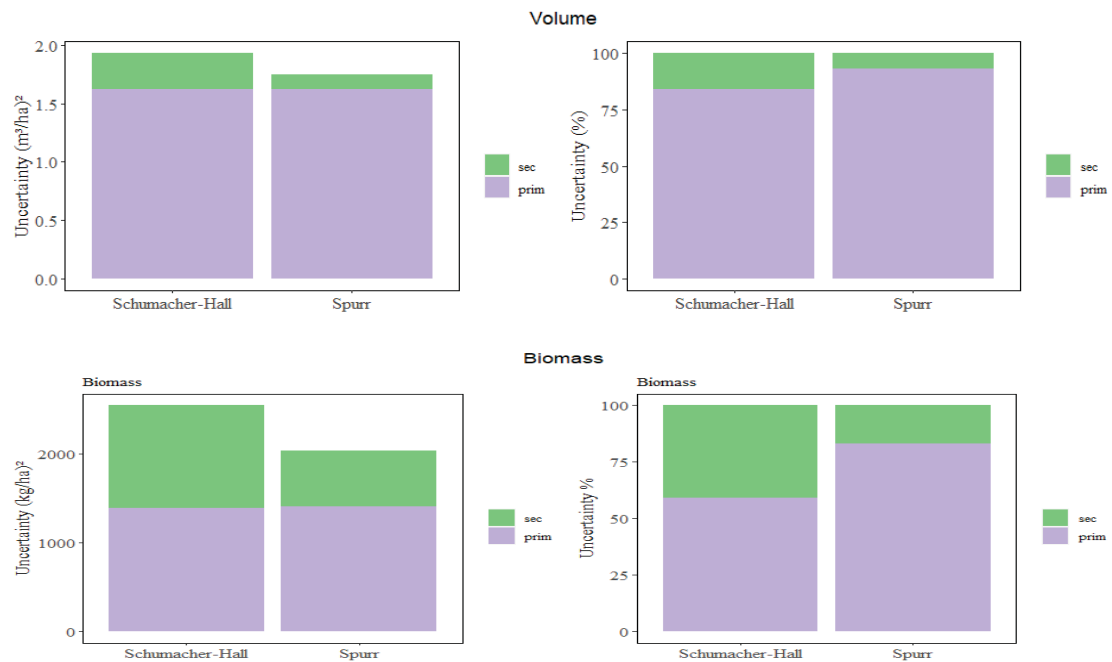
TABLE 4.2 – SAMPLING AND MODEL CONTRIBUTION TO THE TOTAL UNCERTAINTY WITH THE CORRECTION IN THE ANALYTICAL APPROACH AND BOOTSTRAP ESTIMATOR.

	Sampling		Model			
	Analytical	BS	BS+C	A-no covariance	A-covariance	BS
Volume ($m^3 \cdot ha^{-1}$) ²						
Schumacher	40.41	40.33	40.35	0.0001566	1.93	1.78
Spurr	40.37	40.33	40.39	0.00011524	1.7498	1.6724
Biomass $kg \cdot ha^{-1}$						
Schumacher	4.695,94	4.679,373	4.713,952	20.75	2.111,24	2.074,34
Spurr	4.665,46	4.665,46	4.665,66	14.30	1.540,30	1.540,30

Where: BS+C is bootstrap estimator with correction, A-no covariance is analytical approach with no covariance, A-covariance is analytical approach considering covariance

The contribution of the second derivative to the model-related uncertainty was smaller in the Spurr model than in the Schumacher-Hall model (FIGURE 4.1). The second derivative contributed with 7% and 16% of the uncertainty explanation of the Spurr model for volume and biomass, respectively. For the Schumacher-Hall model, this derivative contributed with 16% and 40% for volume and biomass.

FIGURE 4.1 – CONTRIBUTION OF EACH DERIVATIVE FOR THE MODEL-RELATED UNCERTAINTY FOR VOLUME AND BIOMASS ESTIMATES.



SOURCE: The author.

4.4 DISCUSSION

In this study we calculated model- and sampling-related uncertainty jointly, seeking more precise estimators to quantify the total uncertainty in the volume and biomass predicted by non-linear models. Our analysis of model-related uncertainty is based on a correction to the analytical methodology presented by Gertner (1990), in which we consider the correlation between tree-level predictions and plot-level predictions, as they all come from the same model. Additionally, we applied the second derivative of the Taylor series to avoid issues with the lack of precision in the linearisation of non-linear models. The fact our corrected approach performed similarly to the bootstrap estimator by Fortin et al. (2018) confirms that our approach is correct. The main merit of this paper is then to have revisited a classic and yet useful approach to assess uncertainty in forest inventory, and to having made it available without the original bias.

The correlation between the tree-level and plot-level predictions exists due to the fact that the same model was used on different plots. When these predictions are assumed to be independent Gertner (1990), the model-related uncertainty is underestimated. In this respect, Lehtonen et al. (2007) state that biomass model errors and/or stem volume model errors are in fact correlated and because trees from the same plot resemble to each other more than to trees that belong to different plots, and confirm that ignoring these mainly positive correlations leads to error bounds that indeed are narrower than they should.

Other analytical approaches exist to unbiasedly calculate the uncertainty around volume or biomass estimates e.g. Cunia (1987). It is still very useful though to refer to GERTNER (1990) as an approach to quantify the uncertainty due to the parameter estimates, especially in a nonlinear modelling framework, given its wide application in forest sciences (KANGAS, 1996; LEHTONEN et al., 2007; BERGER et al., 2014). In their discussion paper, Moore (1990) already hinted at an "*a posteriori*" solution to correct for at least part of the bias that Gertner's approach may have been incurred. Essentially, the author states that a bias must exist for each plot's predictions, and that if the variation of this bias can be quantified, then it should be included in the total uncertainty.

Although our approach is motivated by the same problem as pointed out by Moore (1990), our development was radically different. We demonstrated that the correlation among predictions has to be considered not only at the plot level, but also at the tree level. More fundamentally, our changes are embedded into the existing mathematical developments and are not just a correction applied to resulting total variance. Despite this, our approach simply is not difficult to implement, with just the variance-covariance matrix of the parameter estimates and the Jacobian matrices of the models with respect to the parameters (and the Hessian if the second-order approximation is considered) as the only elements to handle.

As a result of Schumacher-Hall's higher degree of nonlinearity, the proportion of the second derivative is higher compared to Spurr's to express model-related uncertainty. Second derivatives are more accurate at capturing the curvature of the response of a model (GERTNER, 1991). This curvature of the response surface is the amount by which a curve deviates from being a straight line, or a surface deviates from being a plane (LI, 2015), which is fully affected by the nonlinearity of the model. Additionally, it is common parameters to be correlated with each other (XU–GERTNER, 2008), depending on the degree of nonlinearity (GUJARATI–PORTER, 2011). The correlation between parameter estimates is also part of the uncertainty linked to these estimates through the variance-covariance matrix of the parameters and, therefore, can affect the magnitude of uncertainty.

The first-order uncertainty is quite enough to characterize the uncertainty in the volume model. However, for biomass, the use of the Taylor expansion to derive expressions for uncertainties means that we are estimating a dominant part with the second derivative. This was due to the substantial uncertainty of the biomass equation, characteristic of the variable biomass (CUNIA, 1987; CHAVE et al., 2004; PICARD et al., 2015) and the second derivative managed to capture this uncertainty. The desirable characteristics of the second derivative, such as the detection and characterization of non-linearities, among which are robustness to additive uncertainty, helped to accurately characterize this uncertainty (GU et al., 2021).

We relied on the bootstrap estimator by Fortin et al. (2018) and its extension in Melo et al. (2019) as the reference to compare our corrected analytical estimator, instead of running a simulation study. Fortin et al. (2018) showed how a naive bootstrap estimator would

underestimate the variance and then introduced a correction that deals with this bias. The correction itself was tested in different simulation studies and proved successful in all of them.

With the analytical and bootstrap estimators, it was possible to distinguish the variance related to the model and sampling. We identified that, in our case study, the most significant source of variability was sampling. Our homogeneous forest types of Acacia stands and the appropriate representativeness of the database set to fit the models led us to expect this outcome. The results are in accordance with other volume and biomass research (CUNIA, 1987; GERTNER, 1990; KANGAS, 1996; CHAVE et al., 2004; DJOMO et al., 2010; BERGER et al., 2014; MCROBERTS–WESTFALL, 2014, 2015; SAARELA et al., 2015; PICARD et al., 2015).

4.5 CONCLUSION

Based on the results, we draw the following conclusions:

i) Our hypothesis was proved, and considering the correlation between tree-level and plot-level predictions is necessary to ensure that Gertner's analytical hybrid estimator does not underestimate the total variance.

ii) Our correction in the analytic hybrid estimator generated accurate results, thus the use of this analytic estimator for assessing uncertainty in forest inventories where the variable of interest is predicted is good. This correction is valid for linear and non-linear modelling frameworks.

iii) The effect of the second derivative was non-negligible to assess the model-related uncertainty.

iv) The largest source of error in our case study was the sampling, due to the structure of our inventory. The estimates of the model parameters are however non-negligible and should therefore be considered as a source of uncertainty. This can be done through hybrid estimators so that the total uncertainty is not underestimated.

We focus on correcting only the Gertner (1990) estimator for the model. We did not correct the methodology for measurement error, where the author comments that he also considers measurement errors independently.

4.5 REFERENCES

- BEHLING, A., PÉLLICO NETTO, S., SANQUETTA, C. R., CORTE, A. P. D., AFFLECK, D. L., RODRIGUES, A. L. e BEHLING, M. (2018). Critical analyses when modeling tree biomass to ensure additivity of its components. *Anais da Academia Brasileira de Ciências*, 90:1759–1774.
- BERGER, A., GSCHWANTNER, T., MCROBERTS, R. E. e SCHADAUER, K. (2014). Effects of measurement errors on individual tree stem volume estimates for the austrian national forest inventory. *Forest Science*, 60(1):14–24.

- CHAVE, J., CONDIT, R., AGUILAR, S., HERNANDEZ, A., LAO, S. e PEREZ, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1443):409–420.
- CORONA, P., FATTORINI, L., FRANCESCHI, S., SCRINZI, G. e TORRESAN, C. (2014). Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. *Canadian Journal of Forest Research*, 44(11):1303–1311.
- CUNIA, T. (1987). Error of forest inventory estimates: its main components.in: Wharton eh, cunia t (eds) estimating tree biomass regressions and their error.
- DJOMO, A. N., IBRAHIMA, A., SABOROWSKI, J. e GRAVENHORST, G. (2010). Allometric equations for biomass estimations in cameroon and pan moist tropical equations including biomass data from africa. *Forest Ecology and Management*, 260(10):1873–1885.
- FATTORINI, L. (2012). Design-based or model-based inference? the role of hybrid approaches in environmental surveys. *Studies in Honor of Claudio Scala*, páginas 173–214.
- FORTIN, M., MANSO, R. e SCHNEIDER, R. (2018). Parametric bootstrap estimators for hybrid inference in forest inventories. *Forestry: An International Journal of Forest Research*, 91(3):354–365.
- FORTIN, M., ROBERT, N. e MANSO, R. (2016). Uncertainty assessment of large-scale forest growth predictions based on a transition-matrix model in catalonia. *Annals of Forest Science*, 73(4):871–883.
- GERTNER, G. (1991). Prediction bias and response surface curvature. *Forest Science*, 37(3):755–765.
- GERTNER, G. Z. (1990). The sensitivity of measurement error in stand volume estimation. *Canadian Journal of Forest Research*, 20(6):800–804.
- GU, M.-H., CHO, C., CHU, H.-Y., KANG, N.-W. e LEE, J.-G. (2021). Uncertainty propagation on a nonlinear measurement model based on taylor expansion. *Measurement and Control*, 54(3-4):209–215.
- GUJARATI, D. N. e PORTER, D. C. (2011). *Essentials of econometrics*. SAGE Publications.
- HORVITZ, D. G. e THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685.
- KANGAS, A. (1996). On the bias and variance in tree volume predictions due to model and measurement errors. *Scandinavian Journal of Forest Research*, 11(1-4):281–290.

- KANGAS, A., KORHONEN, K. T., PACKALEN, T. e VAUHKONEN, J. (2018). Sources and types of uncertainties in the information on forest-related ecosystem services. *Forest Ecology and Management*, 427:7–16.
- KIRCHNER, M., MITTER, H., SCHNEIDER, U. A., SOMMER, M., FALKNER, K. e SCHMID, E. (2021). Uncertainty concepts for integrated modeling-review and application for identifying uncertainties and uncertainty propagation pathways. *Environmental Modelling & Software*, 135:104905.
- LEHTONEN, A., CIENCIALA, E., TATARINOV, F. e MÄKIPÄÄ, R. (2007). Uncertainty estimation of biomass expansion factors for norway spruce in the czech republic. *Annals of Forest Science*, 64(2):133–140.
- LI, X. (2015). Curvature of a geometric surface and curvature of gravity and magnetic anomalies. *Geophysics*, 80(1):G15–G26.
- MCROBERTS, R. E. e WESTFALL, J. A. (2014). Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. *Forest Science*, 60(1):34–42.
- MCROBERTS, R. E. e WESTFALL, J. A. (2015). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, 73(3):625–633.
- MELO, L. C., SCHNEIDER, R. e FORTIN, M. (2018). Estimating model-and sampling-related uncertainty in large-area growth predictions. *Ecological Modelling*, 390:62–69.
- MOORE, R. (1991). The sensitivity of measurement error in stand volume estimation: Discussion. *Canadian Journal of Forest Research*, 21(8):1296–1296.
- PARRESOL, B. R. (2001). Additivity of nonlinear biomass equations. *Canadian Journal of Forest Research*, 31(5):865–878.
- PICARD, N., BOYEMBA BOSELA, F. e ROSSI, V. (2015). Reducing the error in biomass estimates strongly depends on model selection. *Annals of forest Science*, 72(6):811–823.
- PÉLLICO NETTO, S.; BRENA, D. (1997). Inventário florestal. 1:316.
- SAARELA, S., SCHNELL, S., GRAFSTRÖM, A., TUOMINEN, S., NORDKVIST, K., HYYPPÄ, J., KANGAS, A. e STÅHL, G. (2015). Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. *Canadian Journal of Forest Research*, 45(11):1524–1534.
- SAMBAKHE, D., FORTIN, M., RENAUD, J.-P., DELEUZE, C., DREYFUS, P. e PICARD, N. (2014). Prediction bias induced by plot size in forest growth models. *Forest Science*, 60(6):1050–1059.

- SCHUMACHER, F. X. e HALL, F. S. (1933). Logarithmic expression of timber-tree volume. *J Agric Res*, 47:719–734.
- SPURR, S. H. (1952). Forest inventory.
- STÅHL, G., SAARELA, S., SCHNELL, S., HOLM, S., BREIDENBACH, J., HEALEY, S. P., PATTERSON, P. L., MAGNUSSEN, S., NÆSSET, E., MCROBERTS, R. E. e OTHERS (2016). Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. *Forest Ecosystems*, 3(1):1–11.
- SUN, Z., QIAN, W., HUANG, Q., LV, H., YU, D., OU, Q., LU, H. e TANG, X. (2022). Use remote sensing and machine learning to study the changes of broad-leaved forest biomass and their climate driving forces in nature reserves of northern subtropics. *Remote Sensing*, 14(5):1066.
- XU, C. e GERTNER, G. Z. (2008). Uncertainty and sensitivity analysis for models with correlated parameters. *Reliability Engineering & System Safety*, 93(10):1563–1573.
- YANAI, R. D., WAYSON, C., LEE, D., ESPEJO, A., CAMPBELL, J. L., GREEN, M. B., ZUKSWERT, J. M., YOFFE, S., AUKEMA, J., LISTER, A. e OTHERS (2020). Improving uncertainty in forest carbon accounting for redd+ mitigation efforts. *Environmental Research Letters*, 15(12):124002.

5 CHAPTER 4: HOW DO DIFFERENT SOURCES OF UNCERTAINTIES AFFECT VOLUME AND BIOMASS ESTIMATES?

ABSTRACT

Estimates of population parameters in forestry are fraught with uncertainties that should always be assessed and reported so that the accuracy of the results is known and understood. The aims of this study were (i) to quantify the total uncertainty on volume and biomass estimates and the share of this total due to different sources of uncertainty, such as design-related uncertainty, modelling-related uncertainty – including that emerging from an embedded height-diameter model– and that due to measurement errors, and to (ii) explore the economic implications of ignoring uncertainty in volume and biomass estimates. For these purposes, we used the case study of *Acacia mearnsii* stands in the South of Brazil. To evaluate how the above uncertainty propagates, we defined five scenarios where the different sources of uncertainty are progressively and incrementally added. Scenarios 1, 2, and 3 focus on the impact of height models on the uncertainties of volume and biomass estimates. We assume that in these scenarios the explanatory variables dbh and h were measured error free. In scenarios 4 and 5, in addition to continuing with the effect of the height equation, we added systematic and random errors on dbh in plots. We fitted Spurr model in linear fashion with mixed-effects to predict individual tree volume and biomass. The greatest uncertainty occurred when we used all predicted heights, both to fit the model and to estimate the volume and biomass in the plots. The systematic error in the inventory plots led to the second largest uncertainty. In the third place is the uncertainty due to random errors in the explanatory variables used in volume and biomass models. One of the main conclusions of this study is that uncertainty sources interact in different ways and are propagated in the final uncertainty calculations in terms of volume and biomass. These sources of uncertainty may be more significant in the model or inventory, but they directly influence the total uncertainty. Understanding the effect of multiple sources of uncertainty in any estimate helps identify the most effective way to direct monitoring effort to improve confidence in those estimates.

Keywords: error propagation chain; predicted height; measurement errors; modelling; sampling.

5.1 INTRODUCTION

Quantification of forest stock plays a key role in global international agreements on carbon cycles and climate change mitigation (STÅHL et al., 2014; SAARELA et al., 2020; FORTIN, 2020). The two most popular attributes used for this quantification are tree (timber) volume, which is key in forest management and planning (FORTIN–DEBLOIS, 2010; HEROLD et al., 2019; BEHLING et al., 2018) and tree biomass (TRAUTENMÜLLER et al., 2021), which provides support for climate and environmental policies (HEROLD et al., 2019). Additionally, these variables can be used to estimate carbon stocks, a difficult variable to assess. Despite the promising advances in remote sensing to directly measure volume and biomass (SAARELA et al., 2015, 2017), these variables are typically predicted through regression models based on easy to measure variables such as diameter at breast height (1.30 m; dbh) and tree height. This is

usually done in a set of sample plots, and subsequently the resulting plot-level predictions are up-scaled to obtain population-level estimates.

The workflow described above leads to two major and widely reported sources of uncertainty: sampling and modelling (CUNIA, 1987; GERTNER, 1990; STÅHL et al., 2014; MCROBERTS–WESTFALL, 2014; FORTIN et al., 2018). Sampling-related uncertainty is due to the intrinsic variance between sample units, which depends on landscape-level heterogeneity (CUNIA, 1987; PICARD et al., 2015). Structural inventory elements such as sample size, sample selection, plot structure, and different type of estimators can be used to handle between-plot variability (CUNIA, 1987; CHAVE et al., 2004; GRAFSTRÖM–RINGVALL, 2013; PICARD et al., 2015; MCROBERTS–WESTFALL, 2015), but they cannot eliminate it, because natural variability is present in the environment and the forest data is a sampling, not a census. Model-related uncertainty is due to regressions describing tree allometry, and consists of two components: the estimates of model parameters (uncertainty in the expected value) and the residual variability around the expected value (CUNIA, 1987; KANGAS, 1996; MCROBERTS–WESTFALL, 2015; PICARD et al., 2015; FORTIN et al., 2018).

Model-related uncertainty is not restricted to volume and biomass models but to any intermediary model used in the process. One example is height-diameter models. Height measurements used in many volume or biomass models, e.g. (SPURR, 1952; SCHUMACHER–HALL, 1933), are not always available for all trees (MAGNUSSEN et al., 2020; DE PETRIS et al., 2022) because measuring heights is a time-consuming job that can be particularly difficult in closed canopy stands (FORTIN–DEBLOIS, 2010). In fact, height is usually measured only on a subsample of trees, and these measurements serve to fit height-diameter models, which are in turn used to predict the heights of unsampled trees from their dbhs (GARBER et al., 2009; FORTIN–DEBLOIS, 2010; MAGNUSSEN et al., 2020). Models that take height into account to predict volume may overall improve the precision of the predictions (SULLIVAN et al., 2018), but substituting predicted heights for actual heights also adds uncertainty to the system. However, the impact of this uncertainty on the model-related uncertainty is typically ignored. (GARBER et al., 2009; FORTIN–DEBLOIS, 2010; KEARSLEY et al., 2013; MAGNUSSEN et al., 2020).

Uncertainty however begins to be propagated from the very first step in any conventional forest inventory. This is down to the notion that no physical quantity can be measured with total certainty (MONTGOMERY, 2020). Associated with this is the fact that measurement errors may occur both due to the precision of instruments and the skills of different operators in the field (KANGAS, 1996; BERGER et al., 2014). These errors are carried over to the input variables either fed to the models to obtain predictions or directly used to fit the models, propagating this way to the population-level estimates.

Uncertainty should always be assessed and reported in order to establish the significance of any results and to make predictions with a known level of confidence (LAROCQUE et al., 2008; PARÉ et al., 2016). The issues covered above need to be considered in order to identify their individual impact on the precision of the population-level estimates and to design

strategies to minimise or mitigate these impacts (PICARD et al., 2015; KIRCHNER et al., 2021). Underestimating or indeed failing to consider uncertainty in decision making could lead to suboptimal management and poor financial choices (PARÉ et al., 2016).

In an attempt to demonstrate the relevance of these concepts, we present in this paper a case study based on a conventional forest inventory in *Acacia mearnsii* stands in Southern Brazil. This forest inventory conducted in a systematic sampling design where the dbh of all trees within the selected sample plots was measured. A volume model and a biomass model that depend on dbh and tree height, plus a height-diameter model, were fitted to a subsample of the data. We then built an array of scenarios to test how population-level predictions of volume and biomass and their model and design-associated uncertainty are affected by

1. The uncertainty linked to height predictions, when these replace heights in plots to predict volume and biomass (scenarios 1 and 2), and are used to fit volume and biomass models (scenario 3)
2. Adding measurement errors to the explanatory variables used to fit the volume and biomass models, and the height model (scenarios 4.1, 4.2 and 4.3)
3. Adding measurement errors to the input variables of the volume and biomass models to predict volume and biomass, respectively (scenarios 5.1 and 5.2)

In addition to that, we evaluated what are the economic consequences of not considering the uncertainty from these scenarios in an exercise where we used current timber prices for *Acacia mearnsii*.

The aims of this study were: i) to quantify different sources of uncertainty in volume and biomass estimates, ii) to evaluate the effect of propagating the uncertainty in height predictions and measurement errors on the total uncertainty population-level on volume and biomass estimates; iii) to explore the economic implications of ignoring uncertainty in these estimates. We projected the objectives supported by the first hypothesis of the study on what are the sources of uncertainty, described above in the different scenarios, affect uncertainty due to sampling and the model in different ways along the uncertainty propagation chain. The second hypothesis is that results reported without adequate uncertainty analysis influence the economic limits of timber and may generate losses when average estimates per hectare underestimate uncertainty.

5.2 MATERIAL AND METHODS

5.2.1 Data and Sampling design

The studied forests were divided in three regions: Cristal, Encruzilhada do Sul and Piratini. In each region four stands of *Acacia mearnsii* were sampled in an age sequence after planting to cover the entire rotation (10 years). In Cristal the stands were sampled at 1.75, 2.75,

5.00 and 10.08 years, in Encruzilhada do Sul at: 1.83, 3.08, 5.75 and 10.75 years and in Piratini at 1, 2.33, 5.25 and 9.83 years.

In each stand, four circular 22.56-m diameter plots of 400 m² were randomly set up, totaling 94 temporary plots. In these plots, the dbh (d-cm) of all trees was measured, totaling 6,867 trees and the total height (h-m) was measured in 20% of the trees in each plot, totaling 2,489 trees with the height measured. These data were used to analyze scenario 1, which will be described later on.

A 10-m diameter subplot (78.54 m²) was established in the center of 48 plots where biomass from branches and crown, and total volume were measured for 670 trees. This subsampling scheme was conducted because it was not possible to measure all variables from all trees in the larger plot, due to the high cost. Towards that end, all trees in the subplots were felled and the following measurements taken: height, total crown height, crown length, diameter at breast height, and the diameter of the cross-sections at 5%, 15%, 25%, 35%, 45%, 55%, 65%, 75%, 85%, and 95% of the total height, using the Hohenadl method. The stem volume was obtained using the Huber method.

Biomass measurements were taken for the stem (stem wood + bark) and crown (living branches and dead leaves, flowers, and fruits). For each tree, these components were separated and weighed with a digital scale (Portable Electronic Scale) with an accuracy of 5 g to determine green biomass. The samples were collected aiming at to calculate stem and crown biomass and to weight them in situ using a digital balance (Hoyle) with an accuracy of 1 g. From the stem, five 2-cm discs were extracted at the cross-sections corresponding to 0%, 25%, 50%, 75%, and 95% of the total height. The samples were dried in an air circulation-assisted oven at 100°C until attaining constant weight, and the material was weighed on a digital scale with an accuracy of 1 g.

5.2.2 Volume and biomass modelling

We fitted the linear version of the Spurr model (Equation 5.1) to predict individual tree volume and biomass. We used for these purposes mixed-effects modelling techniques to attend the assumption of independence, since a within-stand correlation of the residuals is to be expected as a result of the experimental design and the effect of unknown local variables. We used forest stands of different ages as a random effect, which was included in the intercept of the model.

$$y = \beta_0 + u + \beta_1 d^2 h + \epsilon_y \quad (5.1)$$

where y is the variable of interest – volume or biomass –, β_0 and β_1 are fixed-effect parameters, u is the stand random effect, which is assumed normally distributed such that $u \sim N(0, \sigma_u^2)$, where σ_u^2 is the variance of the random effects, and ϵ is a tree-level residual error term, also normally distributed such that $\epsilon_y \sim N(0, \sigma_y^2)$, with σ_y^2 the residual variance.

When heteroscedasticity was observed, we fitted the model by means of the weighted least squares method. The matrix of weights was calculated as presented in PARRESOL (2001) and used by BEHLING et al. (2018).

5.2.3 Height-diameter model

Height predictions were obtained through the model introduced by NÄSLUND (1936) (Equation 5.2). We fitted this model to data from the 670 trees collected in the subplots through least squares. All regression assumptions were met, including independence, which was checked through preliminary tests that showed no correlation at the stand/age.

$$\frac{d^2}{h - 1.3} = \gamma_0 + \gamma_1 d + \gamma_2 d^2 + \epsilon_h \quad (5.2)$$

where h is height, d is the diameter at breast height, γ_0, γ_1 and γ_2 are the model parameters and ϵ_h is the error term, assumed normally distributed such that $\epsilon_h \sim N(0, \sigma_h^2)$, with σ_h^2 the residual variance.

5.2.4 Analysis of uncertainty in different scenarios

To evaluate the uncertainty propagation chain, we defined five separate scenarios that are described throughout this section. We focused here on modelling-related uncertainty (scenarios 1, 2, 3 and 4), including that from the embedded height-diameter model, and the uncertainty due to measurement errors. Design-related uncertainty is jointly evaluated in our analysis, we did consider design-based in scenario 5.

For comparison purposes, we also calculated the total uncertainty without considering the uncertainty due to the height model and measurement errors, which is called scenario 0.

5.2.4.1 Scenarios 1, 2 and 3

These scenarios focus on the impact of the uncertainty linked to the predicted heights. This uncertainty arises from the parameter estimates. We assumed that the explanatory variables d and h were measured error-free in these scenarios.

Scenario 1: From each plot, 20% of the heights fed into the volume and biomass models were measured heights, whereas 80% of them were stochastically predicted through equation 5.2.

Scenario 2: All tree heights in plots used as explanatory variable in the volume and biomass models were stochastically predicted through equation 5.2.

Scenario 3: This scenario aims to explore the effect of using stochastically predicted heights to fit volume and biomass models. In other words, we used the tree actual diameters from the subplots and the predicted heights of these trees as the models' input variables. If we refer to these predicted heights as $h_{predicted}$ and they are substituted into equation 5.1, we obtain:

$$y = \beta_0 + u + \beta_1 d^2 h_{predicted} + \epsilon_y \quad (5.3)$$

5.2.4.2 Scenario 4

In this scenario, we aimed to test the impact of fitting our height, volume and biomass models to data where the explanatory variables have been added measurement errors. Towards this aim, we simulated random errors in d and h , and created three sub-scenarios to study the specific effect on each model. The errors were in principle generated from independent normal distributions such that

$$\begin{aligned} \xi_d &\sim N(0, \sigma_d^2) \\ \xi_h &\sim N(0, \sigma_h^2) \end{aligned}$$

where ξ_d and ξ_h are random measurement errors for diameter and height, and σ_d^2 and σ_h^2 are the variance of the empirical distributions of the observed diameters and heights. In an attempt to limit the extent of the errors within realistic boundaries, only values that fell within the interval $[-0.05\bar{d}, 0.05\bar{d}]$ and $[-0.05\bar{h}, 0.05\bar{h}]$ were retained, which is effectively equivalent to sampling a truncated normal distribution whose notation we omit here for simplicity. The modified diameters and heights were defined as $d_{al} = d + \xi_d$ and $h_{al} = h + \xi_h$. When it comes to using the model to produce predictions, the d from the plot dataset were assumed to be error-free, and all the heights were stochastically predicted.

Subscenario 4.1: Both d_{al} and h_{al} are substituted in 5.2:

$$y = \beta_0 + u + \beta_1 d_{al}^2 h_{al} + \epsilon_y \quad (5.4)$$

Subscenario 4.2: Only d_{al} is substituted in 5.2:

$$y = \beta_0 + u + \beta_1 d_{al}^2 h + \epsilon_y \quad (5.5)$$

Subscenario 4.3: This scenario is similar to 4.1 but the effect of the measurement error is also considered to fit the height model:

$$\frac{d_{al}^2}{h_{me} - 1.3} = \beta_0 + \beta_1 d_{al} + \beta_2 d_{al}^2 + \epsilon_h \quad (5.6)$$

5.2.4.3 Scenario 5

This scenario aimed to evaluate the influence of measurement errors added to input variables fed into the models to predict volume and biomass (as opposed to Scenario 4, where we evaluated the effect of the measurement errors on model fit). We focused on this exercise on dbh

alone, thus all tree heights were considered free of measurement errors and were stochastically predicted with equation 5.2. Two subscenarios were considered where the measurement errors were assumed to be random and systematic, respectively.

Subscenario 5.1: We considered random errors such that

$$\xi_{dp} \sim N(0, \sigma_{dp}^2)$$

where σ_{dp}^2 is the variance of the empirical distribution of the observed diameters. Only values that fell within the interval $[-0.05\bar{d}, 0.05\bar{d}]$ were retained, which is effectively equivalent to sampling a truncated normal distribution whose notation we omit here for simplicity.

Subscenario 5.2: In this case, a systematic error of 5% of the dbh was added to each tree dbh. The error is always positive (i.e. a positive bias).

A summary description of each scenario is shown in Table 5.1.

TABLE 5.1 – SUMMARY DESCRIPTIONS IN EACH SCENARIO.

Scenario 1		
Height model influence on volume and biomass estimates, predicting 80% trees heights in each plot		
Models		Design
Height: equation 2	Errors-free measured dbh	
Volume and Biomass: equation 1	20% of measured heights in each plot	
	80% of the predicted heights with height equation 2 for each variable.	
Scenario 2		
Height model influence on volume and biomass estimates by predicting all tree heights in the plots		
Models		Design
Height: equation 2	Error-free measured dbh	
Volume and biomass: equation 1	Predicted heights with eq 2	
Scenario 3		
Predicted heights to fit volume and biomass models + height model influence on predicting all tree heights in the plots		
Model		Design
Height: equation 2	Error-free measured dbh	
Volume and biomass: equation 3	Predicted heights with equation 2	
Scenario 4		
Influence of random measurement errors on the input variables in models of height, volume, and biomass		
Subscenario 4.1		
Random measurement errors on the dbh and h in database to fit volume and biomass models		
Model		Design
Height: equation 2	Error-free measured dbh	
Volume and biomass: equation 4	Predicted heights with equation 2	
Subscenario 4.2		
Random measurement errors on the dbh in database to fit volume and biomass models		
Model		Design
Height: equation 2	Error-free measured dbh	
Volume e biomass: equation 5	Predicted heights with equation 2	
Subscenario 4.3		
Random measurement errors on dbh in data base to fit height, volume, and biomass models		
Model		Design
Height: equation 6	Error-free measured dbh	
Volume e biomass: equation 5	Predicted heights with equation 2	
Scenario 5		
Influence of systematic and random measurement errors on plots tree diameters		
Subscenario 5.1		
Random measurement errors on plots		
Model		Design
Height: equation 2	Random error of up to 5% of the dbh	
Volume and biomass: equation 1	All predicted heights with equation 2.	
Subscenario 5.2		
Systematic measurement errors on plots		
Model		Design
Height: equation 2	Systematic error of 5% of the dbh added to the dbh	
Volume and biomass: equation 1	All predicted heights with equation 2.	

5.2.5 Uncertainty assessment

We used the bootstrap estimator procedure to assess the uncertainty in population-level volume and biomass estimates. We ran simulations for each scenario, each based on 10,000 realizations ($B=10,000$). For all scenarios we essentially followed the same steps, which only differed in the sources of uncertainty considered in different scenarios. To assess the uncertainty with BS estimator sequential steps are followed:

1. For a given realization, random deviations are generated to account for the model-related uncertainty. This always included the uncertainty due to the parameter estimates through their variance-covariance matrix plus the variance of the random effects. Depending on the scenario random draws were also taken from the distribution of the parameters of the height model and/or the distributions used to generate measurement errors.
 - (a) Scenario 1 & 2: The random deviations are added to the estimated coefficients and random effects of the height, volume and biomass models, as for scenario specification.
 - (b) Scenario 3: The random deviations are first added to the estimated coefficients of the height model. Heights are then predicted stochastically and used to re-fit the volume and biomass models. Random deviations of the resulting variance-covariance matrix are added to the new parameter estimates for these models.
 - (c) Scenario 4: Generated measurement errors are added to their respective variables and these used to re-fit the respective models, as for scenario specifications. Random deviations of the resulting variance-covariance matrix are added to the new parameter estimates for these models.
 - (d) Scenario 5: Generated measurement errors are added to their respective variables and these used as input variables to the models for prediction purposes. The random deviations are added to the estimated coefficients and random effects of the height, volume and biomass models.
2. Predictions of volume and aboveground biomass in the plots are obtained from the models using the generated parameters and modified variables according to (ii).
3. We ran the simulations B times.
4. To ascertain the contribution of each source of uncertainty to the total uncertainty we used the hybrid estimator:

$$Var(\hat{Y}) = \frac{\sum (\hat{Y}_b - \hat{Y}_{BS})^2}{B} + 2Var(\hat{Y}_{\hat{y}_j}) - \frac{\sum \widehat{Var_d} [\hat{Y}_b]}{B} \quad (5.7)$$

Where subindex b is assigned to the realization, and $\hat{Y}_{BS}$ is estimator of the mean in the bootstrap method (i.e. over all realizations), and B is the total number of realizations, in our case 10.000. The second term in equation 4.20 can be obtained as

$$Var(\hat{Y}_{\bar{y}_j}) = \frac{\sum_{j=1}^n (\hat{y}_j - \hat{Y}_{BS})^2}{n(n-1)} \quad (5.8)$$

where $\hat{y}_j = \sum_{b=1}^B \frac{\hat{y}_{j,b}}{B}$.

here Y is the variable of interest, $\hat{V}(\hat{Y}_B)$ is the total variance, $\hat{Y}_b$ is an estimate of the mean volume or biomass for each realization, $\hat{Y}_{BS}$ is the mean for all realizations, $\hat{V}_d(\hat{Y}_b)$ is the contribution of the sampling-related uncertainty, B is the number of realization, in our case 10.000, and b is the subindex for the realisation.

5.2.6 Economic analysis

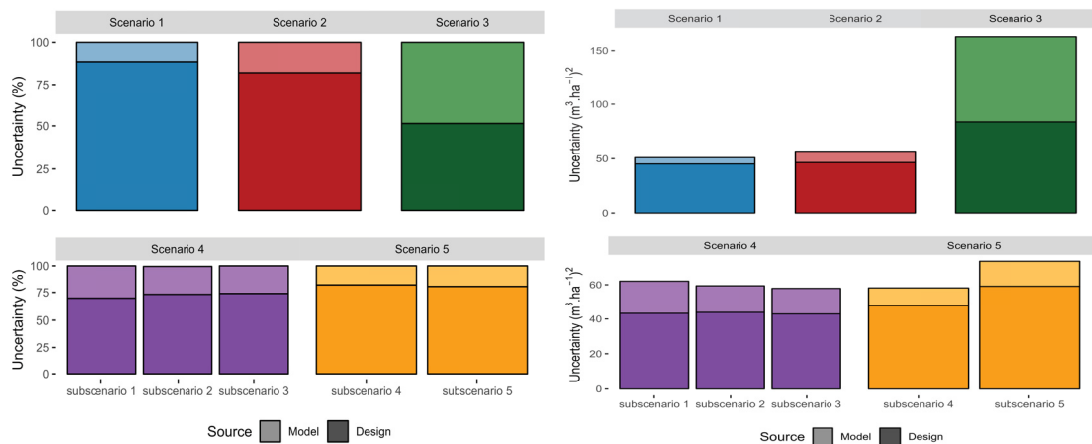
For the purposes of the economic evaluation, we only analyzed the volume per hectare as the variable of interest because it is the most relevant commercial output. After quantifying the volume, and the uncertainty in the different scenarios, we calculated the approximate price of standing timber per unit area, to visualize the consequence of not considering the sources of uncertainty in volume estimates.

According to companies that deal with black wattle wood, the approximate value of a 1-meter stereo of wood volume (mst) is R\$ 180,00. Following this as the only reference, and given that our predictions are in m^3 , we assumed that $1 m^3 = 1.3 mst$ for the sake of the exercise. As a form of comparison between the scenarios, the differences in the economic value between them were analyzed, taking into account the uncertainty calculated in the previous sections.

5.3 RESULTS

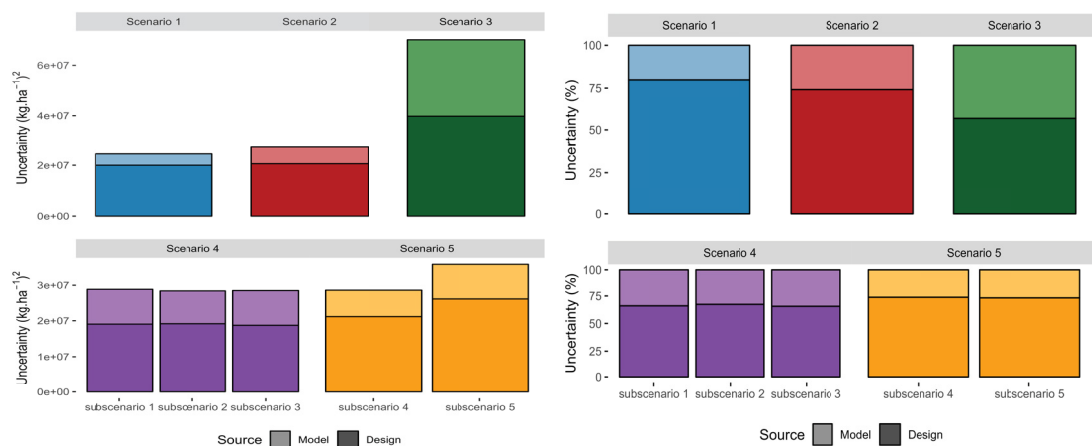
The absolute and relative contribution of the sampling and the model to the total variance is shown in figure 5.1 for volume and in figure 5.2 for biomass. In this context, when the uncertainty related to the height model is not considered as in scenario 0 (e.g. when we use deterministic heights), the uncertainty related to the model and the sampling is $4.081 m^3.ha^{-1}$, $42.71 m^3.ha^{-1}$, respectively in terms of volume. In the case of biomass, these values are $2.107,1 kg.ha^{-1}$ and $4.457,4 kg.ha^{-1}$, respectively. Scenarios 1, 2 and 3 exhibited higher total uncertainty in comparison with this alternative, which showed that the height model-related uncertainty has a noticeable effect on the process and total uncertainty.

FIGURE 5.1 – ABSOLUTE AND RELATIVE CONTRIBUTION OF THE TOTAL UNCERTAINTY DUE TO THE MODEL- AND SAMPLING-RELATED UNCERTAINTY IN VOLUME ESTIMATES FOR EACH SCENARIO.



SOURCE: The author.

FIGURE 5.2 – ABSOLUTE AND RELATIVE CONTRIBUTION OF THE TOTAL UNCERTAINTY DUE TO THE MODEL- AND SAMPLING-RELATED UNCERTAINTY IN BIOMASS ESTIMATES FOR EACH SCENARIO.



SOURCE: The author.

This influence is stronger on model-related uncertainty for scenarios 1 and 2, which exhibited an increase, while sampling-related uncertainty remained almost the same. The volume model-related uncertainty increased from $4.081 \text{ m}^3.\text{ha}^{-1}$ to $5.84 \text{ m}^3.\text{ha}^{-1}$ in scenario 1, whereas in scenario 2 doubled and reached $10.16 \text{ m}^3.\text{ha}^{-1}$. Sampling-related uncertainty marginally increased from $42.71 \text{ m}^3.\text{ha}^{-1}$ to $45.07 \text{ m}^3.\text{ha}^{-1}$ and $46.38 \text{ m}^3.\text{ha}^{-1}$, for scenarios 1 and 2 respectively. In relative terms, total uncertainty increased by 8.8% for Scenario 1 and 20.8% for Scenario 2. For biomass, the relative total uncertainty rose to 3.06% and 17.17%, for scenarios 1 and 2 respectively. The volume model-related variances accounted for 11.47% and 17.97% of the total variance, for scenarios 1 and 2, and the sampling-related uncertainty accounted for 88.52% and 79.80%. Biomass model-related uncertainty accounted for 20.20% and 25.83%, for

scenarios 1 and 2 respectively. Sampling-related uncertainty accounted for 79.80% and 74.16% of the total biomass, respectively.

Thus, the total uncertainty of these estimates decreased when we measured some tree heights in plots and use them as an input variable to predict volume and biomass. In the case study, 20% of the trees measured in each plot resulted in 36% of the total trees sampled in all plots. As a result, the uncertainty propagated by the height equation has decreased, as approximately 64% of tree heights were estimated to be used as an input variable to estimate volume and biomass. When we estimated the tree heights in the plots with the diameter and height relationship equation, a 10% increase in the total uncertainty of the estimates occurs.

Using predicted heights as input variables to fit volume and biomass models greatly increased uncertainty. Scenario 3 presented the highest uncertainty, compared with the other scenarios. In proportion, the model-related variance represented 48.58% of the total variance and sampling-related uncertainty 51.41%. For biomass, the model-related uncertainty was 43.2% and sampling uncertainty contributed 56.79% to the total uncertainty. The model-related uncertainty almost represented the same contribution for the total uncertainty as sampling.

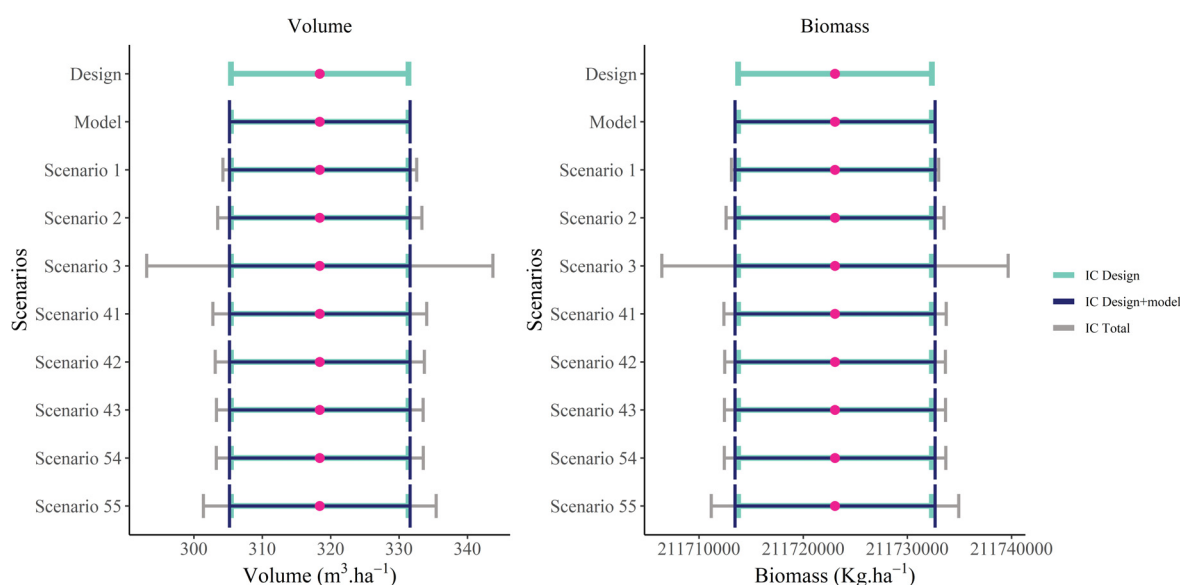
When analyzing the measurement error of the input variables to fit volume and biomass models in scenario 4, sub-scenario 4.1 with measured error in dbh and h, they influence more the parameters. The model-related uncertainty represented 30% of the total uncertainty for volume and 34% for biomass. Compared to scenario 2, as it considers the height of the equation and all the estimated trees, the increase in total uncertainty in volume estimates was 9% and the increase in model-related uncertainty was 15%. In biomass estimates, it was 4% to the total and 3% in model-related uncertainty. However, when considering measurement error just in dbh, it is also high, reaching 26% for model-related uncertainty contribution to the total in volume estimates and 33% for biomass estimates. Considering dbh with a random error to fit the height model, the uncertainty in the volume estimates is similar. The model-related contribution was 26%, the same observed in scenario 4.3. For biomass, although the three subscenarios remained with the same value of absolute uncertainty, however in the subscenario 4.3 the percentage of uncertainty due to the parameters was higher than in the other subscenarios, reaching 33.99%, 32.6%, and 34.38% for scenarios 4.1, 4.2 and 4.3, respectively.

Results for scenario 5 showed that systematic error in dbh of the trees leads to a higher error than random error for allometric variables. The contribution of each source of errors to the total uncertainty is practically the same for systematic and random errors. For volume, sampling-related uncertainty is approximately 81% and model-related uncertainty was 19%. For biomass, sampling-related uncertainty was around 73% due to sampling and 27% due to the model.

Confidence intervals (CI) showed that both allometric variables followed the same trend concerning the behavior of uncertainty underestimation because only considers the sampling-related uncertainty. In figure 5.3 we have the confidence intervals of all the scenarios in comparison to only sampling uncertainty and this plus the model, not considering the height

model and measurement errors. In all scenarios, the ranges are larger than considering only the design and the model. Scenario 3 presented the largest limits of confidence intervals. The higher the total uncertainty associated with the estimates, the larger will be the confidence intervals. For the three subscenarios in scenario 4, the CIs length were the same. As for sub-scenario 5, the systematic error generates a larger CI.

FIGURE 5.3 – CONFIDENCE INTERVALS FOR ALL SCENARIOS.

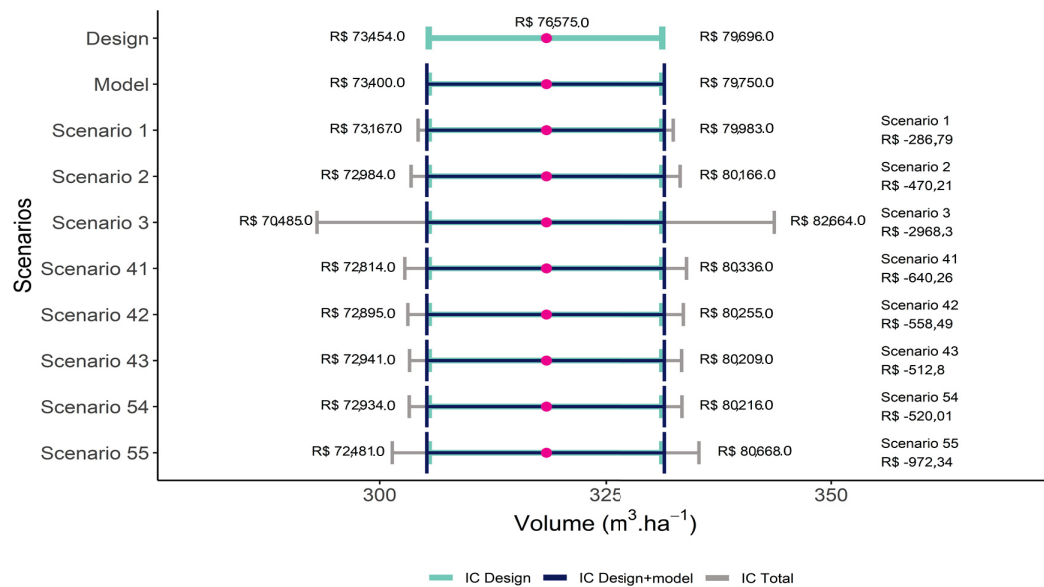


SOURCE: The author.

The mean economic value of the estimated stocking of standing timber is R\$ 76,575.00, with a range of about $\pm$ R\$ 3,000.00 if the uncertainty from the design and the volume model only is considered. The smallest difference in the value limits with the above basic situation is in Scenario 1 and this value almost R\$ 290,00. Scenario 3 presents a much higher difference in the value limit with almost R\$ 3,000.00.

When we assume possible errors of the input variables for the adjustment of the model, the three sub-scenarios in Scenario 4 present a difference in value limits above R\$ 500,00. Already when the errors are in the sampling plots, the sub-scenario with the systematic error shows that this economical difference of up to R\$972,00. For the random error, the difference reaches up to R\$ 520,00.

FIGURE 5.4 – ECONOMICAL DIFFERENT AMONG THE SCENARIOS.



SOURCE: The author.

5.4 DISCUSSION

We presented the error propagation pathways and quantified how much each source influences the total uncertainty in volume and biomass estimates. Both the predicted height uncertainty and measurement errors in input variables influence sampling- and model-related uncertainty of volume and biomass estimates in *Acacia mearnsii* stands. These uncertainty sources impact sampling and models differently, depending on where they fall in the uncertainty propagation chain. After separating these influences into five different scenarios, it became clearer which factors influence and where each source inserts uncertainty in the large error chain of volume and biomass estimates.

The large increase in allometric variables uncertainties when more tree heights are predicted in the plots and when heights are predicted to enter as an input variable in the volume and biomass model highlights the strong impact of tree height predictions on uncertainties in volume and biomass estimates. Furthermore, tree height-related errors influence model- and sampling-related uncertainty depending on the scale at which tree height prediction is applied (MAGNUSSEN et al., 2020). This claim is based on the variation in uncertainty values among scenarios. As all tree heights in plots are predicted, the model-related uncertainty increased considerably while the sampling uncertainty did not. While all tree heights are predicted, including the dataset to fit the model, the uncertainty in both the model and sampling increased. As a consequence, errors associated with height predictions in sample plots propagate to the stand level along with sampling and volume and biomass model errors (FORTIN-DEBLOIS, 2010; MAGNUSSEN et al., 2020).

Our findings contribute to launching a discussion concerning the number of tree heights counted in forest inventory plots. The total uncertainty decreased when some trees in plots were measured and used as an input variable in the volume and biomass equation. This occurs because when all tree heights are predicted in the plot, it increases the number of trees with an error associated. Therefore, all uncertainty related to the predicted height enters into the volume and biomass estimates. Thus, the accuracy of volume and biomass in the plots and then estimated in the stand increases when measuring some tree heights in each plot (FORTIN–DEBLOIS, 2010).

However, measuring tree heights in the field is a difficult task as it requires extensive fieldwork, is time-consuming, and is one of the most expensive data to collect in the forest inventory (JURJEVIĆ et al., 2020). Hence the financial difference between scenarios 1 and 2 for the timber price thresholds in this case study is up to R\$ 183,00. Even if a greater number of trees are measured in plots, financial and time expenditures can still be greater than the value of this difference. Therefore, it is necessary to assess whether it is economically worthwhile to collect height, but never fail to report this source of uncertainty correctly. Besides, we agree with Magnussen et al. (2020) and therefore encourage planners of forest surveys to re-evaluate current guidelines concerning the number and type of selected trees for height measurement.

The largest error propagation resulted when predicted heights were included as an input variable to fit the models. Modeling volume and biomass using predicted heights instead of observed ones at least triples the uncertainty, and model-related uncertainty reaches 50%. In this scenario, the uncertainty about the estimates was greater even than in other scenarios, in spite of addition to the uncertainty related to height prediction, which presents the uncertainty of measurement errors. According to Fortin-DeBlois (2010), when the volume model uses predicted heights, the two errors are closely associated. Uncertainty propagates through models when the model for the interest variable is fitted with an input variable already predicted. Therefore, the influence of residual and parameter-related uncertainty of the model applied to predict the input variable merges with uncertainty related to the variable of interest model (KIRCHNER et al., 2021). This fact doubled the sampling-related uncertainty. Hence, this error is propagated for both model and sampling. This result may be affected due to the fact that the refitted volume/biomass models are making more variability in the plot predictions than they were with the original one.

In this research, we used simulated case studies. However, a common practice is to consider the predicted heights as if they were observed to be used in the model fit (MAGNUSSEN et al., 2020; DE PETRIS et al., 2022). Consequently, the uncertainty in volume and biomass estimates is largely underestimated because they include only the basic components of volume and biomass (FORTIN–DEBLOIS, 2010). The over-optimistic confidence intervals may be erroneously narrow. From an economic perspective, the wood amount represented by an underestimated confidence interval may not be the actual amount, and this will affect the sale prices for wood. In the context of the case study, financial differences can reach up to R\$ 3,000.00.

We assessed the uncertainty related to the height model directly in the final volume and biomass estimates. Therefore, the model- and sampling-related uncertainties increased the uncertainty associated with the predicted heights. Uncertainty analysis in these components in volume and biomass estimates is complex because it involves two different predicted errors, which are related to height and volume or biomass (FORTIN–DEBLOIS, 2010; KIRCHNER et al., 2021). We used a Bootstrap estimator approach to deal with uncertainty to be able to take on board the design-based uncertainty jointly with the model-based uncertainty. Analytical approaches have however been developed to assess the model uncertainty, separating that originating from the parameters of the volume/biomass models and that related to the embedded height model (FORTIN–DEBLOIS, 2010). Furthermore, there is still room for improvement of methodologies that distinguish the exact contribution of height model parameter uncertainties to the total uncertainty of volume and biomass estimates. Further research on uncertainty assessment in volume and biomass models that do not use height as an input variable and in the description of variables per hectare should be explored.

The dbh associated with the random error in the height equation reduces uncertainty because height estimates are based on the same dbh as volume and biomass models. Error tendencies are the same for predicted heights. Since our errors were fully simulated, we acknowledge that measurement errors of actual height may have a more complex structure than in our simulations. Measurement errors in tree height tend to be larger than those of dbh in fieldwork (COLBERT et al., 2002), however, measurement errors in tree height might not affect individual-tree models (QIN et al., 2019). Our results are in line with this statement, especially for the biomass variable which showed less uncertainty due to these measurement errors. Kangas (1996) showed that different measurement-error assumptions of predictor variables affected the regression coefficients of diameter-growth models. The uncertainty caused by measurement errors to fit models should not be overlooked and it is incorrect to assume that certain variables have no error or have an insignificant effect on outputs (QIN et al., 2019).

Random errors contribute more to the model fit than in plots. Our results are in accordance with Qin et al. (2019) which studied measurement error in biomass with data collected for control. The relative uncertainty caused by measurement error in sampling increased from 0.18% to 0.90%, while the relative uncertainty caused by measurement error in the dataset for model fitting increased from 1.89% to 8.6%. Random errors in the model subset and the sample plots propagate in different ways, thus this is a comprehensive comparison. In the case of random errors in the plots, they affect the estimates of individual tree biomass and volume, as well as the estimates of area. When they are added to the model input variables, these errors also interact with parameter uncertainty, increasing this source of error. We noticed the influence of this error on model-related uncertainty, which contributed to the total uncertainty (QIN et al., 2019).

The importance of the effect of measurement error has been studied by some authors (GERTNER, 1990; KANGAS, 1996; MÄKINEN et al., 2010; BERGER et al., 2014;

MCROBERTS–WESTFALL, 2015). This is a difficult source of uncertainty to consider because in operational inventories it is assumed that data collection is error-free. Therefore, with no control data recorded in the field, it is impossible to know whether there are systematic or random measurement errors (DE LERA GARRIDO et al., 2020). However, as the control inventory involves more information collection (QIN et al., 2019), it also involves more time and expenses. We use simulation to have an idea of the propagation of measurement errors. The simulations make it possible to understand how much a measurement error may be interfering with the uncertainty results (MÄKINEN et al., 2010).

The systematic errors associated with dbh plots cause higher uncertainty than random errors. Therefore, the financial difference in the confidence intervals for the average was almost 1,000R\$ per hectare. There is a need for a method to assess the systematic errors of inventories, verify, and try to eliminate distortions that cause systematic errors. This research presented a systematic error of 5% of the diameter of each tree. Therefore, larger trees will receive a more substantial amount of error, related to the proportion of the tree. This is a common feature in the propagation of measurement errors, both systematic errors and random errors (MÄKINEN et al., 2010; CANAVAN–HANN, 2004).

Comparing the two variables under study, volume and biomass, the scenarios followed the same behavior. However, biomass presents more uncertainty in the model than volume in all situations. Biomass presents less influence on the propagation of uncertainty associated with height estimates. Measurement errors related to height also influence biomass less than volume, in agreeing with the statement of Qin et al. (2019) that measurement errors in tree height might not affect models of individual-tree AGB.

We emphasize that our findings are naturally limited to *Acacia mearnsii* stands in experimental conditions with the value of timber being sold at that price in a single company. Financial values will vary for other species and forest types. Also, the effect of uncertainty sources may be different depending on the forest, the variables included in the models (DE LERA GARRIDO et al., 2020), inventory types, and models (CHAVE et al., 2004; PICARD et al., 2015). Our research gives a broad introduction to the possibilities of economic variation at the limits of confidence intervals for the mean. Finally, under this assumption of erroneous financial data and by the aspect of separation between uncertainty we emphasize that uncertainty should not be ignored, not only for the increased precision and accuracy but also for the economic risks of not considering the mean that may be outside the financial limits, consequently a financial loss.

5.5 CONCLUSION

The greatest uncertainty was observed in scenario 3. We used all predicted heights, both to adjust the model and to estimate the volume and biomass in the plots. Given this, we draw the conclusion that if the aim is to reduce uncertainty it is feasible to build a database with measured height to fit a model for volume or biomass. The other option is to use volume and

biomass models that do not present this variable as input whether these models maintain the same accuracy as well as models that consider height as a variable. In relation to measuring tree heights in the plots, it is worthwhile to analyze from a financial perspective because although the uncertainty is lower with some collected heights, the difference does not significantly interfere in the results.

The systematic error in the inventory was the second largest source of error. The third one is caused by random errors in both variables, d_{hb} and h . These systematic and random errors are due to the accuracy of the device as well as individual practices. In order to reduce the possibility of these errors occurring, care should be taken. Furthermore, the use of methodologies to better detect these errors is important.

One of the main conclusions of our study is that uncertainty sources, e.g. height model and measurement errors, interact in different ways and are propagated in the final uncertainty estimations in terms of volume and biomass. Uncertainty in predicted heights and measurement errors may be more significant in the model than in sampling. They thus directly influence the total uncertainty. Using predicted heights as a variable in a volume and biomass model resulted in different uncertainty than using predicted heights in plots. When it enters the model with the predicted tree heights it propagates errors for both the model and the sampling. This occurs because it incorporates errors from another model into the main model. Measurement errors show different behavior when they are in the subset to fit the model than in the plots.

Even with consideration of the different sources of error, the largest source of uncertainty was sampling. In all scenarios, this source of uncertainty was more than half of the total uncertainty. Model-related uncertainty varied more among scenarios with different uncertainty sources. It is also worthwhile to highlight that sampling and the volume/biomass model are the two main sources of uncertainty in forest inventory and they are independent of each other. In other words, one does not influence the other. The other sources of errors presented in the scenarios - measurement errors and height models, together with the predicted heights - influence and interact in different ways with these main sources.

Therefore, understanding the effect of multiple sources of uncertainty in the estimates helps identify where best to direct monitoring efforts to improve confidence in those estimates, but in some cases, the most important thing is just to report the amount of uncertainty related to the main sources appropriately.

5.5 REFERENCES

BEHLING, A., PÉLLICO NETTO, S., SANQUETTA, C. R., CORTE, A. P. D., AFFLECK, D. L., RODRIGUES, A. L. e BEHLING, M. (2018). Critical analyses when modeling tree biomass to ensure additivity of its components. *Anais da Academia Brasileira de Ciências*, 90:1759–1774.

- BERGER, A., GSCHWANTNER, T., MCROBERTS, R. E. e SCHADAUER, K. (2014). Effects of measurement errors on individual tree stem volume estimates for the austrian national forest inventory. *Forest Science*, 60(1):14–24.
- CANAVAN, S. J. e HANN, D. W. (2004). The two-stage method for measurement error characterization. *Forest Science*, 50(6):743–756.
- CHAVE, J., CONDIT, R., AGUILAR, S., HERNANDEZ, A., LAO, S. e PEREZ, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1443):409–420.
- COLBERT, K. C., LARSEN, D. R. e LOOTENS, J. R. (2002). Height-diameter equations for thirteen midwestern bottomland hardwood species. *Northern Journal of Applied Forestry*, 19(4):171–176.
- CUNIA, T. (1987). Error of forest inventory estimates: its main components.in: Wharton eh, cunia t (eds) estimating tree biomass regressions and their error.
- DE LERA GARRIDO, A., GOBAKKEN, T., ØRKA, H. O., NÆSSET, E. e BOLLANDSÅS, O. M. (2020). Reuse of field data in als-assisted forest inventory. *Silva Fennica*, 54(5).
- DE PETRIS, S., SARVIA, F. e BORGOGNO-MONDINO, E. (2022). About tree height measurement: Theoretical and practical issues for uncertainty quantification and mapping. *Forests*, 13(7):969.
- FORTIN, M. (2020). Updating plots to improve the precision of small-area estimates: the example of the lorraine region, france. *Canadian Journal of Forest Research*, 50(7):648–658.
- FORTIN, M. e DEBLOIS, J. (2010). A statistical estimator to propagate height prediction errors into a general volume model. *Canadian journal of forest research*, 40(10):1930–1939.
- FORTIN, M., MANSO, R. e SCHNEIDER, R. (2018). Parametric bootstrap estimators for hybrid inference in forest inventories. *Forestry: An International Journal of Forest Research*, 91(3):354–365.
- GARBER, S. M., TEMESGEN, H., MONLEON, V. J. e HANN, D. W. (2009). Effects of height imputation strategies on stand volume estimation. *Canadian journal of forest research*, 39(3):681–690.
- GERTNER, G. Z. (1990). The sensitivity of measurement error in stand volume estimation. *Canadian Journal of Forest Research*, 20(6):800–804.
- GRAFSTRÖM, A. e RINGVALL, A. H. (2013). Improving forest field inventories by using remote sensing data in novel sampling designs. *Canadian Journal of Forest Research*, 43(11):1015–1022.

- HEROLD, M., CARTER, S., AVITABILE, V., ESPEJO, A. B., JONCKHEERE, I., LUCAS, R., MCROBERTS, R. E., NÆSSET, E., NIGHTINGALE, J., PETERSEN, R. e OTHERS (2019). The role and need for space-based forest biomass-related measurements in environmental management and policy. *Surveys in Geophysics*, 40(4):757–778.
- JURJEVIĆ, L., LIANG, X., GAŠPAROVIĆ, M. e BALENOVIĆ, I. (2020). Is field-measured tree height as reliable as believed—part ii, a comparison study of tree height estimates from conventional field measurement and low-cost close-range remote sensing in a deciduous forest. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169:227–241.
- KANGAS, A. (1996). On the bias and variance in tree volume predictions due to model and measurement errors. *Scandinavian Journal of Forest Research*, 11(1-4):281–290.
- KEARSLEY, E., DE HAULLEVILLE, T., HUFKENS, K., KIDIMBU, A., TOIRAMBE, B., BAERT, G., HUYGENS, D., KEBEDE, Y., DEFOURNY, P., BOGAERT, J. e OTHERS (2013). Conventional tree height–diameter relationships significantly overestimate aboveground carbon stocks in the central congo basin. *Nature communications*, 4(1):2269.
- KIRCHNER, M., MITTER, H., SCHNEIDER, U. A., SOMMER, M., FALKNER, K. e SCHMID, E. (2021). Uncertainty concepts for integrated modeling-review and application for identifying uncertainties and uncertainty propagation pathways. *Environmental Modelling & Software*, 135:104905.
- LAROCQUE, G. R., BHATTI, J. S., BOUTIN, R. e CHERTOV, O. (2008). Uncertainty analysis in carbon cycle models of forest ecosystems: research needs and development of a theoretical framework to estimate error propagation. *Ecological Modelling*, 219(3-4):400–412.
- MAGNUSSEN, S., KLEINN, C. e FEHRMANN, L. (2020). Wood volume errors from measured and predicted heights. *European Journal of Forest Research*, 139:169–178.
- MÄKINEN, A. M., KANGAS, A. S. e TOKOLA, T. (2010). Identifying the presence of assessment errors in forest inventory data by data mining. *Forest Science*, 56(3):301–312.
- MCROBERTS, R. E. e WESTFALL, J. A. (2014). Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. *Forest Science*, 60(1):34–42.
- MCROBERTS, R. E. e WESTFALL, J. A. (2015). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, 73(3):625–633.
- MONTGOMERY, D. C. (2020). *Introduction to statistical quality control*. John Wiley & Sons.
- NÄSLUND, M. (1936). Die durchforstungsversuche der forstlichen versuchsanstalt schwedens in kiefernwald: Primärbearbeitung. meddelanden från statens skogsförsöksanstalt 29 (1). *Swedish with German summary*.

- PARÉ, D., GERTNER, G., BERNIER, P. e YANAI, R. (2016). Quantifying uncertainty in forest measurements and models: approaches and applications. *Canadian Journal of Forest Research*, 46(3):v–v.
- PARRESOL, B. R. (2001). Additivity of nonlinear biomass equations. *Canadian Journal of Forest Research*, 31(5):865–878.
- PICARD, N., BOYEMBA BOSELA, F. e ROSSI, V. (2015). Reducing the error in biomass estimates strongly depends on model selection. *Annals of forest Science*, 72(6):811–823.
- QIN, L., LIU, Q., ZHANG, M. e SAEED, S. (2019). Effect of measurement errors on the estimation of tree biomass. *Canadian Journal of Forest Research*, 49(11):1371–1378.
- SAARELA, S., ANDERSEN, H.-E., GRAFSTRÖM, A., SCHNELL, S., GOBAKKEN, T., NÆSSET, E., NELSON, R. F., MCROBERTS, R. E., GREGOIRE, T. G. e STÅHL, G. (2017). A new prediction-based variance estimator for two-stage model-assisted surveys of forest resources. *Remote Sensing of Environment*, 192:1–11.
- SAARELA, S., SCHNELL, S., GRAFSTRÖM, A., TUOMINEN, S., NORDKVIST, K., HYYPPÄ, J., KANGAS, A. e STÅHL, G. (2015). Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. *Canadian Journal of Forest Research*, 45(11):1524–1534.
- SAARELA, S., WÄSTLUND, A., HOLMSTRÖM, E., MENSAH, A. A., HOLM, S., NILSSON, M., FRIDMAN, J. e STÅHL, G. (2020). Mapping aboveground biomass and its prediction uncertainty using lidar and field data, accounting for tree-level allometric and lidar model errors. *Forest Ecosystems*, 7(1):1–17.
- SCHUMACHER, F. X. e HALL, F. S. (1933). Logarithmic expression of timber-tree volume. *J Agric Res*, 47:719–734.
- SPURR, S. H. (1952). Forest inventory.
- STÅHL, G., HEIKKINEN, J., PETERSSON, H., REPOLA, J. e HOLM, S. (2014). Sample-based estimation of greenhouse gas emissions from forests—a new approach to account for both sampling and model errors. *Forest Science*, 60(1):3–13.
- SULLIVAN, M. J., LEWIS, S. L., HUBAU, W., QIE, L., BAKER, T. R., BANIN, L. F., CHAVE, J., CUNI-SANCHEZ, A., FELDPAUSCH, T. R., LOPEZ-GONZALEZ, G. e OTHERS (2018). Field methods for sampling tree height for tropical forest biomass estimation. *Methods in Ecology and Evolution*, 9(5):1179–1189.

TRAUTENMÜLLER, J. W., NETTO, S. P., BALBINOT, R., WATZLAWICK, L. F., DALLA CORTE, A. P., SANQUETTA, C. R. e BEHLING, A. (2021). Regression estimators for aboveground biomass and its constituent parts of trees in native southern brazilian forests. *Ecological Indicators*, 130:108025.

6 FINAL CONSIDERATIONS: ANSWERS TO THE SCIENTIFIC QUESTIONS

i) What are the concepts and definitions of uncertainty in volume and biomass inventory estimates?

Based on our literature review chapter, we aimed to present a first discussion on the main concepts of uncertainty linked to volume and biomass estimates in forest inventory. Thus, we hold to the concept: “Uncertainty arises from imperfect or unknown information and some sources of uncertainty are intrinsic to the measurement process and can never be completely removed, such as sampling and modelling of variables. This is called irreducible or natural uncertainty. The sources of uncertainty generate a chain of uncertainty propagation up to the final volume and biomass estimates.”

It was necessary to present the main concepts and most important methods for dealing with uncertainty in forest inventories in one document. As a result of this review of published research on the subject, the entire thesis was constructed on a solid foundation. A further benefit of this review is that it may serve as a reference for future research concerning forest volume and biomass estimates that seek to assess uncertainty. However, the literature review presents concepts and methodologies that can be applied to other forest variables as well, such as forest growth, carbon estimates, ecophysiological models, and climate change models.

ii) What is the state of the art in uncertainty studies in volume and biomass estimates?

We gather the main papers about uncertainty analysis in forest inventory. Most of the research on uncertainty analysis were carried out in the countries of the northern hemisphere, with emphasis on (CHAVE et al., 2004) who worked with uncertainty in Panama. Therefore, we identified a lack of research on uncertainty analysis in Brazilian forests. The review chapter comes as another strong point of discussion on uncertainty analysis in forest estimates in Brazil and will help to better understand and quantify uncertainty in forest estimates.

Regarding the state of the art, we note that hybrid variance estimators have been adapted over the last few years (FU et al., 2017; FORTIN et al., 2018; MELO et al., 2019; FORTIN, 2020). We also observed an evolution regarding the methodologies used to assess uncertainty, starting with analytical hybrid estimators applied in linear models for biomass (CUNIA, 1987) and non-linear models for volume (GERTNER, 1990). Bootstrap methods were later used widely in volume and biomass models (MCROBERTS–WESTFALL, 2014, 2015; PICARD et al., 2015). These methods have started to be used combined with forest growth models as well (FORTIN et al., 2016; MELO et al., 2018). It is evident from the existing literature that different methods can identify sources of uncertainty effectively and efficiently. Despite this, there is considerable scope for improvement in uncertainty assessment using reliable methodologies and implementing these analyses in practical applications within this context (YANAI et al., 2020).

iii) How do some factors related to the inventory data, (e.g. number of trees to fit the models, between-plot variance, and model precision) affect the uncertainty?

A key issue in our research was to determine the impact of the sample size to fit the models in the overall uncertainty. We observed that when we used few trees the uncertainty of the model was large and differed between the linear and the non-linear models. Thus, when we increased the number of trees to fit the model, we observed a reduction of uncertainty in the parameters of all models. This is specifically due to the model parameters being estimated more accurately because the parameter estimates represent the conditional mean of the observations and their standard error naturally decrease with increasing sample size.

The analytical estimators used here linearize the model functional form through a Taylor expansion and had a large influence of the number of trees on the participation of the second Taylor derivative. Small amounts of trees to fit the model led to a more model-related uncertainty in the estimates and the second derivative has a larger contribution. We also note that linear and nonlinear models respond differently to the number of trees to fit the models. The linear model presented less uncertainty in their parameters with fewer trees than nonlinear models. The non-linearity of the model contributes to this increase in the model-related uncertainty.

We also emphasize that our aims were to evaluate how these factors affect the uncertainty. Reducing uncertainty comes as a result of increasing sample size to fit the model. However, this goes against financial and practical considerations, as the greater the number of trees collected, the greater the time and cost for data collection. This is an important issue to look at in practical work. We also emphasize that uncertainty does not necessarily need to be reduced in the forest inventory, but its evaluation becomes indispensable for analyses with greater reliability.

Our equation in the first chapter showed good quality of fit and the fit statistic was not modified with sample size to fit the model, while the uncertainty was affected by this. In view of our results, we believe that further in-depth studies should be conducted regarding to the quality of the model. The equations' quality is linked to how well the equation can explain the results of the response variable, while the model-related uncertainty in volume and biomass estimates explains how the errors propagate into these estimates. Furthermore, other factors such as plot shape, sampling intensity and distribution of plots can be studied in further work to verify if there is an influence on uncertainty due to sampling or the model.

iv) Based on these factors, what is the amount and share of uncertainty arising from the model and sampling?

We quantify the uncertainty due to sampling and model and we wrote the results throughout the chapters. The uncertainty related to the volume model is generally smaller than the sampling uncertainty. This is also true for biomass, although the share of model uncertainty is higher than that of volume. Even though sampling error was not affected by the sample size to fit the models, it remained the main source of error. It appears that sampling uncertainty represents

the largest share in total uncertainty for linear models in all sample size groups, and in non-linear models when sample size increases, and model-related variation is low (more than 90% for total uncertainty in volume and 80% for biomass). This was an average result of the survey, but the uncertainty of each source varies across the different case studies and situations presented in the different chapters.

We believe that researchers in Brazil will benefit from this set of studies when attempting to quantify the uncertainty in population-level estimates in forestry. We also expect this research will help to raise awareness of the practical gain and the care due in terms of data reliability, as well as the importance of considering uncertainty in stand-level inventories and NFI.

v) How can variance hybrid estimators be applied, and adapted to quantify uncertainty more accurately for linear and non-linear models?

We answered this question in all chapters of the thesis. In chapter 1 we demonstrated the concepts of analytic and Bootstrap hybrid estimators. In the chapter 2 we achieved the objectives of quantifying uncertainty with the analytical approach for linear model, presented by (CUNIA, 1987). In chapter 3 we needed to calculate the uncertainty with an analytical approach for non-linear models. The first results following GERTNER (1990) presented a large underestimation in the model-related uncertainty, thus, we started studying the within and between-plot correlation for predicting errors. Therefore, we developed a correction for the inclusion this correlation, and we strongly recommend considering it to quantify the uncertainty in non-linear models if an analytical approach is preferred.

Another challenge we faced working with non-linear models in an analytical framework was the accuracy of the linearization of the model functional form needed in these methods. This is done through a (typically first-order) Taylor expansion. The first derivative in the Taylor expansion proved insufficient to properly propagate the uncertainty in non-linear models tested. Thus, we studied how to quantify and add the second derivative to the first one. We emphasize that it is important to investigate procedures to represent exact uncertainty and achieve its more precise uncertainty estimation.

The approach with bootstrap estimators was flexible for all models analyzed in the thesis, especially when we fitted the model with a random effect in chapter 4. All results with the analytical hybrid estimators were close to the Bootstrap estimator. We also checked whether the correction proposed to the classic bootstrap estimator (MCROBERTS–WESTFALL, 2015) to deal with inherent underestimation of the sampling part of the uncertainty (FORTIN et al., 2018), made a big difference in our case study. Although it did not, we recommend not ignoring this correction, as it is straightforward to implement. In other cases it may fit.

We emphasize that variance hybrid estimators need to be free from bias. Also, they must be accurate and provide reliable estimates of the uncertainty in volume and biomass estimates. Hybrid estimators presented in the thesis provided accurate uncertainty estimates.

As a result, we compiled a collection of methodologies to address uncertainty and, with the necessary adaptations, we were able to create more reliable variance estimators. Consequently, it is important to analyze which methodology represents all sources of uncertainty in a more accurate way before applying variance hybrid estimators.

vi) How do the different sources of uncertainty interact in the error propagation chain and influence the total uncertainty?

We considered for this research the volume or biomass model, sampling, height model and measurement errors as the sources of uncertainty in the error propagation chain. The sources of uncertainty interact in different ways, according to whether each source of uncertainty enters into the volume and biomass estimates, e.g. into the input variables to fit the model or into the variables of the trees in the plots.

However, all sources are propagated in the final uncertainty calculations in terms of volume and biomass. Using the stochastically predicted heights as an input variable in a volume and biomass model resulted in different uncertainty than using the stochastically predicted heights in the plots only. When entering predicted tree heights into the model, they propagate errors for both the model and the sampling. This occurs because it incorporates errors from another model into the main model as well as plot-level predictions seem to be more varied when using predicted height. Measurement errors show different behavior when they are in the subset to fit the model than in the plots.

Another point we would like to discuss is that some factors affect the different sources of uncertainty differently. In our study we developed these two aspects: factors affecting uncertainty and some sources of uncertainty in separate evaluations. However, in future research it could be analyzed how the sources of uncertainty interact, having the influence of factors such as the number of trees to estimate the model, the number of trees in the inventory plots, the size, intensity, and distribution of the inventory plots in the forest.

vii) How does the equation for estimating height affect uncertainty in volume and biomass estimates?

We discuss this scientific issue in chapter 4. We used all predicted heights, both to fit the model and to estimate the volume and biomass in the plots. Given this, we conclude that if the objective is to reduce uncertainty, it is preferable to build a database with measured height to fit a model for volume or biomass, rather than using stochastically predicted heights. However, it is important to consider whether this is financially convenient in terms of data collection.

The other option is to use volume and biomass models that do not present height as an input variable in the model if these models keep a similar level of the same accuracy as models that do consider height as a variable. However, in our situation we did not analyze this type of model, not being able to indicate the merit of this option. Thus, future studies need

to be conducted to verify this possibility. Furthermore, more specific studies focusing on the quantification of uncertainty in the height equation need to be analyzed. An alternative for this is to seek methodologies already existing in the literature or even to develop estimators to quantify the exact value of the uncertainty related to the height model.

viii) How do measurement errors in auxiliary variables -dbh and h- affect the uncertainty in volume and biomass estimates?

The systematic error in the inventory plots caused the largest uncertainty related to measurement errors. After that, the random errors in our two explanatory, diameter and height, also had an impact on the uncertainty. These systematic and random errors would be imitating those due to device accuracy and operator skills and good practice.

In order to reduce the possibility of these errors occurring in the different scenarios we set, and thus affecting the uncertainty in the estimates, care should be taken. Furthermore, the use of methodologies to better detect these errors are necessary for an improvement in data collection that totally affects the processing and final results.

*ix) What is the effect of uncertainty on the management of *Acacia mearnsii* stands and what is the influence for the economic issues?*

The effects of uncertainty in *Acacia mearnsii* management are seen directly in the reliability of volume and biomass estimates in stands. Thus, uncertainty is associated with loss of information, precision and reliability of volume and biomass estimates. Quantifying and assessing uncertainty is therefore crucial to maintain the quality and determine the accuracy of estimates reported in the forest inventory.

Thus, when uncertainty is not properly evaluated in forest inventory estimates, the total variance will tend to be underestimated and the resulting confidence intervals narrower what they should report, which would lead one to believe that the calculated estimate is more realistic than it actually is. Consequently, forest managers may have a misleading perception of precision, with potential impacts on decision-making and in the formulation of forest management. Understanding the effect of multiple sources of uncertainty in any estimate helps identifying where best to direct monitoring effort to improve confidence in those estimates.

6.0 REFERENCES

- CHAVE, J., CONDIT, R., AGUILAR, S., HERNANDEZ, A., LAO, S. e PEREZ, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1443):409–420.
- CUNIA, T. (1987). Error of forest inventory estimates: its main components.in: Wharton eh, cunia t (eds) estimating tree biomass regressions and their error.

- FORTIN, M. (2020). Updating plots to improve the precision of small-area estimates: the example of the lorraine region, france. *Canadian Journal of Forest Research*, 50(7):648–658.
- FORTIN, M., MANSO, R. e SCHNEIDER, R. (2018). Parametric bootstrap estimators for hybrid inference in forest inventories. *Forestry: An International Journal of Forest Research*, 91(3):354–365.
- FORTIN, M., ROBERT, N. e MANSO, R. (2016). Uncertainty assessment of large-scale forest growth predictions based on a transition-matrix model in catalonia. *Annals of Forest Science*, 73(4):871–883.
- FU, Y., LEI, Y., ZENG, W., HAO, R., ZHANG, G., ZHONG, Q. e XU, M. (2017). Uncertainty assessment in aboveground biomass estimation at the regional scale using a new method considering both sampling error and model error. *Canadian Journal of Forest Research*, 47(8):1095–1103.
- GERTNER, G. Z. (1990). The sensitivity of measurement error in stand volume estimation. *Canadian Journal of Forest Research*, 20(6):800–804.
- MCROBERTS, R. E. e WESTFALL, J. A. (2014). Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. *Forest Science*, 60(1):34–42.
- MCROBERTS, R. E. e WESTFALL, J. A. (2015). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. *Annals of Forest Science*, 73(3):625–633.
- MELO, L., SCHNEIDER, R. e FORTIN, M. (2019). The effect of natural and anthropogenic disturbances on the uncertainty of large-area forest growth forecasts. *Forestry: An International Journal of Forest Research*, 92(3):231–241.
- MELO, L. C., SCHNEIDER, R. e FORTIN, M. (2018). Estimating model-and sampling-related uncertainty in large-area growth predictions. *Ecological Modelling*, 390:62–69.
- PICARD, N., BOYEMBA BOSELA, F. e ROSSI, V. (2015). Reducing the error in biomass estimates strongly depends on model selection. *Annals of forest Science*, 72(6):811–823.
- YANAI, R. D., WAYSON, C., LEE, D., ESPEJO, A., CAMPBELL, J. L., GREEN, M. B., ZUKSWERT, J. M., YOFFE, S., AUKEMA, J., LISTER, A. e OTHERS (2020). Improving uncertainty in forest carbon accounting for redd+ mitigation efforts. *Environmental Research Letters*, 15(12):124002.

APÊNDICE A – VARIANCE COVARIANCE MATRIX AND PARAMETERS STANDARD ERROR

A.1 VOLUME

Variance-covariance matrix

Group	Model					
	1	2	3			
G1	$\begin{bmatrix} 7.02E-07 & -1.88E-10 \\ -1.88E-10 & 9.10E-14 \end{bmatrix}$	$\begin{bmatrix} 1.37E-11 & -2.72E-08 \\ -2.72E-08 & 5.44E-05 \end{bmatrix}$	$\begin{bmatrix} 5.05E-11 & 1.67E-07 & -4.45E-07 \\ 1.67E-07 & 1.52E-03 & -2.36E-03 \\ -4.45E-07 & -2.36E-03 & 4.75E-03 \end{bmatrix}$			
G2	$\begin{bmatrix} 3.68E-07 & -1.11E-10 \\ -1.11E-10 & 6.22E-14 \end{bmatrix}$	$\begin{bmatrix} 1.16E-11 & -2.62E-08 \\ -2.62E-08 & 5.94E-05 \end{bmatrix}$	$\begin{bmatrix} 3.15E-11 & 4.94E-08 & -2.38E-07 \\ 4.94E-08 & 8.39E-04 & -1.09E-03 \\ -2.38E-07 & -1.09E-03 & 2.48E-03 \end{bmatrix}$			
G3	$\begin{bmatrix} 1.58E-07 & -5.41E-11 \\ -5.41E-11 & 3.84E-14 \end{bmatrix}$	$\begin{bmatrix} 6.76E-12 & -1.47E-08 \\ -1.47E-08 & 3.22E-05 \end{bmatrix}$	$\begin{bmatrix} 1.30E-11 & 2.03E-08 & -9.93E-08 \\ 2.03E-08 & 5.45E-04 & -6.36E-04 \\ -9.93E-08 & -6.36E-04 & 1.21E-03 \end{bmatrix}$			
G4	$\begin{bmatrix} 7.92E-08 & -2.49E-11 \\ -2.49E-11 & 2.11E-14 \end{bmatrix}$	$\begin{bmatrix} 4.54E-12 & -9.78E-09 \\ -9.78E-09 & 2.12E-05 \end{bmatrix}$	$\begin{bmatrix} 7.55E-12 & 9.15E-09 & -5.49E-08 \\ 9.15E-09 & 3.78E-04 & -4.12E-04 \\ -5.49E-08 & -4.12E-04 & 7.25E-04 \end{bmatrix}$			

TABLE A1-1: standard errors of the parameter estimates in to volume equations.

Model	Parameter	G1	G2	G3	G4
2*1	$\hat{\beta}_1$	0,001007	5,38E-04	3,36E-03	0,0002757
	$\hat{\beta}_2$	3,169E-07	2,17E-07	0,00003719	1,356E-07
2*2	$\hat{\beta}_1$	0,000003699	2,97E-06	0,00000222	0,000001936
	$\hat{\beta}_2$	0,007376496	6,00E-03	0,00449	0,00397
3*3	$\hat{\beta}_1$	0,00000711	5,02E-06	0,000003071	0,000001886
	$\hat{\beta}_2$	0,03898204	2,32E-02	0,01908	0,017674654
	$\hat{\beta}_3$	0,06889915	3,92E-02	0,02795	0,02094766

A.2 ABOVEGROUND BIOMASS

Group	Model		
	1	2	3
G1	$\begin{bmatrix} 0.827285741 & -2.41E-04 \\ -0.000241485 & 1.40E-07 \end{bmatrix}$	$\begin{bmatrix} 1.58E-05 & -6.73E-05 \\ -6.73E-05 & 2.88E-04 \end{bmatrix}$	$\begin{bmatrix} 0.000107075 & 0.000432344 & -0.001137797 \\ 0.000432344 & 0.00514387 & -0.0077008 \\ -0.001137797 & -0.0077008 & 0.014943929 \end{bmatrix}$
G2	$\begin{bmatrix} 1.92E-01 & -6.87E-05 \\ -6.87E-05 & 5.74E-08 \end{bmatrix}$	$\begin{bmatrix} 5.50E-06 & -2.57E-05 \\ -2.57E-05 & 1.20E-04 \end{bmatrix}$	$\begin{bmatrix} 2.91E-05 & 9.74E-05 & -0.00035659 \\ 9.74E-05 & 2.07E-03 & -0.002818835 \\ -3.57E-04 & -2.82E-03 & 0.005889469 \end{bmatrix}$
G3	$\begin{bmatrix} 8.98E-02 & -2.77E-05 \\ -2.77E-05 & 3.08E-08 \end{bmatrix}$	$\begin{bmatrix} 3.57E-06 & -1.81E-05 \\ -1.81E-05 & 9.17E-05 \end{bmatrix}$	$\begin{bmatrix} 1.58E-05 & 4.92E-05 & -0.000205527 \\ 4.92E-05 & 1.47E-03 & -0.001883378 \\ -2.06E-04 & -1.88E-03 & 0.003852586 \end{bmatrix}$
G4	$\begin{bmatrix} 4.20E-02 & -1.49E-05 \\ -1.49E-05 & 2.32E-08 \end{bmatrix}$	$\begin{bmatrix} 2.79E-06 & -1.35E-05 \\ -1.35E-05 & 6.53E-05 \end{bmatrix}$	$\begin{bmatrix} 7.50E-06 & 1.74E-05 & -7.74E-05 \\ 1.74E-05 & 1.29E-03 & -1.33E-03 \\ -7.74E-05 & -1.33E-03 & 1.86E-03 \end{bmatrix}$

TABLE A1-2: standard errors of the parameter estimates in to biomass equations.

Model	Parameter	G1	G2	G3	G4
2*1	$\hat{\beta}_1$	12.952.915	4,38E-01	3,00E-01	0,204867
	$\hat{\beta}_2$	0,0004075	2,40E-04	0,0001756	0,0001523
2*2	$\hat{\beta}_1$	0,004052	2,34E-03	0,00189	0,00167
	$\hat{\beta}_2$	0,0018202	1,10E-02	0,009576	0,008079
3*3	$\hat{\beta}_1$	0,01034772	5,40E-03	0,003977	0,00273857
	$\hat{\beta}_2$	0,07172078	4,55E-02	0,038306	0,03588246
	$\hat{\beta}_3$	0,12224536	7,67E-02	0,062069	0,04311246

APÊNDICE B – UNCERTAINTY ASSOCIATED IN THE PREDICTIONS

B.0.1 Volume

Grupo	Model	Sampling		Model Uncertainty		Total Um		Under (%)		IC		Mean ($m^3.ha^{-1}$)
		Analytical	MC	Analytical	MC	Analytical	MC	Analytical	MC	Analytical	MC	
G1	1	42,81	43,28	2,98	3,05	45,80	46,33	6,51	6,59	13,43	13,51	320,36
	2	39,97	39,83	6,36	6,273	46,33	46,10	13,72	13,60	13,51	13,48	313,66
	3	39,95	39,42	15,41	15,75	55,36	55,18	27,84	28,55	14,77	14,75	313,63
G2	1	42,70	43,17	2,19	2,06	44,90	45,23	4,88	4,55	13,30	13,35	318,66
	2	40,92	40,74	4,55	4,38	45,48	45,13	10,01	9,72	13,39	13,34	314,82
	3	40,94	40,59	8,76	8,01	49,71	48,61	17,64	16,49	14,00	13,84	314,72
G3	1	42,64	43,09	1,57	1,65	44,21	44,75	3,55	3,69	13,20	13,28	318,54
	2	40,39	40,33	2,88	2,85	43,27	43,19	6,66	6,61	13,06	13,05	313,66
	3	40,41	40,22	3,53	3,32	43,94	43,54	8,03	7,62	13,16	13,10	313,63
G4	1	42,71	43,16	0,990	1,02	43,70	44,18	2,26	2,311	13,12	13,20	318,39
	2	40,37	40,33	1,74	1,67	42,12	42,003	4,15	3,98	12,88	12,86	313,64
	3	40,41	40,33	1,93	1,78	42,35	42,11	4,56	4,23	12,92	12,88	313,62

Where: Under is the relative underestimation of the total variance when the uncertainty due to the model is not considered.

B.0.2 Biomass above-ground

Group	Model	Sampling		Model Uncertainty		Total Uncertainty		Under (%)		IC		Mean ($Kg.ha^{-1}$)
		Analytical	MC	Analytical	MC	Analytical	MC	Analytical	MC	Analytical	MC	
G1	1	22478,90396	22720,94517	5187,130246	5124,85084	27666,03421	27845,79601	18,74	18,4	329,3356	330,4038	230519,3
	2	22848,1675	22444,85248	14446,05613	17614,12401	37294,22363	40058,97649	38,73	43,97	382,3719	396,2918	230835,6
	3	21443,00535	20522,25728	43603,89534	69517,16552	65046,90069	90039,4228	67,03	77,2	504,985	594,1301	223405,1
G2	1	21323,73994	21553,30154	2579,20107	2533,700759	23902,94101	24087,0023	10,79	10,51	306,1194	307,2958	224134,8
	2	21765,11293	21616,84828	5463,380818	5000,874337	27228,49375	26617,72262	20,06	18,78	326,721	323,0358	224746,2
	3	22081,51093	21637,8253	12714,24313	14526,14047	34795,75406	36163,96577	36,53	40,16	369,3417	376,5331	224839,9
G3	1	21217,50988	21449,75512	1622115,384	1658,971117	1643332,894	23108,72624	7,1	7,17	2538,212	300,9908	223797,7
	2	22164,56284	22030,959	3200828,494	3518,254758	3222993,057	25549,21376	12,61	13,77	3554,634	316,4856	225517,9
	3	22477,12994	22167,39238	6307,504187	7096,670667	28784,63413	29264,06305	21,91	24,25	335,9275	338,7135	225334,2
G4	1	20932,65485	21155,83366	1317,44592	1223,301227	22250,10077	22379,13489	5,92	5,46	295,3461	296,2012	222481,6
	2	21766,5438	21766,5438	2372,534355	2372,534355	24139,07816	24139,07816	9,82	9,82	307,6278	307,6278	224234,5
	3	22051,85	19013,81464	4457,373764	4302,91403	26509,22376	23316,72867	18,94	18,45	322,3767	302,3424	211723,1