Parameter testing and application of the 3PG model for *Eucalyptus grandis x urophylla* on the Zululand coastal plain, South Africa

by

Oluwaseun Friday Gakenou

Thesis presented in partial fulfilment of the requirements for the degree of
Master of Forestry and Natural Resources Science

at
Stellenbosch University
Department of Forestry and Wood Sciences, Faculty of AgriSciences

Supervisor: Professor David Michael Drew
Co-supervisor: Doctor Ilaria Germishuizen

December 2021

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 06/12/2021

Summary

This study aims to calibrate the 3PG (Physiological Processes Predicting Growth) model for *Eucalyptus grandis x urophylla* growing in the coastal Zululand region, South Africa. Parameter values developed for this hybrid across regions in Brazil were used as baseline parameters. To generate a set of reliable point estimates of weather data for growth modelling, we evaluated the performance of two spatial interpolation techniques (Random Forest and the R package "Meteoland") using Mean Absolute Error, Root Mean Square Error, Coefficient of Determination, Index of Agreement and Nash Sutcliffe Model Efficiency Index. We collected observed long-term weather data from the South African Weather Services (SAWS) and the South African Sugarcane Research Institute (SASRI). Weather stations spread across the KwaZulu-Natal region were used for the performance analysis. Both models showed great potential. However, the Random Forest model was the best performing model used to generate weather data in this study for growth modelling. Parameter estimation of the model was based on 17 permanent sample plots (PSPs) managed by two forestry companies, Mondi Ltd and Sappi Ltd. Allometric parameters for stem mass as a function of stem diameter at breast height were calibrated using biomass harvest data from sampling undertaken in 2018. Eleven parameters were selected from the list of base parameters to be adjusted using a parsimonious optimization approach. A novel method for ranking the parameter set combinations, called *extended Root Mean Square Error (eRMSE),* was created and used to select the optimal parameter set. Using the new parameter set resulted in good predictions of three key output variables (Mean stand height (H, m), stand basal area (BA, m^2/ha), and mean stem diameter at breast height (DBH, cm)) which were then used to calculate stand volume (V, m^3/ha). Model performance at 15 independent validation sites allowed the comparison with three other Brazilian parameter sets. Overall, the 3PG model gave a good but slightly overestimated stand volume prediction at the validation sites. We compared the 3PG model with three simpler models. The forest companies' statistical growth and yield models outperformed all other models in terms of all metrics used, followed by a very simple model using the cumulative rainfall model. Although the 3PG gave similar growth predictions, it demonstrates its usefulness in simulating growth patterns in response to environmental changes.

Dedication

This thesis is dedicated to my dear parents, Mr Paul Gakenou and Mrs Hedessiawa Gakenou, who sacrificed everything for me to study

Acknowledgements

I wish to express my sincere gratitude and appreciation to the following persons and institutions:

- My supervisors: Prof David Michael Drew for bringing me on this project, for his patience, support, and encouragement, for spending time with me in my office to discuss this research work, for taking time to edit each chapter immediately they were ready and for believing in me. Dr Ilaria Germishuizen (ICFR) for teaching me spatial modelling, helping me with packages used for spatial interpolation of weather data, always responding to my email and WhatsApp messages during weekends and when she is on leave, and for her words of encouragement.

- Dr Kim Martin for her invaluable support in assisting me build algorithms in R. She encouraged me to use R in cleaning and large handling data, helped me set up a GitHub account for easy collaborative work, and spent time discussing my work with me.

- Gugu Gama for assisting me with my thesis formatting, encouraged me to drink water and eat fruits whenever I am stressed out.

- Dr Leandra Moller, for always assisting with technical issues in the office.

- Mondi and Sappi for providing funding for this research

- Mr Henys Kotze for providing the Site index-based model and always responding to my queries via email

- The whole EucXylo team for their support and contributions during our weekly meetings.

- My parents for always calling to pray for me, for their ending love and continual encouragement

- To my Fiancée, Olufunke, for staying strong and looking after our baby in my absence and to my baby, Michelle, for giving me a reason to love more and be happy.

- To my Siblings, thank you for believing in me.

- To God Almighty, the giver of life!

Table of Contents

Table of Contents

List of Figures

List of Tables

Table of Equations

List of acronyms

3PG – Physiological Processes Predicting Growth
AI – Aridity index
APAR – Absorbed photosynthetically active radiation
ASW – Available soil water
BA – Basal area
CRAN – Comprehensive R archive network
DBH – Diameter at breast height
eRMSE – extended root mean square error
ET – Evapotranspiration
FR – Fertility rating
GDP – Gross domestic product
GPP – Gross primary productivity
HD – Dominant height
LAI – Leaf area index
MAE – Mean absolute error
MBE – Mean bias error
MinASW – Minimum available soil water
MaxASW – Maximum available soil water
MODIS – Moderate Resolution Imaging Spectroradiometer
NDVI – Normalized Difference Vegetation Index
NPP – Net primary production
OOB – Out-of-bag error
PAR – Photosynthetically active radiation
PE – Potential evapotranspiration
PSPs – Permanent sample plots
RF – Random Forest
RMSE – Root mean square error
SASRi – South Africa Sugar Cane Research Institute
SAWS – South Africa Weather Services
SI – Site index
SOC – Soil Organic Carbon
SSE – Sum of square error
SSF – Sum of square fit
SLA – Specific leaf area
SRAD – Solar radiation
TPH – Tree per hectare
TPP – Tree per plot
VPD – Vapor pressure deficit

Chapter 1 General overview

1.1 Introduction

Although it only represents a small land area in the country (0.97%) (DAFF, 2019), the South African commercial forestry sector is a modern and essential industry contributing about 1% to South Africa's GDP (*https://www.gov.za/about-sa/forestry*). Plantations of exotic tree species, mainly pines and eucalypts, were established to meet domestic and international demands (Campion, 2005; Albaugh et al., 2013). Of the total commercial plantations in South Africa, hardwood species comprising mainly eucalypts occupy about 0.5 million ha (43.7% of the total commercial forestry plantation). Of the eucalypt resource, about 53.4% are clonal and are prominently managed for pulpwood production (DAFF, 2019).

Most commercial plantations are in warm temperate areas (about 57%), followed by cool temperate areas (34%). A minority are in the subtropical areas (9%), which have been historically productive, thus a crucial forestry growing area (Figure 1.1). However, the afforestation of new plantation areas in these regions is restricted by the availability of suitable land and water legislation (Naidoo et al., 2013; DAFF, 2019). This, combined with rising demand for wood and fibre, has compelled the industry to seek new ways to increase the productivity of existing plantations while maintaining low-cost wood production (du Toit et al., 2010).

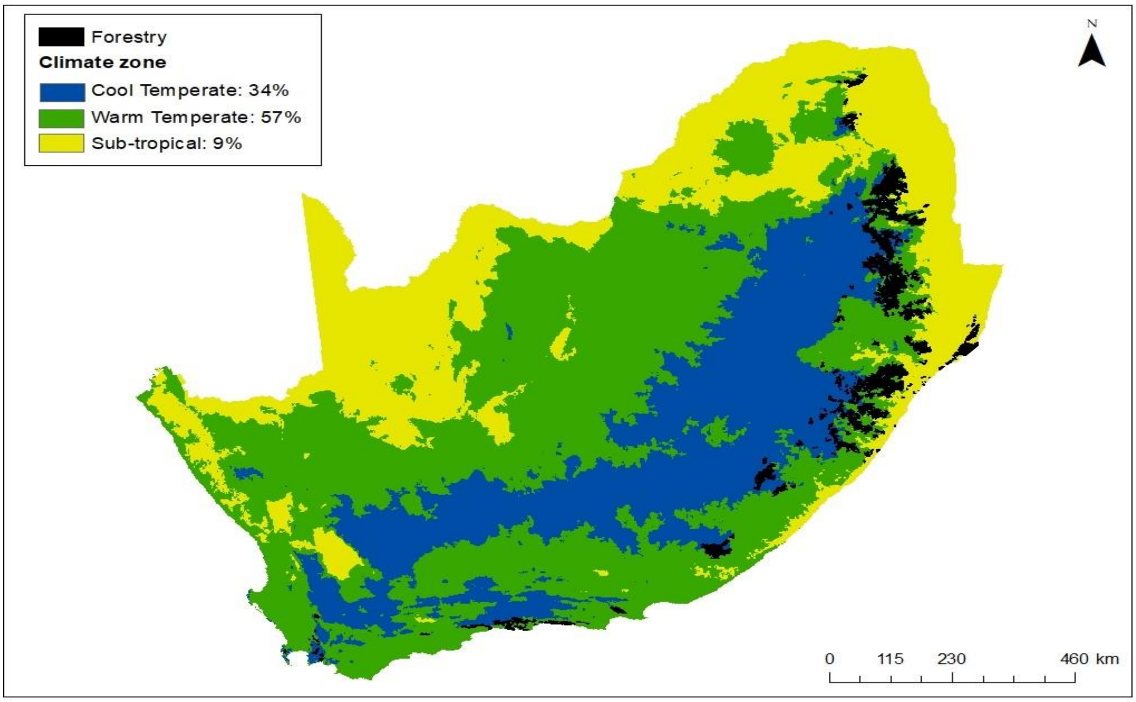


Figure 1.1 Distribution of forestry areas across different climatic zones in South Africa

(14)

One of the forestry industry's strategies was the implementation of hybrid clonal forestry in the subtropical region of coastal Zululand, one of South Africa's major plantation forestry areas and historically very productive (Gardner, 2012). *Eucalyptus grandis*, the popular tree species for commercial pulp production in South Africa, was replaced by a clonal hybrid, *E.grandis x E.urophylla (Egxu),* due to its susceptibility to fungal and pest diseases in this region (Retief & Stanger, 2009; Stanger et al., 2011; Van Den Berg, 2017). This hybrid, which is also commonly planted in other countries such as Brazil (Rezende et al., 2014), combines the fast-growing traits of *E. grandis* with the good survival, disease tolerance, and higher wood density characteristics of *E. urophylla* (Retief & Stanger, 2009). Furthermore, the coppicing ability, short rotation length (8–12 years), high productivity, and suitability for pulp and paper production makes the hybrid an essential raw industrial wood material among South African pulp growers (Melesse & Zewotir, 2017). Productivity gains due to superior genetics of this hybrid have been reported (Gardner, 2012; Melesse & Zewotir, 2015; Melesse & Zewotir, 2017). As a result, the research presented in this thesis was conducted in collaboration with two forestry companies interested in quantifying the effect of climate variation x site, specifically drought, on survival, growth, and uniformity, thereby fibre yield of *Egxu* species planted in the coastal Zululand (see Figure 1.2).

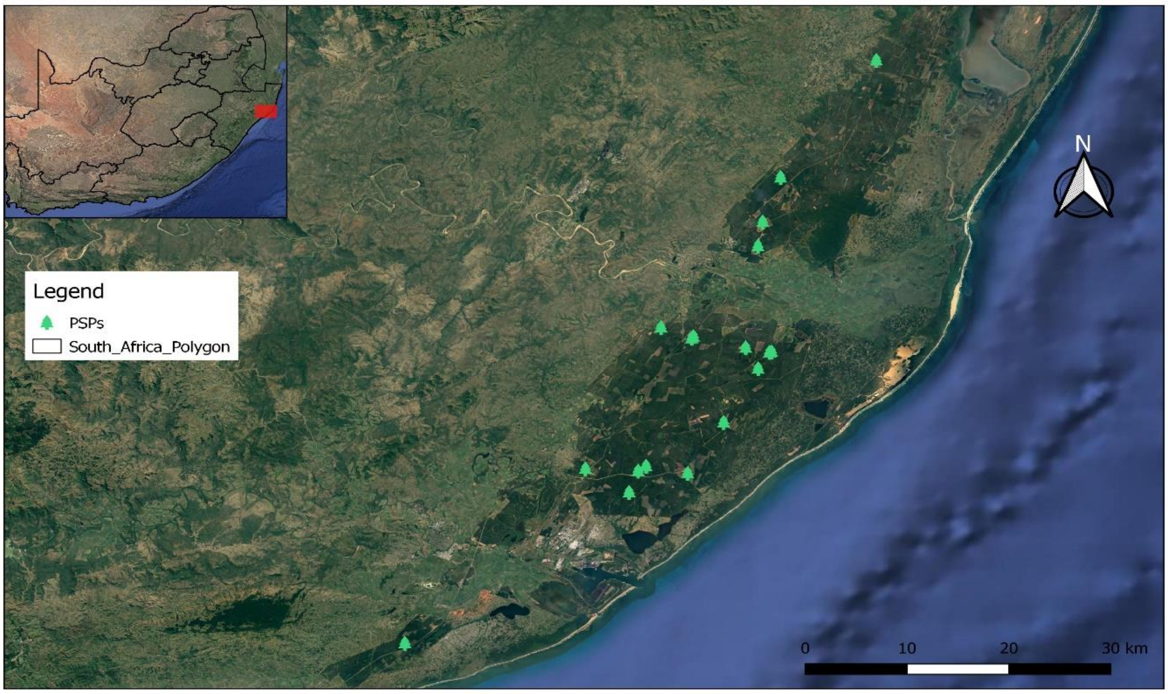


Figure 1.2 Google map showing the extent of the plantations

However, despite these gains, unpredictable climate change and shifts in productivity remain an issue in planning horizons for commercial forest managers. Climate is the only factor foresters cannot directly influence out of the three main factors (climate, genetics, and soil management) responsible for increased productivity levels in *Eucalyptus* plantations (Binkley et al., 2017; Elli, 2020). The impact of climate on forest plantation productivity is caused by changes in atmospheric CO_2 concentration, temperature, rainfall regimes, and extreme events such as pest and disease prevalence, wildfires, and drought (Alig et al., 2004; Warburton & Schulze, 2006). Although the physiological response of trees to the interplay of these changes is still uncertain (Warburton & Schulze, 2006), numerous studies have projected the impact to be site and species-specific (Warburton & Schulze, 2008; Almeida et al., 2009; Pinkard et al., 2010; Naumberg et al., 2001; Booth, 2013). Therefore, it is imperative to consider climatic risk assessment as a valuable tool for improving forestry planning and management (Elli, 2020). South Africa is naturally vulnerable to drought (Gibberd et al., 1996; Baudoin et al., 2017) and has long records of recurrent droughts (Xulu et al., 2018). The region suffered a particularly severe drought combined with an extreme El Niño event in 2015 – 2016 (AgriSA, 2016; Baudoin et al., 2017).

Climate-related changes have been a particular issue in the coastal Zululand region. The impact of drought on plantation forests (*Eucalyptus grandis* and its hybrids) in Zululand was investigated by Xulu et al. (2018). According to Warburton & Schulze (2006), the impact of drought on commercial forestry can be long-term, cost-expensive, and irreversible. As a result, it is necessary to forecast short rotation forestry's long-term and large-scale response to a rapidly changing environment.

1.2 Forest models for decision and planning support

In this context, it is evident that forest scientists need support in managing and planning these forest resources. An important part of working towards a solution is to have access to reliable modelling systems. Forest simulation models are essential decision tools for forest managers; they are helpful in understanding and predicting the long-term impact (in terms of yield-related, financial, and ecological consequences) of management practices and global change on future forest productivity (Pretzsch, 2009). Models have been described and defined differently by ecologists and modellers in various contexts. However, Landsberg & Sands (2011) described them as *"a practical tool designed to simulate the behavior of a system in response to change or stimuli, so that managers or decision-makers can assess the probable consequences of those changes or stimuli."*. This definition can be considered appropriate for the work presented in this thesis.

For many decades, growth and yield models used in forest management in the South African Forestry industry have relied on statistical relationships derived from historical stand growth records measured on sample plots. The main strength of the statistical growth and yield model is in describing the best relationship between measured data and growth-determining variables using a predefined mathematical function or curve (Peng, 2000). Their development is predicated on the notion that the future growth of a stand is determined by the same conditions under which the historical data were collected (Kimmins, 1990). Therefore, these models determine site productivity using the concept of a site index, which is influenced by the past growth condition variables such as climatic conditions, soil fertility, and management practices (Landsberg & Gower, 1997; Landsberg & Sands, 2011). Forest managers primarily use these models because they are considered practical and simple tools for forest management (Esprey, 2006), as the model predicts future yield averaged across stands in a region using readily available or easily measured variables such as stem diameter and height, tree age, stocking, and site quality. These models are primarily used to provide information about log sizes and size distribution, predict timber volumes and yield over short periods for which historical conditions are assumed not to change significantly (Landsberg, 2003b), for organizing harvest scheduling and optimizing timber supplies to mills, for management planning, and updating stand inventory (Almeida, 2003).

Although these models achieved great simplicity from the site-index concept, it also limits the model (Johnsen et al., 2001). The concept implicitly assumes that these variables are constant over the entire rotation or that changes likely to occur will have no significant difference from the growth patterns described by the model (Landsberg & Coops, 1999). Under conditions of substantial change and increasing variability, they are less reliable (Johnsen et al., 2001). In South Africa, for example, where rainfall is a critical factor influencing forest growth, a single drought event is enough to render the application of this model useless (Chauke, 2018). Therefore, the model lacks flexibility in terms of predicting growth response to fluctuating weather conditions, the effect of environmental stresses, and changes in management practices (Landsberg, 2003b; Esprey, 2006). Additionally, these models are not generic across all sites, as site index varies between and within regions (e.g., Goulding, 1994 developed seven distinct growth models for *P.radiata* for different regions in New Zealand) (Landsberg & Sands, 2011). For this reason, the question of how accurate we can predict future yield and minimize risk on short rotation forestry is important for consideration.

An approach that has attracted widespread interest from forest scientists is the mechanistic approach, in which process-based models (PBMs) are used in conjunction with weather or climate data (Elli, 2020). Process-based models are created by scaling up a mathematical representation of the physiological and ecological processes affected by changes in available resources for growth to the tree or stand growth level (Johnsen et al., 2001). The logical reason behind this is that a tree's/stand's growth rate and biomass accumulation is an integral of the rates and activities of the physiological processes such as photosynthesis, respiration, carbon allocation, stomatal conductance, light interception, nutrient cycling, and water use (Johnsen et al., 2001; Landsberg & Sands, 2011). Therefore, this modelling approach aims to simulate the growth of stands in terms of the underlying physiological processes that determine growth and how stands are affected by the physical conditions they are subjected to (Landsberg & Gower, 1997).

Apart from the significant advantage of the PBMs to increase our understanding of the cause-effect relationship of physiological factors determining forest growth, their structure presents great flexibility and generality (Kimmins, 1990; Battaglia & Sands, 1998). This means they have the greatest potential to predict forest growth under changing environmental conditions, predict the productivity of a site that has no previous field data, describe sensitivity to various changes, formulate a hypothesis to test new conditions as a result of climate change, new management practices, or when a practical experiment is not feasible (Korzukhin et al., 1996; Pretzsch, 2009). Several process-based models have been developed as forest management tools. However, some factors affecting the adoption and usage of PBM models as operational tools by the forestry industries include the conservative nature of the forestry sector, cost of generating required data (Almeida, 2018), model documentation, ease of calibration and evaluation, demonstration of utility and most importantly does not produce outputs of interest to forest managers (Mäkelä et al., 2000; Landsberg, 2003a). Almeida (2003) and Esprey (2006) conducted a PBM selection suitable for use as a practical and analytical tool, following a set of criteria laid out by Sands (1988). Out of all existing PBMs reviewed by both authors, the 3PG (Physiological Processes Predicting Growth) model emerges as the most suitable tool for forest management.

1.3 3PG hybrid growth modelling system

The 3PG model, sometimes described as process-based and sometimes as hybrid, is perhaps one of the most well-known in forest science. Recently its creators were awarded the prestigious 2020 Marcus Wallenberg Prize. 3PG is a simple, process-based, generic stand-level model, originally developed by (Landsberg & Waring, 1997), with a deliberate attempt to bridge the gap between conventional growth and yield models (empirical based) and carbon balance models (process-based). This model has been calibrated and tested for different species across a wide range of forest types in different regions/countries and applied widely as research and operational tools (Gupta & Sharma, 2019). There are 119 published scientific articles (from 1997 – 2020, and still counting) on the application of the 3PG model (see https://3pg.forestry.ubc.ca/3pg-studies and Gupta & Sharma, 2019 for a detailed review). In Brazil, Aracruz Cellulose implements the 3PG model as an analytical and decision tool in their *Eucalyptus grandis* hybrid plantation at an operational scale (Almeida, 2003). In Australia, CSIRO uses 3PG for contracts ranging across site selection, forest productivity, water use assessment, and land amelioration (Landsberg & Sands, 2011). In South Africa, the model was incorporated into a decision support system to assess water yield and productivity of eucalypt plantations (Dye et al., 2002). The wide adoption of the 3PG model can be attributed to its simple structure, free source code, ability to incorporate remotely sensed data, and express objective to address questions of relevance to forest managers (Sands, 2004; Landsberg & Sands, 2011).

Despite the 3PG model's potential utility as a management decision support tool with scenario-based capability for predicting future growth and risk management in short rotation forestry, its application as an operational forest management tool in the South African forestry industry is still limited. The forestry industry's renewed interest in process-based models was sparked by the decrease in accuracy of traditional growth and yield models in predicting reliable stand productivity estimates, which may be attributed, at least in part, to more erratic weather patterns experienced in plantation growing areas. Furthermore, the development of 3PG into a spatially explicit model with scenario-based capability made it a very appealing option for investigating the potential impact of climate change on planted forests in support of adaptation strategies to ensure the industry's long-term sustainability.

However, in regions such as Southern Africa, where the network of weather stations is scarce and sparsely distributed (Lynch, 2004), process-based growth modelling will always be constrained by the availability of reliable meteorological data. Furthermore, due to the steep climatic gradient experienced at the coastal Zululand (du Plessis & Zwolinski, 2003; Louw et al., 2011), high-resolution meteorological data are needed to represent environmental variability in process-based modelling accurately. This is further discussed in Chapter 2.

1.4 General Study Area

Due to the availability of historical site data, a strong precipitation gradient, and similar genetics planted across sites, the important Zululand region was chosen for the study reported in this thesis. The data for this study were obtained from fast-growing *Eucalyptus grandis x urophylla (Egxu)* permanent sample plots (PSP) owned and managed by two forestry companies in South Africa: Mondi Forests (https://www.mondigroup.com) and Sappi (https://www.sappi.com). The PSPs are widely spread across the Zululand coastal plain, situated along the eastern seaboard of the KwaZulu-Natal region, South Africa (Figure 1.1 and 1.2). As well as 155 weather stations distributed across the province (Figure 1.3). KwaZulu-Natal province is in the southeastern part of South Africa, and it covers 94,360 km^2 or 7.7% of the country's total area. The province is bounded to the east by the Indian Ocean, and the topography ranges from sea level at the coast to over 3300 m along the Drakensberg escapement in the west. The slope is not gradual but rather features undulating terrain in steps. Because of these complex physiographic features, the province has a wide range of climatic conditions, ranging from a subtropical climate along the coast to a temperate climate inland (Ndlovu & Demlie, 2020). Precipitation increases from the inland to the coastal areas, also from North to South (du Plessis & Zwolinski, 2003). The province receives the most rainfall during the hot and humid summer months (November – February) (Ndlovu et al., 2021).

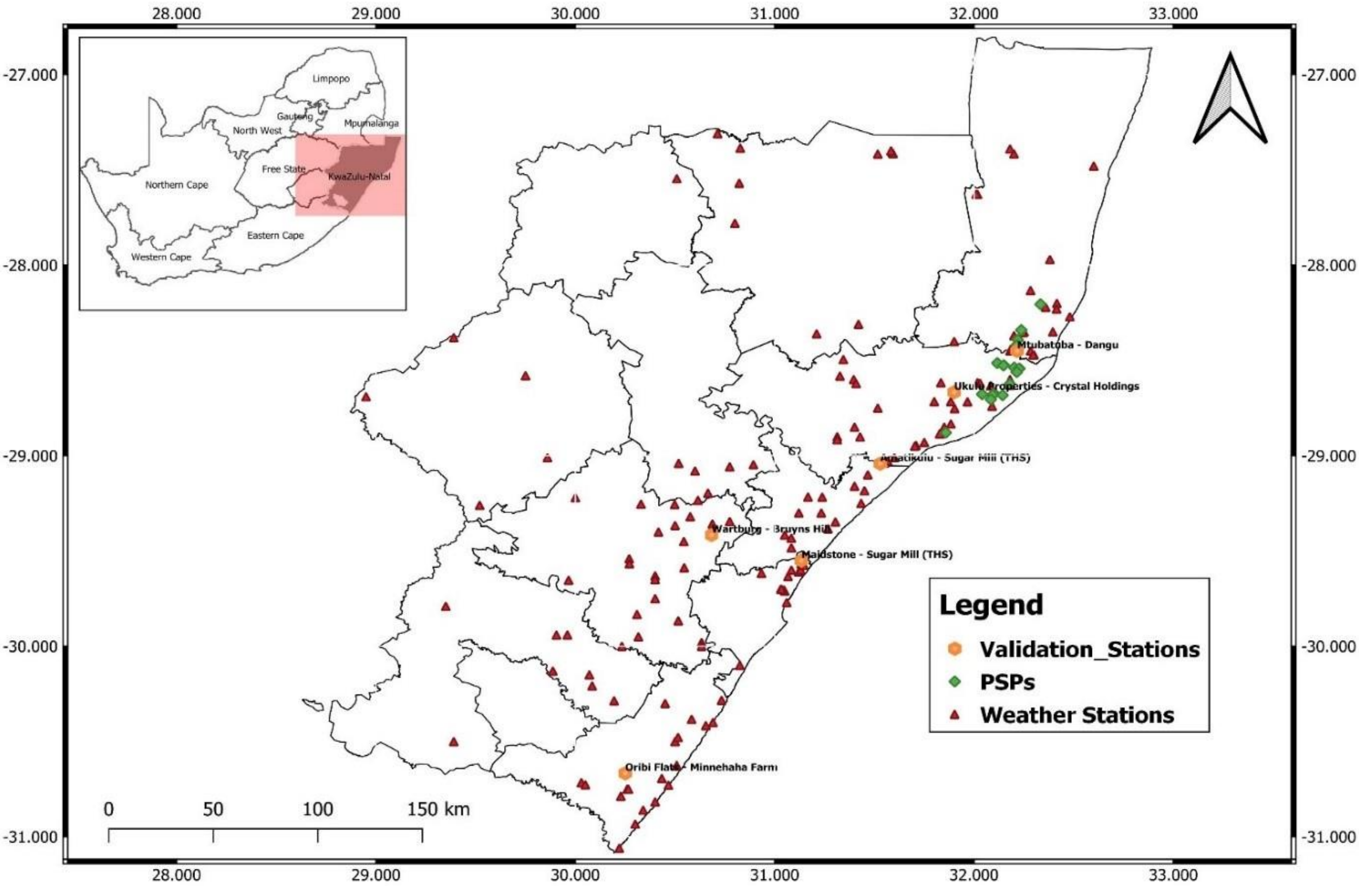


Figure 1.3 Map of the province of KwaZulu-Natal (inset showing location within South Africa) showing weather stations used to estimate weather data for the *Egxu* plantations.

(20)

1.5 Aim and Objectives

This study aimed to compare the performance of the 3PG model to other simpler growth modelling approaches in predicting the productivity of *Eucalyptus grandis x urophylla* across a range of sites in coastal Zululand, South Africa

To achieve this, the following specific objectives were set:

1. To select the best performing spatial interpolation technique for predicting weather data for the ungauged plantations using statistical error and indices.
2. To test, calibrate and validate the 3PG model for *Eucalyptus grandis x urophylla* species under South African conditions.
3. To run and validate three simpler models for the same region.
4. To compare the performance of these modelling approaches using statistical error and indices.

1.6 Key Research Questions

1. To what extent do spatial interpolation techniques accurately predict weather data for ungauged sites?
2. What parameter set gave the best prediction for *Egxu* species growing in South African conditions?
3. What was the difference in the performance of the 3PG model and other simple modelling approaches in predicting yield?

Chapter 2 Spatial Interpolation of Weather Data for Forestry Plantations in KwaZulu-Natal, South Africa

2.1 Introduction

In South Africa, water availability has been identified as the most significant limiting factor to forest growth (Dye, 2000; Edwards & Roberts, 2006). On average, the country receives approximately 460 mm/year of mean annual precipitation (MAP), but with an evaporative demand of 1400-3000 mm/year (Scott & Gush, 2017). Most commercial forestry plantations, however, are in a small region of the country that generally receives more than 750 mm/year MAP (Van Der Zel, 1995).

The Zululand region in KwaZulu-Natal is one of these regions, which has 66 885 ha planted to eucalypt species and has been a significant source of eucalypt fibre for many years (DAFF, 2019). This region has distinct precipitation and temperature gradient (as we move from North to South, and from the Coast to Inland) (du Plessis & Zwolinski, 2003; Louw et al., 2011). Due to this inherent spatiotemporal variability of precipitation, accurate information about the spatial distribution of precipitation is required to scientifically understand global or regional changes in water-related processes (Hu et al., 2019). They are an important input into process-based growth models such as 3PG for understanding the impact of climate change on forest growth, site classification, and monitoring the hydrological impacts of forest plantations (De Cáceres et al., 2018). As a result, real-time meteorological data corresponding to actual managed rotations rather than long-term mean climate data are critical for making practical and sound management decisions based on modelled estimates.

The South African forestry industry recognizes the importance of improving the prediction of stand growth and tree water use by using process-based models that are sensitive to changes in weather conditions, especially rainfall (Dye et al., 2004). Rainfall has always been the most crucial meteorological element measured in the South African weather stations, with records dating back to 1850 (Lynch, 2004). The number of active rainfall stations peaked at 3841 in 1938, then declined steadily, this trend becoming more pronounced after 1980 (Lynch, 2004). The decline impacted the spatial coverage of the rainfall monitoring network (Lynch, 2004), resulting in fewer stations being available. The issue of incomplete observational records and the spatial distribution of recording stations will have a negative impact on the model's output or limit its use (Jeffrey et al., 2001). This is important because high-resolution climate data are required to capture environmental variability, especially in areas with a steep climate gradient, such as South Africa (Hijmans & Parra, 2005).

Unfortunately, no "off-the-shelf" products like Australia's SILO resource (https://www.longpaddock.qld.gov.au/silo/) exist for South Africa. So, to be able to make use of interpolated weather (particularly rainfall) data across both space and time surfaces must be derived.

Several spatial interpolation methods have been used to generate gridded climate surfaces with varying spatial resolution to estimate point data at ungauged sites. For example, Fick & Hijmans (2017) developed 1 km spatially interpolated monthly climate data for global land areas known as "WorldClim", using thin-plate splines with covariates, The Climate Hazard Group (Funk et al., 2015) created CHIRPS (The Climate Hazard Group Infrared Precipitation with Stations), a 0.05 arc-degree daily precipitation dataset, using moving window regression and inverse distance weighting (IDW) techniques. In South Africa, Lynch (2004) interpolated annual, monthly, and daily rainfall data on a one-arc-minute gridded surface using geographically weighted regression (GWR). However, the limitation of these gridded datasets is that they are produced at a coarse spatial resolution and can hardly capture the high level of spatiotemporal variability of rainfall locally. Therefore, better point estimates of climate data are required.

Many spatial interpolation methods are available, and numerous studies have been conducted to determine the "best" of these (Goovaerts, 1999; Apaydin et al., 2004; Li & Heap, 2014; Chen & Guo, 2017). They conclude, however, that there is no optimal method for all circumstances and that each method has its strengths and weakness (Chen & Liu, 2012; Hu et al., 2019). Some factors that may influence the performance of any chosen spatial interpolation method include sampling design and spatial distribution of samples, data nature and quality, a correlation between primary and secondary variables, and factor interaction (Li & Heap, 2014). It is thus strongly advised to choose the appropriate interpolation methods based on the application objective, the geographic and gauge conditions of the study area, as well as the spatiotemporal scales (Hu et al., 2019).

The main objective of this chapter is to describe the approach to generate a set of reliable point estimates for growth modelling. This was done specifically to address the difficulties in obtaining high-resolution climate data for ungauged plantations by evaluating and comparing the accuracy of two interpolation methods, Random Forest (RF) and an R package "Meteoland", in predicting rainfall data from 2008 to 2018. The techniques chosen were based on the findings of a similar study conducted in KwaZulu-Natal, South Africa, by Burengengwa (2020).

2.2 Materials and Method

2.2.1 Data Sources

Observed long-term daily weather data such as maximum temperature, minimum temperature, precipitation, and solar radiation were obtained from the South African Sugarcane Research Institute (SASRI) and the South African Weather Services (SAWS) from January 2008 to September 2018. One hundred fifty-five (155) weather stations were obtained from these two databases: one hundred and eight (108) from SASRI and forty-seven (47) from SAWS. The spatial coordinates, longitude, and latitude of the weather stations were available from the datasets. In this study, the covariables used for modelling were aspect, elevation, slope, and distance from the ocean. Aspect and slope were derived from the GISCOE 20 m Digital Elevation Model (GISCOE, 2001) raster data, using the ArcGIS tools 'Aspect' and 'Slope' in the Spatial Analysis toolbox. The distance from the ocean was calculated from the polyline of the African continent using the Nearest Neighbor Join (NNjoin) plugin tool in QGIS.

2.2.2 Preparing Input File

Given the differences in the file format of the weather data retrieved from each weather station database, as well as the input file format required by the two models, it was necessary to create a pipeline in *R* software for efficient and effective data handling, processing, and preparation of input files used for modelling. The pipeline was designed to handle and report missing dates, missing values, duplicate data, name mismatches, and generate input files for both models as output. The complete R scripts and template file for cleaning and preparing input files are available on GitHub (https://github.com/EucXylo/Random_Forest_weather_prep).

2.3 Spatial Interpolation Models

Spatial interpolation is defined as predicting the values of a primary variable at points within the same region of sampled location (Li & Heap, 2014). A variety of approaches to spatial interpolation have been developed. They are broadly classified as 1) geostatistical methods, this includes Kriging and its derivatives such as Simple kriging (SK), Ordinary kriging (OK), Indicator kriging (IK), 2) non-geostatistical methods, e.g., Nearest Neighbors (NN), Inverse Distance Weighting (IDW), Thin-Plate Splines (TPS), Thiessen polygons, and 3) combined methods, e.g., Regression kriging, Trend surface analysis combined with kriging, Gradient plus inverse distance square (see Li & Heap, (2014) for a list and description of methods). Furthermore, some machine learning methods, such as random forest (RF), neural network, boosted decision tree (BDT), and so on, have been recently introduced (Li & Heap, 2014).

2.3.1 Random Forest

The application of machine learning (ML) techniques such as the Random Forest (RF) algorithm is becoming more common in spatial interpolation. They are used in many fields, including climatology, geology, land use mapping, spatial planning, and soil science (Sekulić et al., 2020). For example, Youssef et al. (2015) and Chen et al. (2017) reported RF to be the best performing spatial interpolation technique for landslide susceptibility mapping. The RF technique was also used by Sekulić et al. (2021) to create MeteoSerbia1km (the first daily gridded meteorological dataset with a 1 km spatial resolution in Serbia). Leirvik & Yuan (2021) discovered that RF outperforms conventional interpolation techniques for interpolating solar radiation observations.

Random Forest is an effective ensemble-learning method developed by (Breiman, 2001). Its approach is based on the general principle of creating multiple random decision trees from a training dataset (via ***bootstrapping***) and then aggregating the output generated by each decision tree (known as ***bagging***, an acronym for **B**ootstrap **agg**regat**ing**) (Genuer et al., 2010; Genuer & Poggi, 2020). Bagging has proven beneficial for decision trees by reducing the high variance of individual trees and combining them into a single process (James et al., 2013; Sekulić et al., 2020). The bagged trees must be as diverse as possible to achieve good predictive performance because ensembling a set of very similar predictors would result in a similar predictor (Genuer & Poggi, 2020).

Therefore, RF reduces correlation in bagged trees by selecting random samples of *m explanatory variables* from the entire set of *p predictors* each time a split is considered in decision tree construction (see James et al. (2013) for detailed explanation). The number of trees (*ntree*) used in the forest and the number of explanatory variables (*mtry*) used in each tree are two important parameters in RF that can be tuned to improve the final model accuracy (Breiman, 2001; Youssef et al., 2015). The *randomForest* package (Liaw & Wiener, 2002) in R software (R Core Team, 2021) was used for RF modeling.

2.3.1.1 Input

Monthly averages for precipitation, minimum temperature, maximum temperature, and solar radiation for each year (2008 – 2018), temporal stamps (month of observation), spatial coordinates (latitude and longitude), and co-variables (slope, elevation, aspect, distance from coast) are all input variables. The spatial coordinates, co-variables, and temporal stamps were used as the explanatory variables. The original dataset was randomly divided into two parts for the development of the random forest model: *training set (70%) and testing set (30%)*. In the pipeline (section 2.2.2), rainfall months with missing values were returned as NA, therefore stations with NA values for a specific month were excluded when developing the RF model.

2.3.1.2 Parameter Calibration

The *Random Search* method was used to tune *mtry* (the number of explanatory variables randomly sampled as candidates at each split) using *10-fold cross-validation*. The *Grid Search* method was used to tune the number of trees (*ntree*) to obtain the optimized *mtry* and *ntree* values for each year data. The tuning of these parameters was repeated for each year because of the number of station observations.

2.3.2 Meteoland

Thornton et al. (1997) proposed a method that is both an inverse-distance algorithm and a smoothing filter. The method aims to overcome the limitation of the inverse-distance method, which produces surfaces with the spatially anomalous distribution. As a result, the Gaussian filter was developed to reduce the number of observations used in predictions at a given point. However, *Meteoland*, an R package (De Cáceres et al., 2018) to implement the daily weather interpolation and estimation algorithms introduced by Thornton et al. (1997), will be used for this study.

The Meteoland package estimates the following daily surface weather variables: mean, maximum, and minimum temperature, precipitation, mean, maximum and minimum relative humidity, incident solar radiation, wind speed, and wind direction. This approach defines spatial weights W(r) at radial distance r from a target point using:

$$\mathbf{W(r)} = e^{-\alpha(r/R_p)^2} - e^{-\alpha}$$

Equation 2. 1

(26)

If r < R_p and W(r) = 0 otherwise. Where r is the radial distance from p, R_p is the truncation distance, and α is the shape parameter. This filters spatial convolution with a set of weather station locations produces a vector of weights associated with observations for each target point (De Cáceres et al., 2018). R_p is automatically adjusted to be smaller in data-rich areas and larger in data-poor areas. The estimation of R_p is based on N, the average number of observations to be included for each target point (see De Cáceres et al. (2018) for detailed calculations). Additionally, well-detailed documentation of Meteoland calculation routines, ***the meteoland reference book***, is available online via the GitHub repository (www.https://emf-creaf.github.io/meteolandbook/index.html).

2.3.2.1 Input

For the interpolation of daily weather in the *Meteoland* package, the following text files (.txt) were generated from the pipeline in section 2.2.2. For each ***weather variable*** (precipitation, minimum temperature, maximum temperature, solar radiation), a text file containing stations in rows and dates in columns, ***station info*** file containing latitude and longitude for each station, ***station topography*** file containing elevation, slope and aspect, and the ungauged site's spatial coordinate and topography.

2.3.2.2 Parameter Calibration

In this study, the ***initial_Rp*** (default = 140 km), which specifies the initial radius for the truncated spatial Gaussian kernel, was reduced to ~11 km using the ***trace (defaultInterpolationParams, edit = TRUE)*** function in the package. The interpolation parameters α (shape parameter) and N (average number of stations to be used) vary for each variable to be interpolated. Therefore, calibration was performed using the ***interpolation.calibration()*** function by specifying a sequence of ***N = seq(3, 5, by = 1)*** and ***α = seq(0.5, 10, by = 0.5)*** for minimum and maximum temperature, precipitation amount and precipitation event. The ***f_max*** parameter used in the estimation of precipitation was also calibrated using the same function with ***fmax_seq = seq(0.05, 0.95, by = 0.05)***. The calibration function returns a set of results. The most important are the minimum mean absolute error value, N, and alpha values corresponding to the minimum MAE. This calibration exercise was performed for each year following the recommendation that it should be performed more than once if the number of stations available differed temporally.

2.4 Performance Assessment

The RF algorithm performs model assessment checks using an unbiased estimation of generalization error known as out-of-bag (OOB) error (Breiman, 2001). The Meteoland package assesses its performance using the leave-one-out cross-validation method (De Cáceres et al., 2018). However, this study used a pairwise comparison of model-predicted and observed monthly rainfall data from six weather stations (validation stations) (Figure 1.3) between 2008 and 2018 for the performance analysis. Due to the inherent spatiotemporal variability of rainfall, which makes interpolation difficult, only rainfall data was used to evaluate the performance of both models. To observe the general precipitation pattern of the ungauged plantations along the Zululand coastal plain, the total cumulative rainfall from January 2008 to September 2018 was divided into three categories: dry (8101 mm – 9795 mm), medium (9795 mm – 10232 mm), and wet (10232 mm – 11774 mm) as shown in Figure 6.

The following statistical errors and indices from the Agricultural and Meteorological software (AgriMetSoft, 2019) were used to compare the predicted and observed precipitation data: mean absolute error (MAE), root mean square error (RMSE), mean bias error (MBE), Willmott index of agreement (d), coefficient of determination (R^2), and Nash Sutcliffe model efficiency index (E).

2.4.1 Statistics

Mean Absolute Error (MAE): This is the average absolute error between estimated and observed values without considering their direction.

$$\text{MAE} = \frac{\sum_{i=1}^{n} |o_i - p_i|}{n}$$

Equation 2. 2

Root Mean Square Error (RMSE): This is the standard deviation of the residuals. It is a measure of how spread out these residuals are. Therefore, it tells how concentrated the data is around the line of best fit.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (o_i - p_i)^2}{n}}$$

Equation 2. 3

Mean Bias Error (MBE): It captures the average bias in the prediction. Bias is the tendency of a statistic to overestimate or underestimate a parameter.

$$\text{MBE} = \frac{\sum_{i=1}^{n} (o_i - p_i)}{n}$$

Equation 2. 4

Willmott Index of Agreement (d): Willmott (1981) developed this index to compensate for the insensitivity of the R^2 to additive and proportional differences between observed and predicted means and variance (Legates & McCabe, 1999; Dawson et al., 2007; Moriasi et al., 2008). This index is a standardized measure of the degree of model prediction, which varies between 0 and 1. It represents the ratio of the mean square error and the potential error. One indicates a perfect match, and zero indicates no agreement at all.

$$\mathbf{d = 1 -} \frac{\sum_{i=1}^{n}(o_i - p_i)^2}{\sum_{i=1}^{n}(|p_i - o| + |o_i - o|)^2}$$

Equation 2. 5

Coefficient of Determination (R^2): This describes the degree of collinearity between predicted and observed data (Moriasi et al., 2008). This metric describes the proportion of the total variance in the observed data that the model can explain. It ranges from 0 (poor agreement) to 1 (perfect agreement). It is calculated as the square of the Pearson correlation coefficient (r), given as

$$\mathbf{r =} \frac{n(\sum(o_i p_i) - (\sum o_i)(\sum p_i)}{\sqrt{[n\sum o_i{}^2 - (\sum o_i)^2][n\sum p_i{}^2 - (\sum p_i)^2}}$$

Equation 2. 6

Model Efficiency Index (E): Also known as Nash-Sutcliffe Efficiency, was developed by Nash & Sutcliffe (1970). It is a normalized statistic that determines the relative magnitude of the residual variance compared to the observed data variance (Moriasi et al., 2008). This index is widely used to evaluate the performance of hydrological models. It also presents an improvement over the coefficient of determination (Legates & McCabe, 1999). It ranges from $-\infty$ to 1, where E = 1 indicates perfect model, $0 > E < 1$ are generally acceptable performance, and $E \leq 0$ indicates unacceptable performance, which means the mean observed value is a better predictor than the simulated value.

$$\mathbf{E = 1 -} \frac{\sum_{i=1}^{n}(o_i - p_i)^2}{\sum_{i=1}^{n}(|o_i - o|)^2}$$

Equation 2. 7

Where, o_i = observation values

p_i = predicted values

o = average observation value

n = number of observations

2.5 Results and Discussion

2.5.1 Performance analysis

There was a good agreement between the observed and predicted precipitation trends, indicating adequate calibration of both models over the range of measured precipitation (Figure 2.1 and 2.2) (Singh et al., 2004). When rainfall was totalled over the time studied, Random Forest predicted-rainfall data was found within the range of the observed cumulative rainfall (and standard deviation) compared with the Meteoland model (Table 2.1).

Overall, the RF model exhibited lower error when compared to Meteoland (Figure 2.3). Generally, the RMSE and MAE values are considered low if they are less than the observed values' standard deviation (Singh et al., 2004). Therefore, predictions by both had a low level of errors (Table 2.1 and 2.2). There were some tendencies towards bias. The MBE statistic measures the average tendency of the models to overestimate (MBE > 0) or underestimate (MBE < 0) the observed data. Based on MBE and evident visually, the RF-based model performed better in this regard than Meteoland (Table 2.2; Figure 2.2). In general, both models overestimated annual rainfall data for the Oribi – flat weather station. For the Wartburg station, Meteoland underestimated rainfall data from 2009 – 2016, while RF underestimated from 2015 – 2017 (Figure 2.2). The relative stability of RF's performance can be explained by its use of the bagging (bootstrap aggregation) scheme, which reduces prediction error variance and improves accuracy (Biau & Scornet, 2016; Sekulić et al., 2020).

The dimensionless statistical indexes used to evaluate the performance of both models (Coefficient of Determination (R^2), Willmott Index of Agreement (d), and Model Efficiency Index (E)) confirmed that both models could be used to generate good predictions. However, RF slightly outperformed Meteoland (Figure 2.3). RF and Meteoland produced high levels of correlation R^2 (>0.80). According to Moriasi et al. (2008), values greater than 0.5 are considered as acceptable. However, this statistic has been reported to be limited for two reasons (1) R^2 is overly sensitive to outliers, which can mask the true overall relationship (2) R^2 is insensitive to additive and proportional differences between the observed and predicted values, allowing for high scores even when the predicted values are significantly different from the observed values in terms of magnitude and variability (see Legates & McCabe (1999) and Dawson et al. (2007)). These limitations can be seen in the graphical results (Figure 2.4 and 2.5), where RF was penalized for one extreme value in the **Oribi Flats – Minnehaha Farm** weather station (R^2 = 0.756) and Meteoland has good R^2 values for stations where they overestimated and underestimated but followed the observed data.

(30)

The index of agreement (d) values for both models were high (>0.90). Also, both models had very good (E) values, but the RF model (E = 0.84) outperforms the Meteoland model (E = 0.79). This demonstrates that, despite the inherent spatiotemporal variability of rainfall, both models were satisfactorily calibrated to simulate rainfall data for the period under consideration in this region.

2.5.2 Interpolated rainfall data

For growth simulation, it was important to understand the spatiotemporal variability of rainfall along the Zululand coastal plain, where the ungauged plantations are located. Two models were tested for this purpose, using different techniques of achieving optimal spatial interpolation. An important outcome of the work was to ensure that the coastal areas where subsequent growth modelling was to be done would be objectively accurately estimated at a broad scale. Both models capture a distinct pattern of wetness from north to south (Figure 2.6). However, the pattern that also exists from coast to inland was better captured by RF and is consistent with previous research (du Plessis & Zwolinski, 2003; Louw et al., 2011; Ndlovu & Demlie, 2020; Ndlovu et al., 2021). This can be explained by the negative correlation between distance to the coast and weather patterns (du Plessis & Zwolinski, 2003; Burengengwa, 2020).

Overall, the results show that both models can predict long-term rainfall in the study area reasonably well. In this study, the RF model performed best, which is consistent with previous research that compared the performance of RF with other interpolation methods (Li et al., 2011; Chen et al., 2017; Burengengwa, 2020; Leirvik & Yuan, 2021). Nevertheless, the Meteoland package in R does provide a novel spatial interpolation technique that has shown potentials in predicting daily weather data (Germishuizen, 2018; Karavani et al., 2018; Sánchez-Pinillos et al., 2018). Aside from the preliminary work of Burengengwa (2020), no other research to the author’s knowledge has evaluated and compared the performance of this technique using real-date weather data. Furthermore, both RF and Meteoland meet our requirements of employing a simple model that uses readily available data, is available in R software, and is easily integrated into the forest simulation model used in this study, which is also available in R software.

Table 2.1 Cumulative observed and predicted rainfall from January 2008 to September 2018

Station	Observed sum (SD), mm	RF sum (SD), mm	ML sum (SD), mm
Amatikulu -Sugar Mill	8761.9 (63.27)	10181.3 (62.44)	10828.5 (74.55)
Maidstone - Sugar Mill	10999.4 (71.18)	10025.3 (61.62)	9993.7 (65.59)
Mtubatuba -Dangu	8987.6 (56.11)	8803.7 (53.32)	8484.4 (58.88)
Oribi Flats - Minnehaha Farm	8292.3 (61.60)	10501.1 (60.40)	11194.7 (81.91)
Ukulu Properties - Crystal Holdings	8312.1 (55.16)	8540.3 (51.50)	10093.0 (61.90)
Wartburg - Bruyns Hill	9715.2 (55.40)	8783.5 (49.73)	8479.6 (50.69)

RF – Random Forest; ML – Meteoland; SD – Standard deviation

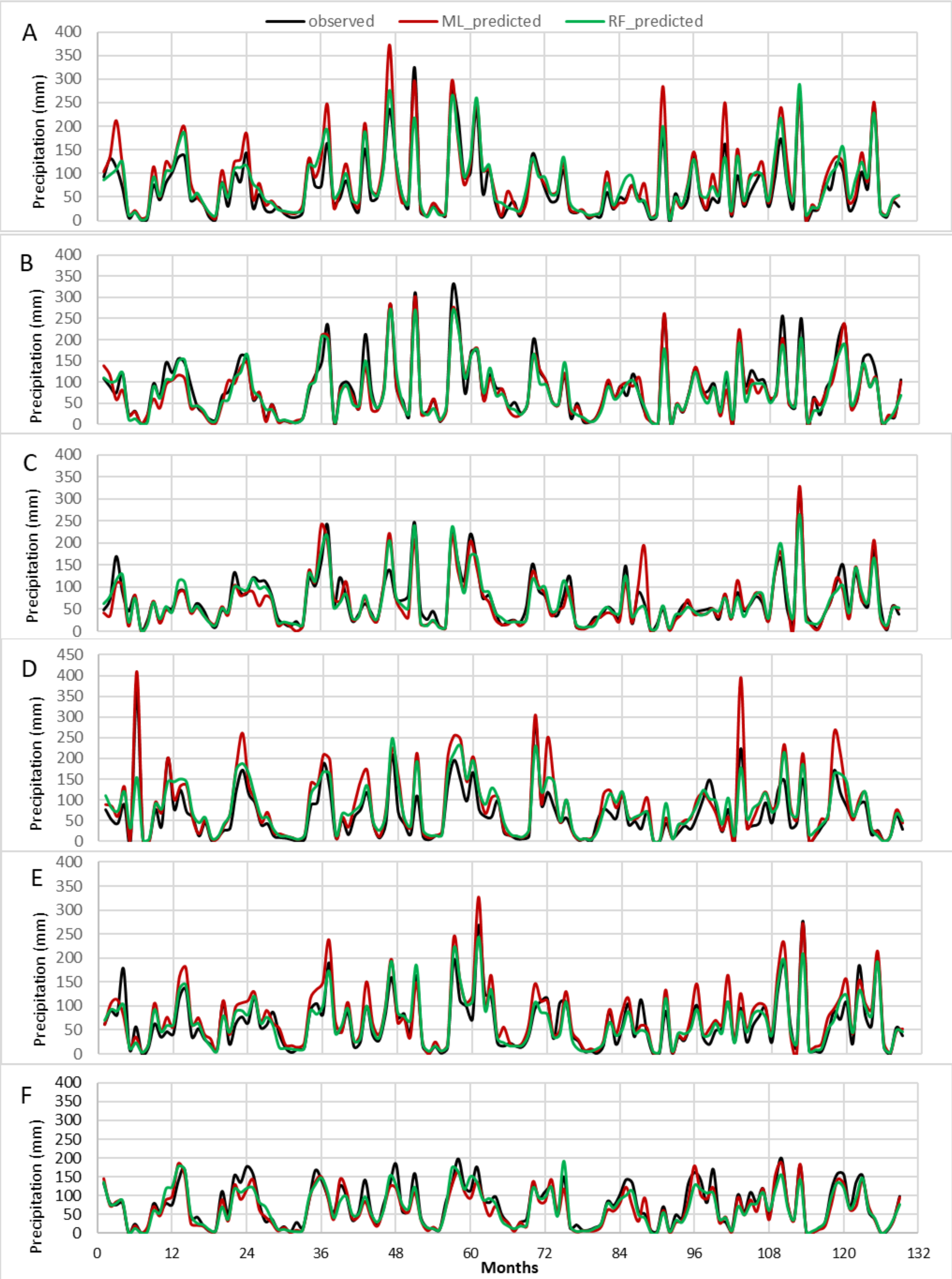


Figure 2.1 Monthly time-series of the observed and model-predicted rainfall from 2008 -2018. A= Amatikulu – Sugar Mill, B= Maidstone – Sugar Mill, C= Mtubatuba – Dangu, D = Oribi Flats, E= Ukulu Properties, F= Wartburg- Bruyns Hill. Prefix RF = Random Forest, ML = Meteoland.

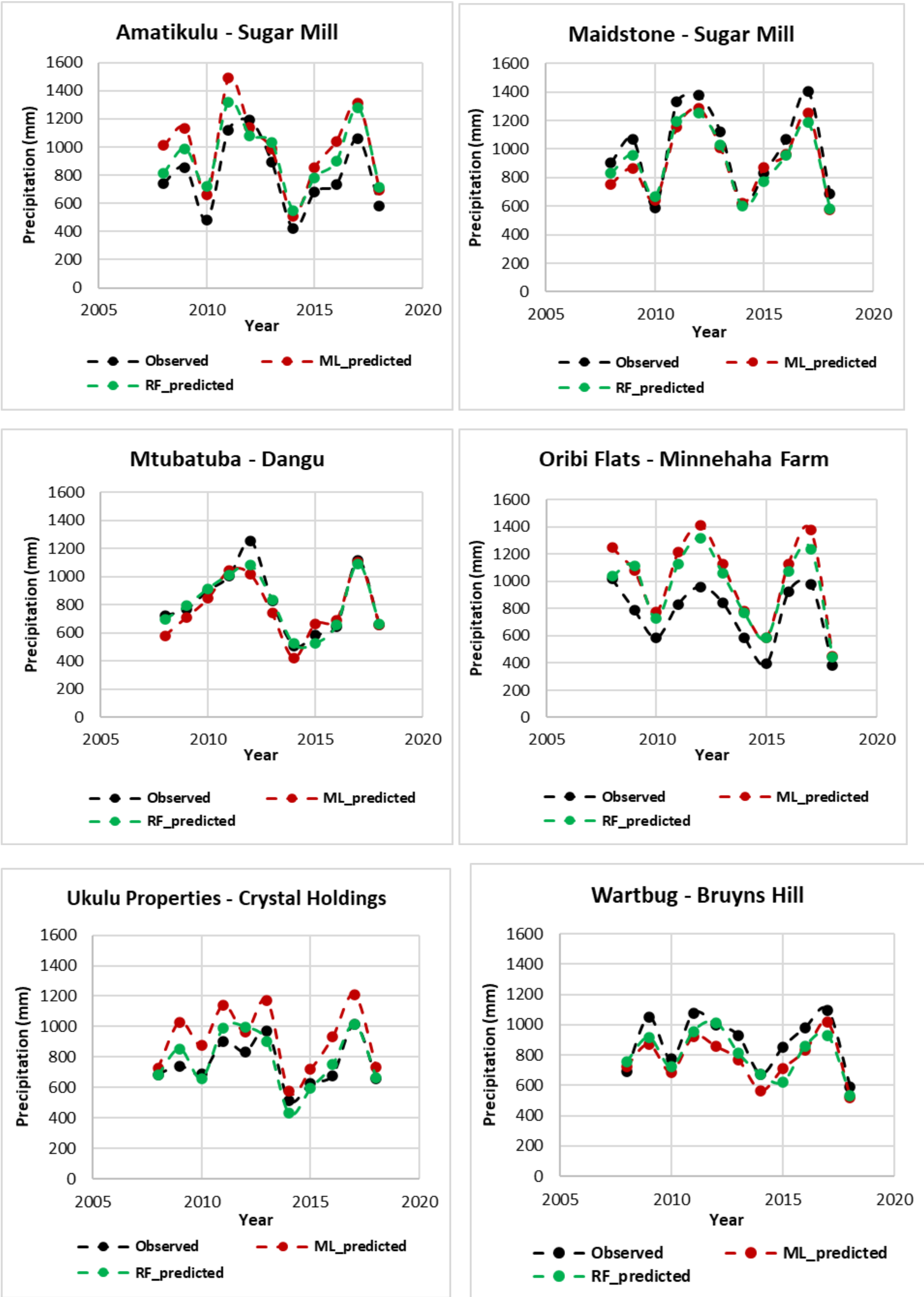
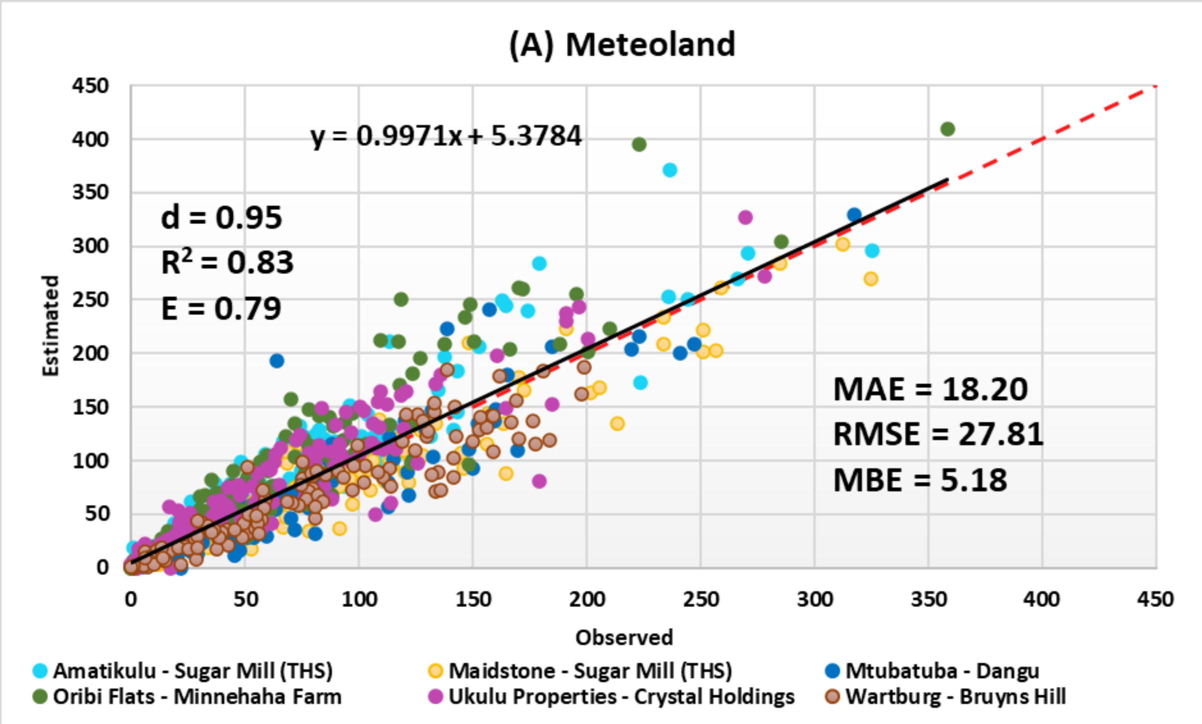


Figure 2.2 Annual time-series of the observed and model-predicted rainfall from 2008 – 2018 for each validation stations. Prefix ML = Meteoland, RF = Random Forest.



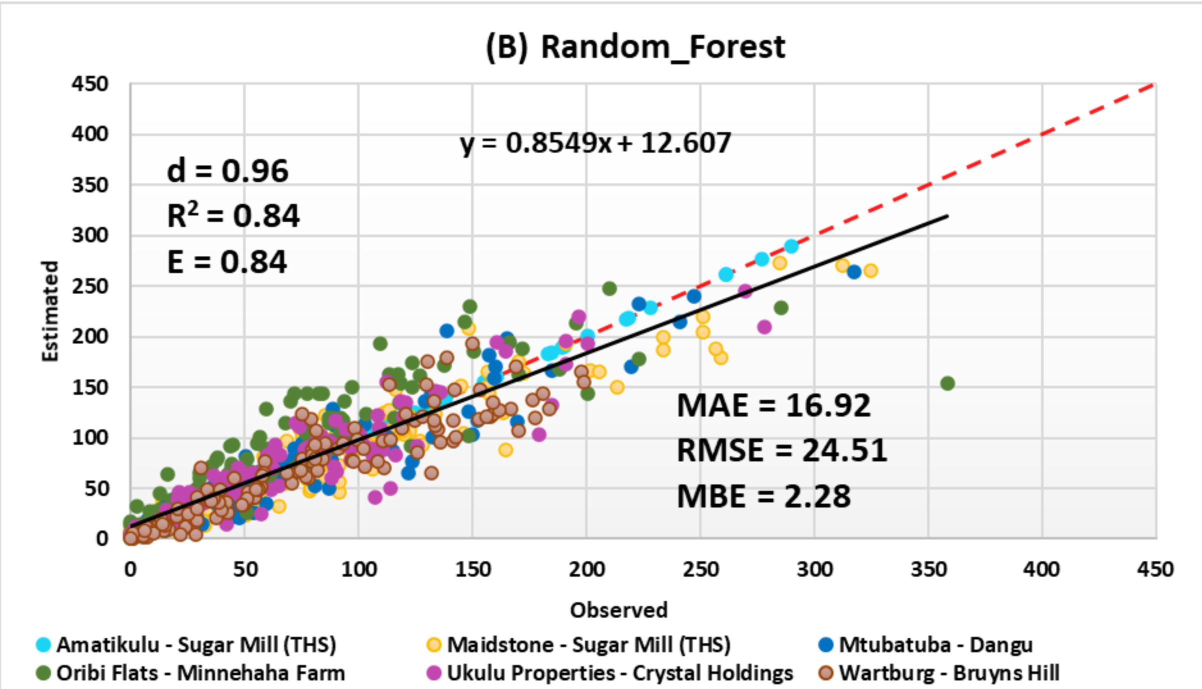


Figure 2.3 Comparison of observed and model-predicted monthly rainfall for all six validation stations by (A) Meteoland and (B) Random Forest. Colors represent different weather stations. d = Index of Agreement, R^2 = Coefficient of Determination, E = Model Efficiency Index, MAE = Mean Absolute Error (mm), RMSE = Root Mean Square Error (mm), MBE = Mean Bias Error (mm).

Table 2.2 Statistical evaluation of models for each station's rainfall prediction

Station	Statistics	Random Forest	Meteoland
Amatikulu – Sugar Mill	**MAE**	17.24	19.46
	RMSE	23.81	30.10
	MBE	11.00	16.02
	d	0.96	0.95
	R^2	0.89	0.89
	E	0.86	0.77
Maidstone – Sugar Mill	**MAE**	15.98	14.31
	RMSE	23.08	22.05
	MBE	-7.55	-7.80
	d	0.97	0.97
	R^2	0.91	0.92
	E	0.89	0.90
Mtubatuba - Dangu	**MAE**	13.48	14.67
	RMSE	18.87	23.72
	MBE	-1.43	-3.90
	d	0.97	0.96
	R^2	0.89	0.84
	E	0.89	0.82
Oribi Flats – Minnehaha Farm	**MAE**	25.13	25.36
	RMSE	35.47	38.91
	MBE	17.12	22.50
	d	0.91	0.93
	R^2	0.76	0.88
	E	0.67	0.60
Ukulu Properties – Crystal Holdings	**MAE**	14.20	20.42
	RMSE	20.03	26.81
	MBE	1.77	13.81
	d	0.96	0.95
	R^2	0.87	0.86
	E	0.87	0.76
Wartburg – Bruyns Hill	**MAE**	15.48	14.97
	RMSE	22.10	21.29
	MBE	-7.22	-9.58
	d	0.95	0.96
	R^2	0.86	0.88
	E	0.84	0.85

MAE – mean absolute error; RMSE – root mean square error; MBE – mean bias error; d – Willmott index of agreement; R^2 – coefficient of determination; E – Model Efficiency Index.

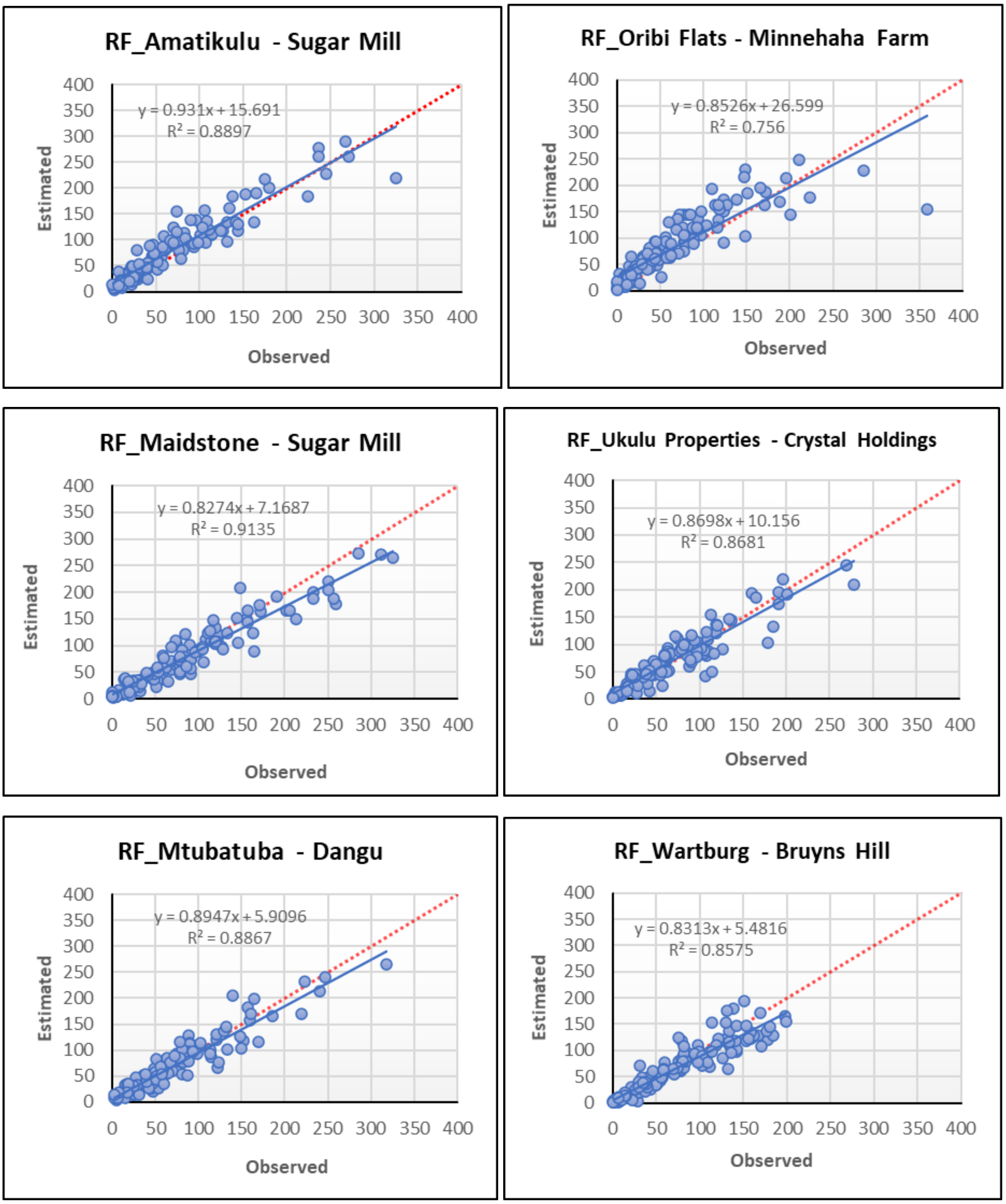


Figure 2.4 Comparison of observed and model-predicted monthly rainfall for all six validation stations. Prefix RF = Random Forest.

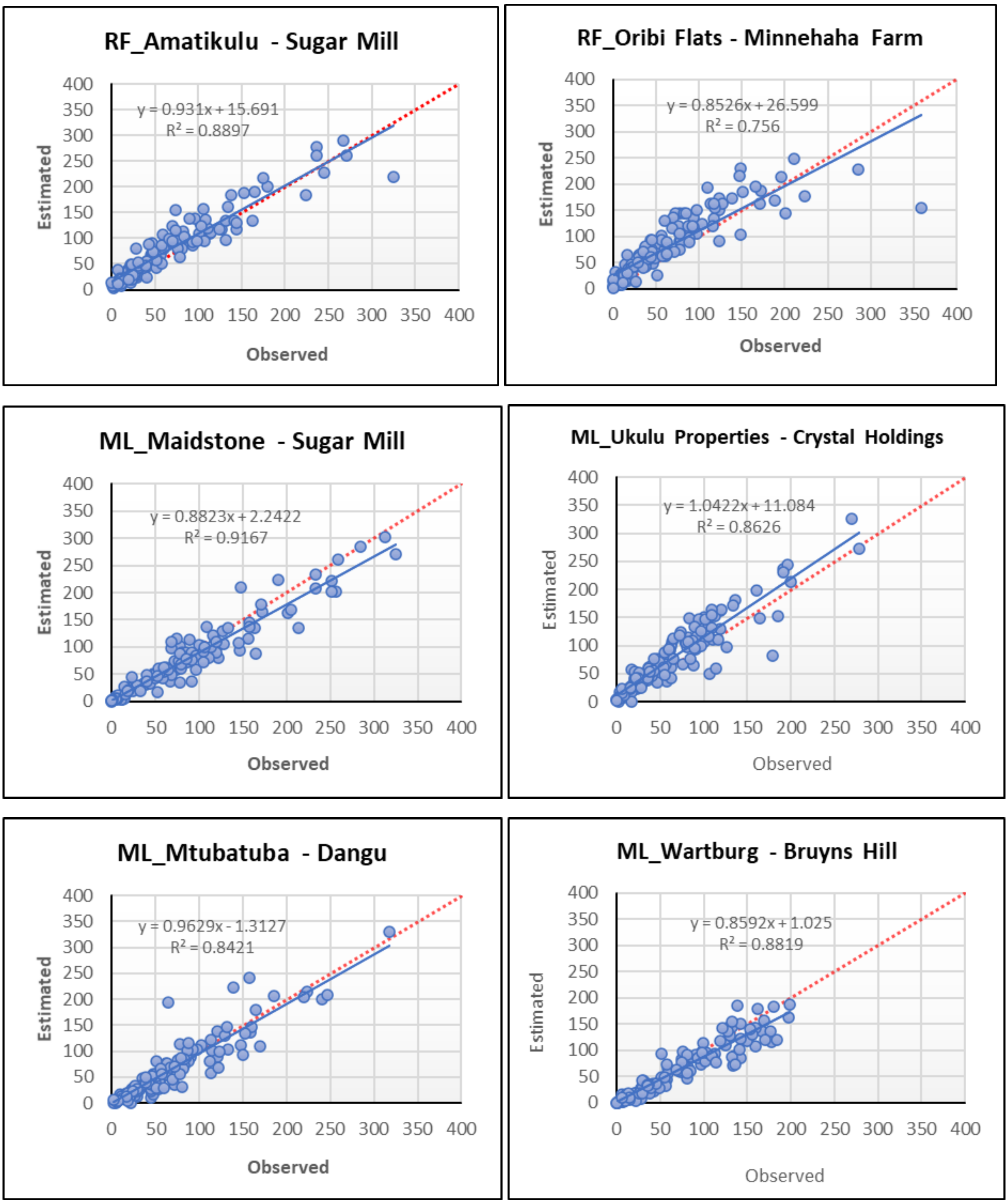
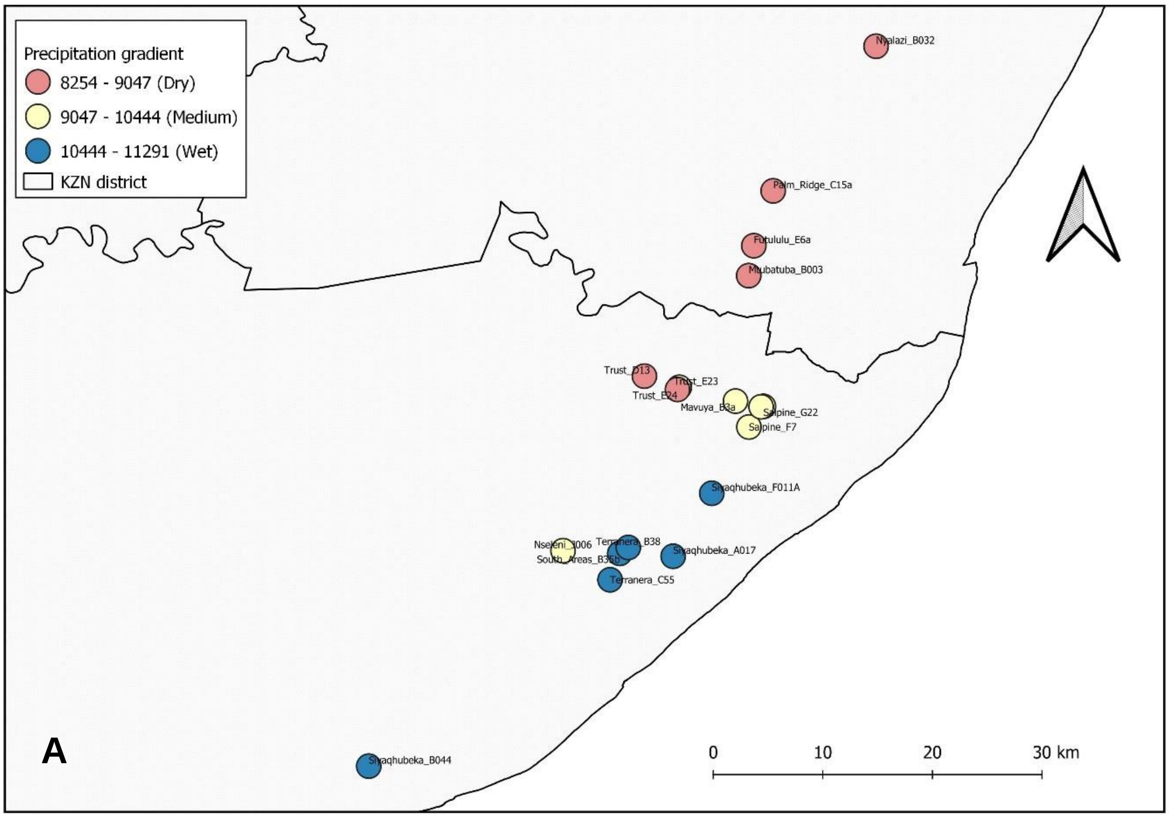


Figure 2.5 Comparison of observed and model-predicted monthly rainfall for all six validation stations. Prefix ML = Meteoland.



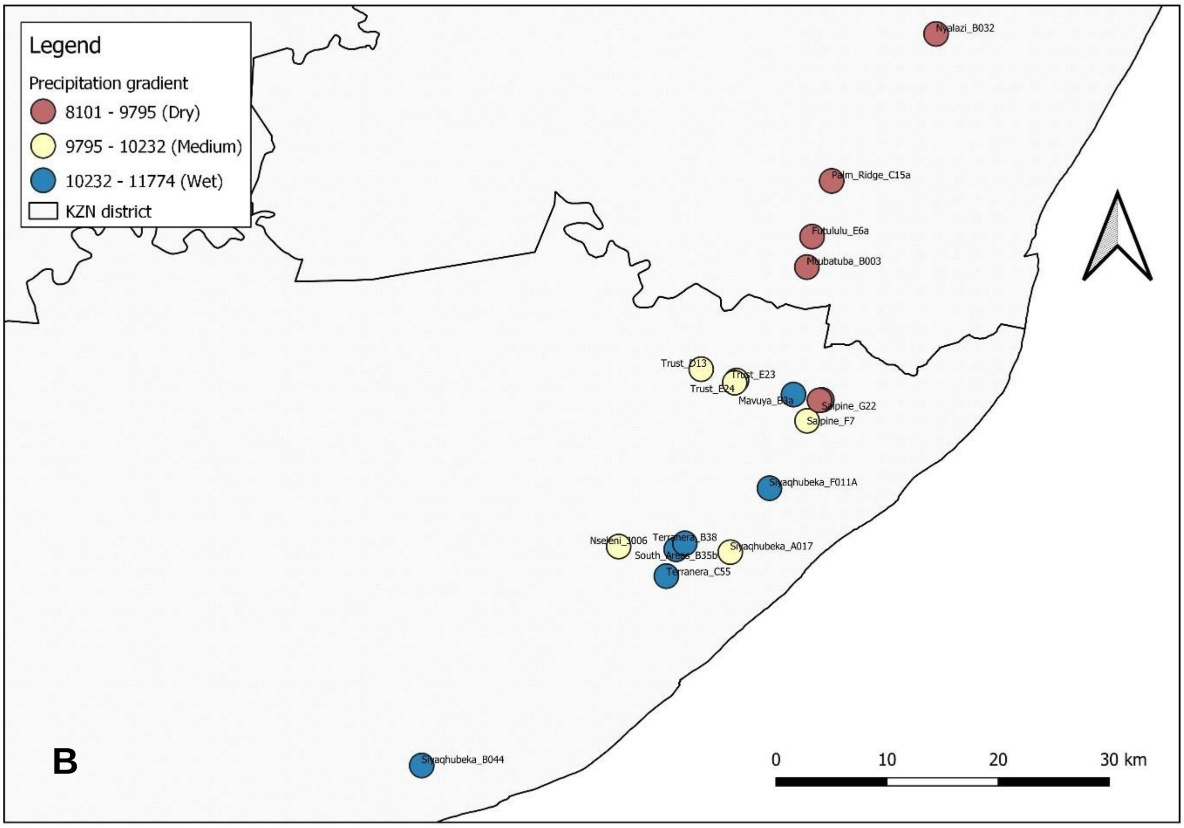


Figure 2.6 Precipitation gradient of interpolated rainfall for the 18 PSP located along Zululand Coastal plain by (A) Random Forest (B) Meteoland.

Chapter 3 Calibration of the 3PG model for *Eucalyptus grandis* x *urophylla* growing under South African conditions.

3.1 Introduction

The structure of the 3PG model is based on well-established principles underlying plant growth and development, including expressions of physiological concepts as light absorption, photosynthesis, water balance, and carbon allocation (Landsberg, 2003; Weiskittel et al., 2011; Forrester et al., 2021). The approach to estimation in 3PG, while process-based, is simple (Landsberg & Waring, 1997). Although the model structure is generic, it must be parameterized for individual species (Sands & Landsberg, 2002), distinguishing in simulation runs between the physiological responses of different species (Landsberg & Sands, 2011). The physiological measurements required to calculate all these parameters can be difficult to obtain or even unavailable (Forrester et al., 2021). This may result from lack of data, problems of scaling up, or poor understanding due to the process being inaccessible for measurement (deep water/nutrient uptake), or too complex to measure (Mäkelä et al., 2000b). Therefore, apart from calibrating the model using parameter values from physiological measurements, they can also be parameterized to address uncertainty in the model's behavior (Forrester et al., 2021).

According to Landsberg & Sands (2011), *"calibration is the process of assigning parameter values by direct measurements in an independent experiment, whereas parameterization is the process of estimating parameter values by adjusting their values to minimize the sum of squares of the residual between the observed and predicted data"*. Both processes are greatly aided by a sound understanding of the model, its parameter space, and knowledge of the sensitivity of its outputs to species-specific parameters (Sands, 2004). Guidelines and procedures for assigning and estimating parameter values are described in detail by Sands & Landsberg (2002), Landsberg et al. (2003), and Sands (2004).

Several studies have published parameter values calibrated and parameterized for various species in different regions (see Gupta & Sharma, (2019)). In many of these studies, parameter estimation was optimized either manually (stepwise adjustment) (Sands & Landsberg, 2002; Landsberg et al., 2003; Almeida et al., 2004) or automatically using estimation software such as *PEST*$_{XL}$ (Esprey, 2006), statistical fitting using the SAS software (Gonzalez-Benecke et al., 2016), or Bayesian calibration using the *BayesianTools* R package (Forrester et al., 2021).

Although the 3PG model has been parameterized for *Eucalyptus grandis x urophylla* in Brazil (Stape et al., 2004; Almeida et al., 2004; Borges et al., 2012), concerns have been raised about the generality of these parameters and the accuracy of model predictions in other regions (like South Africa). Therefore, the objective of this chapter was to parameterize the 3PG model for *Eucalyptus grandis x urophylla* in South Africa using published parameter values from the Brazilian sites and validate it using independent growth data. We also explore a new goodness-of-fit approach called *"extended RMSE"* (*eRMSE*) to optimize parameter values.

3.2 Description of the 3PG model

The 3PG model (an acronym for Physiological Processes Predicting Growth) is a simple, process-based, stand-level model, initially designed for monospecific, even-aged, and evergreen forest (Landsberg & Waring, 1997), but which has since been further developed for deciduous, uneven-aged and mixed-species forests (Forrester & Tang, 2016). The model runs on a monthly time-step, and data required to run the 3PG model can be divided into four classes.

- **Weather data:** Inputs are monthly averages of mean minimum and maximum air temperature (°C), monthly averages of solar radiation (MJ m^{-2} d^{-1}), monthly total rainfall (mm/month), monthly averages of atmospheric vapor pressure deficit (mbar), and the number of frost days in a month (days per month).
- **Site Information:** site-specific information describing the physical properties of the site includes latitude, soil texture, atmospheric CO_2 (ppm), available soil water (minimum and maximum) (mm/m), and a simple fertility rating.
- **Stand initialization data:** In 3PG, stands can be initialized at a selected age, and data required include initial stocking (t/ha), initial stem, root, and foliage biomass (t_{DM}/ha), and the initial available soil water (mm) at the initial age (years). Suppose a stand is initialized at planting age (age = 0): in this case, it is appropriate to use typical seedling weight, where the biomass of the seedling is divided among the biomass pools, or to use default values provided by Sands & Landsberg (2002).
- **Species-specific parameters:** The main 3PG parameters consist of 6 major parameter classes: biomass partitioning and turnover, Net Primary Productivity (NPP) & conductance modifiers, stem mortality, and stand characteristics. However, new parameters were added by Forrester & Tang (2016) to enable the model to work for mixed-species forests, deciduous species, and forests where the stand density is reduced. This is beyond the scope of this study and will not be considered.

The model is divided into five sub-models: biomass production, biomass allocation, soil water balance, stem mortality and stand characteristics. Several authors have described the processes, mathematical equations, and structural representation of the sub-models (Sands & Landsberg, 2002; Almeida, 2003; Sands, 2004; Landsberg & Sands, 2011). Therefore, only key points in the processes will be highlighted in this section. Figure 3.1 shows the 3PG framework.

The biomass production sub-model calculates light absorbed by the canopy using Beer's law. Gross primary production (GPP) is calculated based on a species-specific canopy quantum efficiency (α_c), which accounts for the effect of site and environmental factors on stand-level GPP through a series of growth modifiers (see Almeida (2003) and Landsberg & Sands (2011) for the detailed description of the growth modifiers). GPP is then converted to Net Primary Productivity (NPP) by multiplying GPP by a constant carbon use efficiency factor ($Y = 0.47 \pm 0.04$) (Waring et al., 1998), derived from an empirically determined ratio of NPP to GPP. The biomass allocation sub-model partitions NPP to the roots first followed by stems and then foliage, and partitioning is influenced by site fertility, water availability, vapor pressure deficit, and tree size. Water balance is performed using a simple single-layer soil water balance. Monthly precipitation is balanced against monthly evapotranspiration calculated using the Penman-Monteith equation (Monteith, 1965) such that if evapotranspiration is greater than precipitation, then water balance is negative.
Furthermore, suppose precipitation is greater than the maximum available soil water (ASW). In that case, water is assumed lost as runoff or drainage. The mortality sub-model calculates tree mortality as either density-independent (environmental or stress-induced) or density-dependent (using the -3/2 self-thinning law) (Drew & Flewelling, 1977). The stand characteristics sub-model converts biomass into output variables of interest to forest managers, such as basal area, mean tree diameter, height, and stand volume.

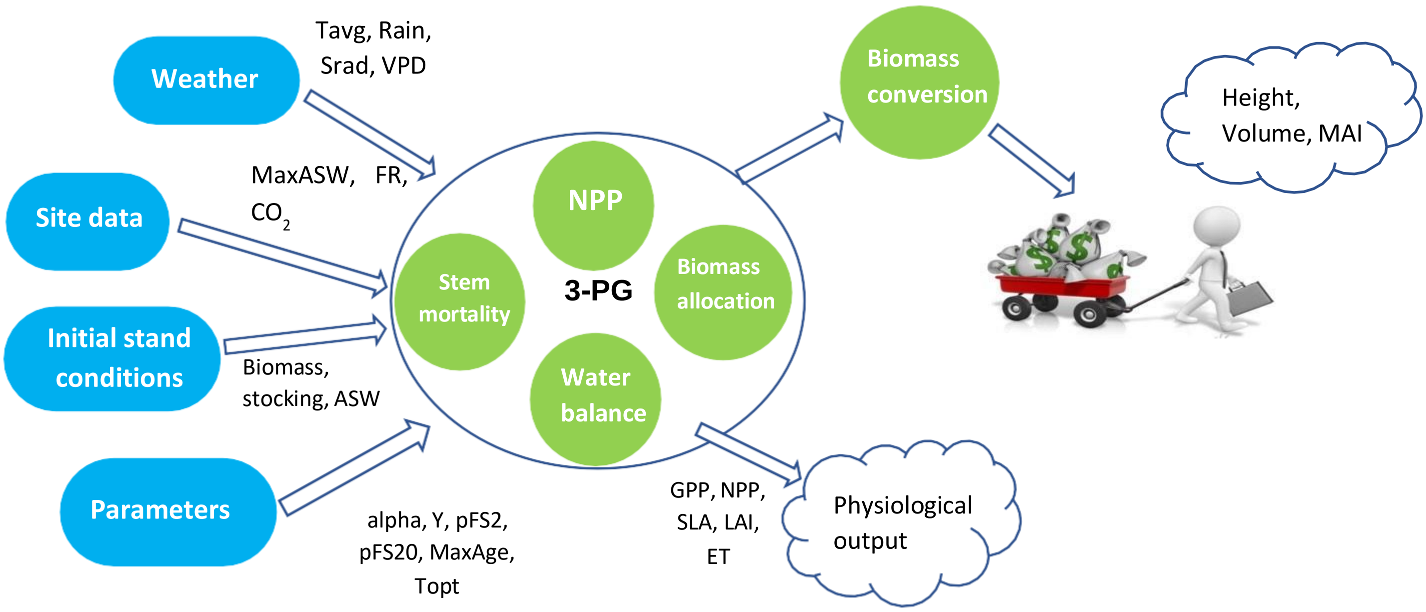


Figure 3.1 Schematic representation of 3PG inputs (blue), processes (green), and outputs

3.3 Materials and Method

3.3.1 Simulation software

Preliminary 3PG runs were done using 3PGpjs vsn. 2.7 (Sands, 2010), the Excel version of the model (available at [http://3pg.forestry.ubc.ca/software/)](http://3pg.forestry.ubc.ca/software/) to develop familiarity with the tool. However, the simulation runs and optimization were ultimately undertaken using a package (*R3PG)* developed by Trotsiuk et al. (2020) in *R* (R Core Team, 2018). The package offers users a flexible switch between various options and submodules to use the original 3PGpjs (Landsberg & Waring, 1997) and 3PGmix (Forrester & Tang, 2016). To run the original 3PGpjs, we used *settings = list(light_model = 1, transp_model = 1, phys_model = 1, height_model = 1, correct_bias = 0, calculate_d13c = 0)*. The function *run_3PG* was designed for *SingleSite* run type. Therefore, we developed a loop function to run *R3PG* for *MultiSite* run type. Figure 3.2 shows the flow chart for this loop function created using the *flow* R package (Antoine, 2021).

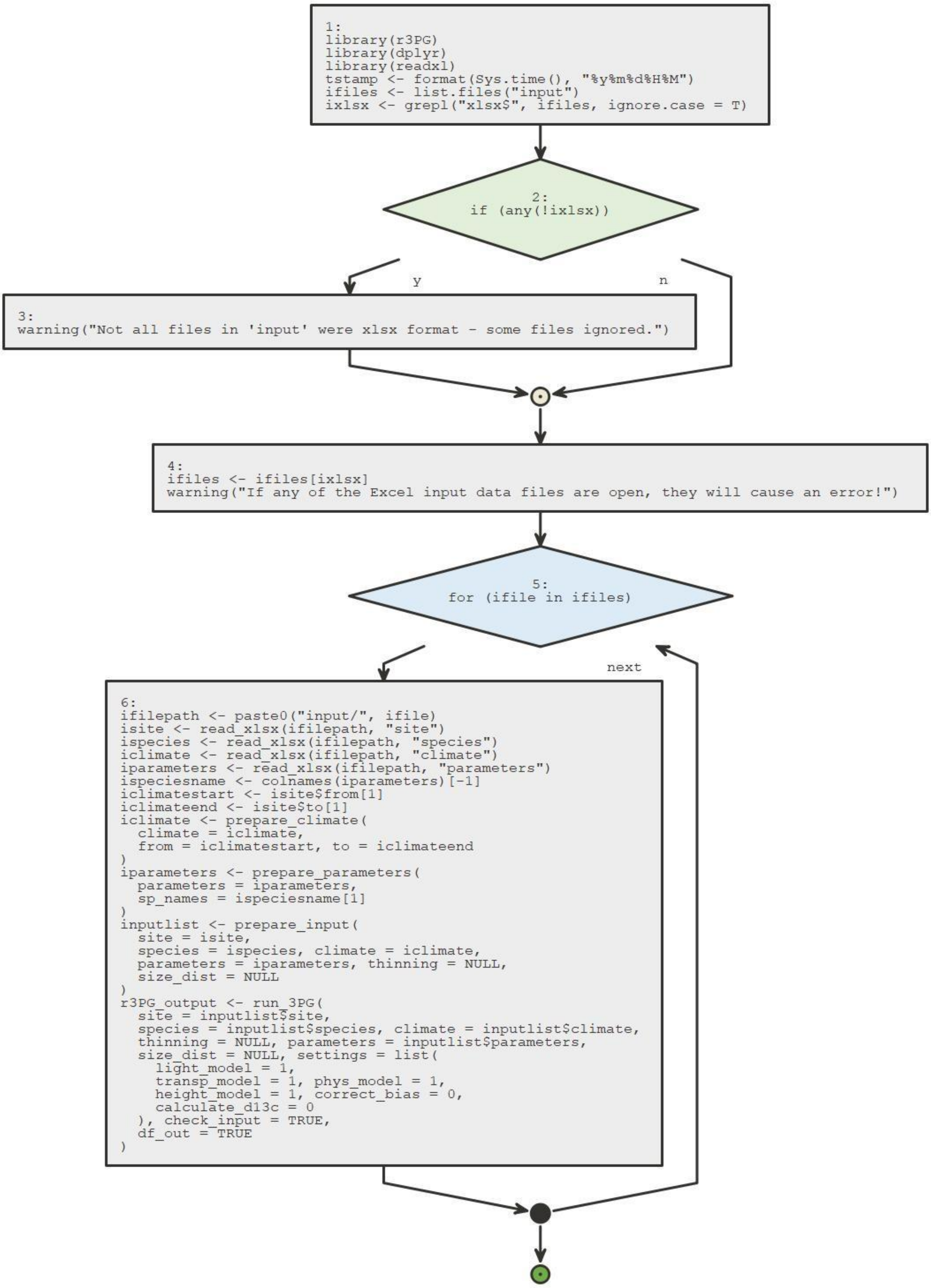


Figure 3.2 Representation of *R3PG* loop function to perform *MultiSite* run type

3.3.2 Site and stand information

Forest stands data were obtained from two companies that manage eucalypt resources for pulp in the Zululand area (see maps in section 1.1). All sites used the hybrid Egxu, but not all stands were of the same parent trees. Plots ranged from 0.043 to 0.066 ha but were all established as square or rectangular plots within the stands. Summary of the site and stand information is presented in Table 3.1.

3.3.3 Stand growth data

Tree-level diameter at breast height (DBH) and height data for the study plots were obtained from various sources. First, from routine PSP re-measurements undertaken by the two forest companies involved in the study (Table 3.2). This was based on annual measurements from the second year after establishment. Second, data for five plots were available in greater detail from a set of band dendrometers installed in December 2013 (see description in Table 3.3). Measurement of DBH and stem number were carried out biweekly from 2013 when the dendrometers were installed, while height measurements were done annually. For each plot, age was calculated based on planting date, and height measurements were carried out on a subset of trees. Dominant height was calculated as the average height of the 20 largest trees (based on DBH). Mean height was derived from the dominant height by multiplying it by a factor (the ratio of the mean height to dominant height from a complete measurement) (Equation 3.1). Third, in August 2018, Stellenbosch University students involved in the project took the final set of measurements. During this inventory, DBH and height of all trees were measured in the sample plots (see Table 3.3 and 3.4 for information on measurement dates for each stand).

$$Hmean = dom\ height * Factor$$

Equation 3. 1

Where Factor = 0.95 in this study, which corresponds to the value derived for *Eucalyptus grandis* (Tesfamichael et al., 2010)

Measurements of DBH, tree height, and stem numbers from the three sources were used to estimate quadratic mean diameter (Dq, cm) (Equation 3.2), Basal area (BA, m^2/ha) (Equation 3.3), and stand volume (V, m^3/ha) using a stand volume estimator by Burkhart & Tomé (2012) (Equation 3.4).

$$D_q = \sqrt{\frac{\sum DBH^2}{n}}$$

Equation 3. 2

$$BA = \frac{\pi * (D_q)^2 * TPH_i}{40000}$$

Equation 3. 3

$$V = BA \times H_{mean} \times f$$

Equation 3. 4

Where V is utilizable volume/ha (m^3/ha), *BA is* the basal area (m^2/ha), *Dq* is the quadratic mean diameter (cm), DBH is the stem diameter at breast height (cm), n is the number of observed trees per plot, TPH_i is the number of stems (t/ha), H_{mean} the mean stand height (m), and f the species-specific form factor (for *E. grandis x urophylla* by Kassier (2005)).

This stand volume equation was used throughout the study, including the derivation of volume from simulated BA and height.

Table 3.1 Summary of sites information used for model calibration

Company	Compartment name	Latitude	Longitude	Site index	Elevation (m)	Spacing (m x m)	Plot area (ha)	Soil form
Sappi	E6a	-28.39	32.22	16.7	53	3.0 x 2.2	0.0514	Fw
Sappi	B3a	-28.54	32.20	18.3	39	3.0 x 2.2	0.0508	Fw
Mondi Forest	B003	-28.42	32.21	17.1	45	2.5 x 3.0	0.0479	Fw1210
Mondi Forest	J006	-28.68	32.04	22.8	43	2.5 x 3.0	0.0433	Hu2200
Mondi Forest	B032	-28.21	32.33	16.1	29	2.5 x 3.0	0.0483	Cv21
Sappi	C15a	-28.34	32.24	16.8	59	3.0 x 2.2	0.0515	Fw
Sappi	F7	-28.56	32.21	21.2	54	3.0 x 2.2	0.0535	Fw
Sappi	G22b	-28.54	32.23	15.8	20	2.7 x 2.2	0.0594	Fw
Sappi	G33b	-28.54	32.23	19.2	20	3.0 x 2.2	0.0660	Fw
Mondi Forest	A017	-28.68	32.14	27.8	32	2.5 x 3.0	0.0491	Vf2110
Mondi Forest	B044	-28.97	31.64	25.9	55	2.5 x 3.0	0.0500	Hu2100
Mondi Forest	F011A	-28.62	32.18	28.8	63	2.5 x 3.0	0.0488	Ct2100
Sappi	B35b	-28.68	32.09	20.2	60	3.0 x 2.2	0.0535	Fw
Sappi	B38	-28.67	32.10	20.0	62	3.0 x 2.2	0.0535	Fw
Sappi	C55	-28.70	32.08	22.9	41	3.0 x 2.2	0.0535	Fw
Sappi	D13b	-28.51	32.12	15.1	63	3.0 x 2.2	0.0660	Fw
Sappi	E23f	-28.52	32.15	14.0	44	3.0 x 2.2	0.0660	Fw
Sappi	E24g	-28.53	32.15	14.4	44	2.7 x 2.4	0.0648	Fw

Table 3.1 Stand information for the KwaMbonambi compartment

Company	Compartment name	Planted date	Measurement date	Age (yrs.)	Clone	TPH (t/ha)	Initial trees per plot
Sappi	E6a	2010/08/15	2018/08/22	8.02	GU W1830	1515	81
Sappi	B3a	2010/04/16	2018/08/21	8.35	GU W1700	1515	81
Mondi Forest	B003	2011/07/04	2018/08/22	7.13	GGRAURO	1336	64
Mondi Forest	J006	2010/07/01	2018/08/21	8.14	GGRAURO	1479	64
Mondi Forest	B032	2012/08/01	2018/08/22	6.06	GGRAURO	1347	64
Sappi	C15a	2010/06/16	2018/08/17	8.17	GU W1700	1515	81
Sappi	F7	2010/11/16	2018/08/14	7.74	GU W1830	1515	81
Mondi Forest	A017	2011/07/01	2018/08/11	7.11	GGRAURO	1366	64
Mondi Forest	B044	2011/06/01	2018/08/21	7.22	GGRAURO	1280	64
Mondi Forest	F011A	2012/08/02	2018/08/11	6.03	GGRAURO	1312	64
Sappi	B35b	2012/04/16	2018/08/17	6.34	GU SGU1932	1515	81
Sappi	B38	2011/03/16	2018/08/17	7.42	GU W1830	1515	81
Sappi	C55	2010/06/16	2018/08/20	8.18	GU SGU1932	1515	81

Table 3.2 Stand information for the Marie Curie sites

Company	Compartment name	Planted date	Measurement date	Age (yrs)	Clone	TPH (t/ha)	Initial trees per plot	MAP range (mm)
Sappi	G22b	2008/04/15	2018/08/10	10.32	GU W1022	1684	100	<1000
Sappi	G33b	2011/03/16	2018/08/10	7.40	GU W1022	1515	100	1000-1100
Sappi	D13b	2011/05/16	2018/08/07	7.23	GU W1830	1515	100	1000-1100
Sappi	E23f	2012/06/16	2018/08/07	6.14	GU W1700	1515	100	>1100
Sappi	E24g	2008/03/15	2018/08/07	10.39	MIXED	1543	100	>1100

3.3.4 Soil data

Soil information, including physical and chemical properties, was made available from previous experimental sampling conducted on the study site during fieldwork in 2018. The center of each PSP was determined, and pits were cored, as shown in Figure 3.3. For soil textural analysis, soil samples were cored at the middle pit (grey pit in Figure 3.3) using a 1.2 m manual steel auger at 10 cm intervals until a soil depth of 1.2 m was reached. For soil chemical analysis, soil samples were cored at three different points (shown in orange in Figure 3.3), and samples were collected at three depths (0-10 cm, 10-20 cm, and 20-50 cm soil depth) from each point, using an auger. Undisturbed soil samples used to determine bulk density were collected 1 m away from the first soil pit (the yellow pit in Figure 3.3). Samples were taken at 0-10 cm, 10-20 cm, and 20-50 cm soil depths by pressing a steel ring with a diameter of 7.5 cm and a height of 6.5 cm into undisturbed soil. Soil textural and chemical analysis was performed at the Institute for Commercial Forest Research (ICFR) (https://www.icfr.ukzn.ac.za). Soil textural information and chemical properties from the experiment are presented in Table 3.4 and 3.5. The soil class was then determined from the South African texture triangle chart (Schulze, 2007). There was very little heterogeneity in soil class, as expected in this region.

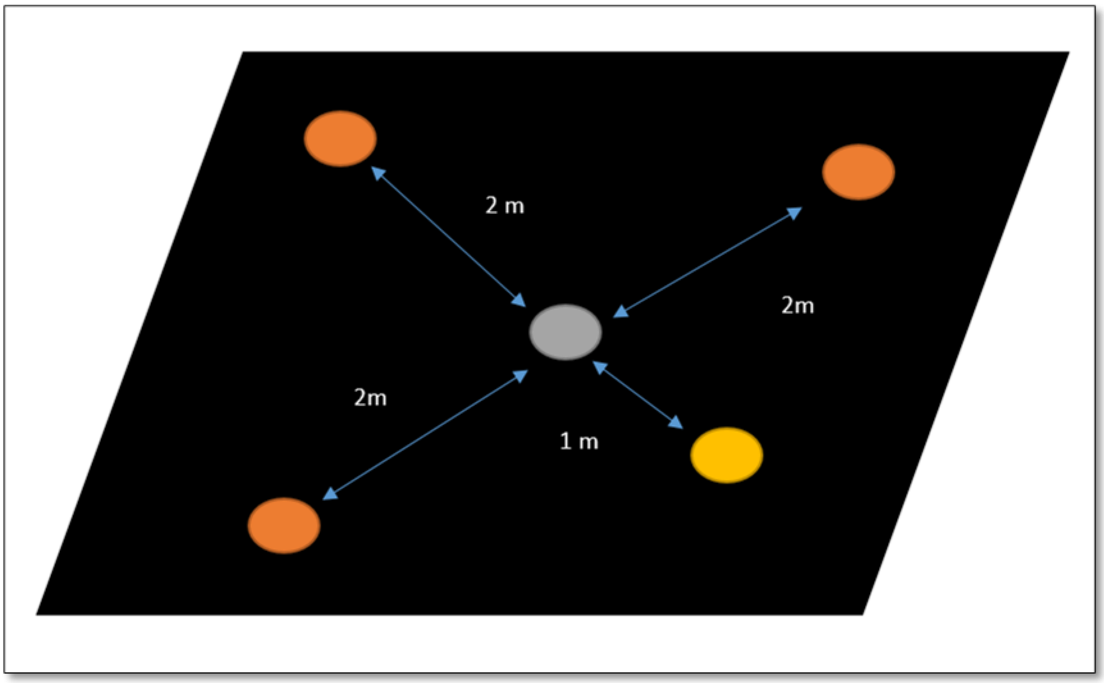


Figure 3.3 Diagram showing the soil pits for soil sampling. The Grey circle represent the soil pit for textural analysis, the tree orange circles represent soil pits for chemical analysis, and the yellow circle represent soil pit for bulk density

Table 3.3 Average soil textural properties averaged across the soil depths (0 – 50 cm)

Compartment name	Silt (%)	Clay (%)	Sand (%)	Soil class
Futululu E6a	0.03	0.04	0.93	Sandy
Mavuya B3a	0.06	0.05	0.89	Sandy
Mtubatuba B003	0.06	0.08	0.85	Sandy
Nseleni J006	0.04	0.08	0.87	Sandy
Nyalazi B032	0.10	0.12	0.78	Loamy Sand
PalmRidge C15a	0.04	0.04	0.92	Sandy
Salpine F7	0.04	0.06	0.91	Sandy
Salpine G22b	0.04	0.04	0.92	Sandy
Salpine G33b	0.09	0.06	0.85	Sandy
Siyaqhubeka A017	0.04	0.07	0.89	Sandy
Siyaqhubeka B044	0.04	0.06	0.89	Sandy
Siyaqhubeka F011A	0.03	0.06	0.92	Sandy
SouthAreas B35b	0.02	0.03	0.95	Sandy
Terranera B38	0.04	0.04	0.92	Sandy
Terranera C55	0.03	0.07	0.91	Sandy
Trust D13b	0.02	0.04	0.94	Sandy
Trust E23f	0.03	0.04	0.93	Sandy
Trust E24g	0.05	0.10	0.84	Sandy

Table 3.4 Chemical properties of the soil (0 – 50 cm) for the 18 sites

Site	SOC (%)	N (%)	S (%)	P (%)	pH (%)	C:N (%)
Futululu E6a	1.82	0.10	0.02	10.47	6.35	54.59
Mavuya B3a	3.07	0.13	0.03	3.26	4.91	71.25
Mtubatuba B003	1.19	0.08	0.02	15.40	5.85	44.67
Nseleni J006	1.39	0.09	0.02	33.87	6.69	43.60
Nyalazi B032	0.79	0.05	0.02	10.61	5.72	44.86
PalmRidge C15a	1.41	0.07	0.02	3.79	4.86	61.52
Salpine F7	0.72	0.04	0.01	2.88	5.38	56.94
Salpine G22b	0.58	0.03	0.01	3.72	4.86	63.93
Salpine G33b	1.01	0.07	0.02	3.08	5.31	42.04
Siyaqhubeka A017	1.54	0.09	0.03	4.20	5.34	52.36
Siyaqhubeka B044	1.22	0.10	0.03	4.48	4.94	33.67
Siyaqhubeka F011A	1.27	0.07	0.02	4.17	4.67	53.11
SouthAreas B35b	0.73	0.05	0.02	3.99	5.46	46.56
Terranera B38	0.89	0.05	0.02	4.97	5.67	54.41
Terranera C55	1.28	0.09	0.03	4.98	5.16	41.98
Trust D13b	1.08	0.07	0.02	21.42	5.85	39.00
Trust E23f	1.12	0.07	0.02	6.23	6.25	47.18
Trust E24g	1.16	0.07	0.02	24.62	5.81	48.87

SOC – Soil organic carbon, N – Nitrogen, S – Sulphur, P – Phosphorus, C:N- carbon to Nitrogen ratio

3.3.4.1 Available Soil Water (ASW)

Available soil water was estimated from the soil texture and soil organic matter described in section 3.3.4, using the soil water characteristics equation by Saxton & Rawls (2006). Maximum available soil water was calculated as the product of soil depth and derived available water capacity (Table 3.6). In 3PG, the minimum available soil water is usually set as a default value of zero. However, it is used to account for access to water table (Sands, 2004). Additionally, according to a comment by David Forrester (pers. Comm) (*Soil Water Availability · Issue #51 · Trotsiuk/R3PG*, n.d.), *"…Occasionally, it can also be reasonable to set MinASW > 0, e.g., when the user knows that the plants have access to a deep permanent water source, or when there is irrigation. In these situations, it needs to be calculated based on information about how much water is being applied (irrigation) or how much the plants can potentially access from a deep-water table after precipitation has been used".* For this reason, the sites were visually inspected using *Google Earth Pro. S*ites planted adjacent to the indigenous forest conservation area, which almost invariably grow along perennial watercourses, were set to half the initial ASW. This is because, it seem very likely that these plots had higher-than-normal access to ground water. This might not be a pragmatic approach in a much broader application of 3PG, but for the calibration work in this case study it was deemed appropriate to avoid any major bias due to major omission of key information.

3.3.4.2 Fertility Rating (FR)

Soil fertility varies over time and space, and determining it requires complex and time-consuming procedures (Binkley & Fisher, 2013; McGrath et al., 2014). According to Landsberg et al. (2001), *"many years of research effort to describe site fertility in terms usable in quantitative models of plant growth has been extremely limited"*. Therefore, the 3PG model relies on a fertility rating index (FR) to relate soil fertility to stand productivity. The fertility rating is an important species-specific variable in 3PG. It affects canopy quantum efficiency and biomass allocation to the root (Landsberg et al., 2001). It ranks soil fertility from 0 (extreme nutritional limitation) to 1 (No nutritional limitation). The empirical nature of the fertility rating has been criticized; however, assigning fertility rating to a site is still problematic (Landsberg & Sands, 2011).

Several studies have presented different approaches to estimate this variable. Some authors depend on arbitrary fixed FR values (Almeida et al., 2004; Campion et al., 2005; Esprey, 2006; Coops et al., 2010), or values derived from soil properties (Xenakis et al., 2008; Vega-Nieva et al., 2013), fertilization trials (Stape et al., 2004), or direct/indirect correlation with site index (Dye et al., 2004; Gonzalez-Benecke et al., 2014; Subedi et al., 2015).

(52)

In this study, to explore the likely variability in FR, FR was varied iteratively to obtain the optimized values for each site. Stepwise regression was performed using the optimized FR values as the independent variable. Soil physical and chemical properties described in section 3.3.4, the total amount of rainfall received, ASW, and site index values, as the explanatory variable. None of the explanatory variables appeared to correlate clearly with FR (data not shown). Consequently, and given that the region is characterized by relatively homogenous soils, FR was set to a constant value of 0.5. Under these circumstances, this is a practical solution, but it may not be the case in more heterogeneous soils.

3.3.5 Weather data

Weather data were obtained using the interpolation methods described in Chapter 2. VPD was not input but rather calculated using the formula embedded in the 3PG model from minimum and maximum temperature. Figure 3.4 shows the mean annual total rainfall variation compared to the long-term (1959-1999) mean rainfall. Note the very dry years 2014 and 2015, which were the region’s driest on record.

Table 3.5 Total soil depth and available soil water for the 18 sites

Compartment name	Available soil water (mm)	Soil depth (m)	Maximum available water (mm)
Futululu E6a	39.1	1.2	47
Mavuya B3a	48	1.2	58
Mtubatuba B003	49.2	1.5	74
Nseleni J006	46.4	1.5	70
Nyalazi B032	60.7	1.2	73
PalmRidge C15a	40.5	1.2	49
Salpine F7	40.5	1.2	49
Salpine G22b	39.0	1.2	47
Salpine G33b	51.4	1.2	62
Siyaqhubeka A017	44.2	1.5	67
Siyaqhubeka B044	43.1	1.5	65
Siyaqhubeka F011A	39.3	1.5	59
SouthAreas B35b	34.7	1.2	42
Terranera B38	39.9	1.2	48
Terranera C55	40.4	1.2	49
Trust D13b	36.7	1.2	44
Trust E23f	38.6	1.2	46
Trust E24g	50.4	1.2	61

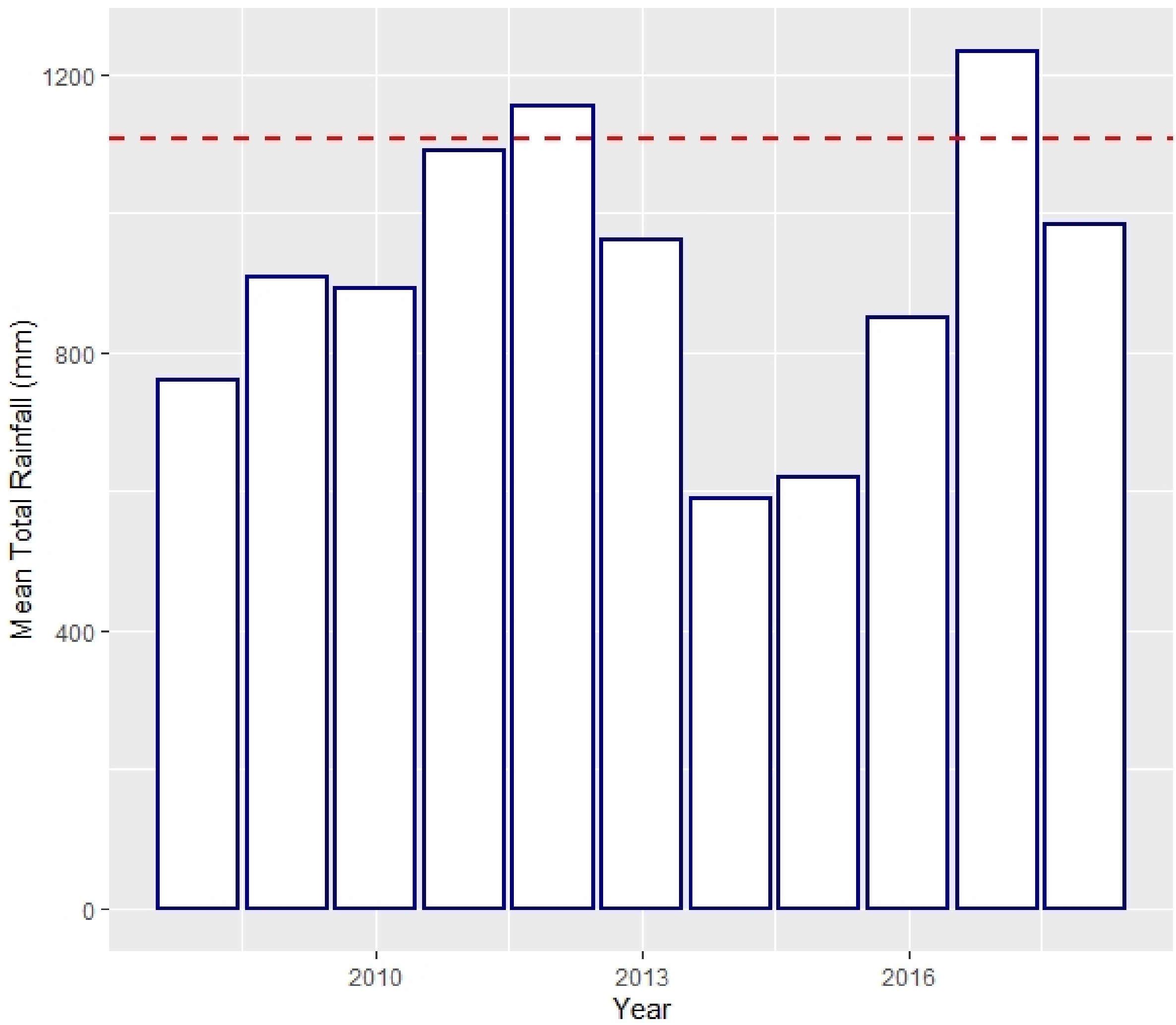


Figure 3.4 Mean annual total rainfall for all 18 study sites from 2008 – 2018. The red dotted line indicates the long-term mean rainfall, as estimated by Schulze (2007), and serves as a point of reference.

3.4 Calibration of 3PG model

In general, the data needed to parameterize and test 3PG can be divided into biomass harvest, field data, literature, mensuration, and physiological data (Esprey, 2006). Direct measurement, analogy with other species, and parameter estimation are three methods for assigning values to parameters in a model (Landsberg & Sands, 2011). Calibration should be done through direct measurement of parameter values whenever possible, which can be done from some experimental measurements or indirectly through fitting a simple model or calculation using another model (Landsberg & Sands, 2011). However, many parameters are essentially generic and can be assigned values based on analogy with other species, such as the conversion of solar radiation to PAR (molPAR_MJ = 2.3 mol/MJ) (Landsberg & Sands, 2011). Failing both options, parameter values can be adjusted through parameter estimation.

Following the parameterization guidelines presented by several authors (Sands, 2004; Esprey, 2006; Landsberg & Sands, 2011), where parameters could not be calibrated because of lack of suitable data, or parameterized due to low sensitivity ratings, default parameter values were chosen from Sands & Landsberg (2002) and Borges et al. (2012).

3.4.1 Allometric parameters for stem mass as a function of DBH

Biomass harvest data used were measured from a subset of sites in destructive samples taken in 2018. The five sites utilized for obtaining these data were the same as those with installed dendrometer bands described in section 3.4.2. The sites were established to investigate short-term growth variability in young and old stands at various levels of drought. These compartments were designed to represent three rainfall classes based on their mean annual precipitation, namely wet (>1100mm), moderate (1000 – 1100 mm), and dry (<1000 mm). Trees received 60 kg/ha LAN fertilizer at planting, with no weeding, pruning, or thinning performed throughout the growth period.

Three trees representing the first quartile (Q1), third quartile (Q2), and maximum in the diameter distribution were destructively harvested in each compartment. Measurements collected were total height, diameter at breast height (DBH), aboveground biomass (stemwood, branch, and foliage). Parameters for the allometric relationship between tree-level biomass (w_s, kg tree^{-1}) and DBH were then estimated for Equation 3.5 as specified by Sands & Landsberg (2002).

$$w_s = a_s B^{n_s}$$

Equation 3. 5

where *B* is stem diameter at breast height, a_S is the coefficient, and n_s is the power in the allometric relationship

The allometric parameter derived using the fifteen harvested trees from the five sites was used to calculate the individual tree mass as a function of DBH for each of the trees measured at the 18 sites during the fieldwork in 2018. The average w_s and Dq (quadratic mean diameter) of the trees at each of the 18 sites were calculated. These 18 pairs of w_s and Dq were combined to develop a single stand-based allometric relationship representing all sites. This was done to upscale the parameter values to stand level to be consistent with the 3PG calculation (Esprey, 2006).

3.4.2 Density-independent mortality coefficients

Some of the sites experienced mortality at post-planting. Therefore we fitted the parameter values for density-independent mortality. One of the sites (B044) which experienced high mortality was used. The Clutter and Jones mortality function (Clutter & Jones, 1980) was used to estimate tree survival per year, then data modelled was fitted using a gaussian function with a non-zero asymptote (Sands and Landsberg, 2002) (Equation 3.6).

$$\boldsymbol{\gamma(t) = \gamma Nx + (\gamma Nx - \gamma N0)e^{-(ln2)t/t\gamma N}}$$

Equation 3. 6

Where, γNx = 0.60, $\gamma N0$ = 1.01, $t\gamma N$ = 3.39

3.4.3 Parameter estimation for Zululand Egxu

Eleven parameters (test parameters) (Table 3.7) were selected from the list of 3PG standard parameters, and parameter values published by Borges et al. (2012) were used as the base parameters. These test parameters were selected because they could not be calibrated from the data available in this study, and 3PG outputs have shown sensitivity to them (Almeida et al., 2004; Esprey et al., 2004; Forrester & Tang, 2016). Published parameter values for Egxu by Almeida et al. (2004) and Borges et al. (2012) were set as biologically plausible bounds (to give three test values: low, medium, high) in the estimation process. An algorithm was developed as part of an *R3PG_Parameter_Testing* pipeline using *R* (R Core Team, 2021) to generate all the possible combinations of the test parameter values (number of combinations = number of test values ^ number of test parameters). The different combinations of test parameter values were then combined with the rest of the base parameters. The routine produces an output as a **.csv* file with each row representing a full 3PG parameter set (pset), with the column names corresponding to parameter names. Another script in this pipeline was created to run the different combinations of 3PG parameters (psets) for multiple sites (i.e., each site had R3PG run *n* times, where *n* is the number of psets, and the process was repeated for each site using a loop construction). The R3PG calibration simulations used 17 sites out of the 18 sites presented in Table 3.2. There was tree theft in one of the sites (Salpine G22) at an early age, and inventory data provided were from the adjacent compartment. As a result, this site was dropped in the parameter estimation process. Site information and weather data required for 3PG simulations were the same as described in sections 3.3.2 to 3.3.5.

According to Sands (2004), parameter estimation should be based on observed values of all state variables (W_F, W_S, W_R, N, and θ_S). However, surrogates for stem biomass such as DBH, stem height, or volume can be used. Dq, mean stem height, and basal area were used in this study. Basal area was selected as it is also a function of stocking. For foliage biomass, Leaf area index (LAI) is a surrogate. Although we did not have access to observed ground-based time-series LAI data for this study, we performed a qualitative comparison by comparing the 3PG LAI values with the Landsat 8 Collection 1 Tier 1 Normalized Difference Vegetation Index (NDVI) product. We used the 8-Day NDVI composite dataset retrieved from Google Earth Engine (GEE) environment. Also, in the *R3PG_Parameter_Testing* pipeline, output other than Dq, mean stem height, and basal area were discarded. The complete R scripts and the template file for this algorithm are available on GitHub at https://github.com/EucXylo/R3PG_parameter_testing.

3.4.4 Selecting the optimized parameter set

All candidate psets generated in section 3.5.2 were evaluated by matching their predicted stem diameter, height, and basal area values to corresponding observed data (for all observed values where stand age > 3). The following statistics were considered to select the best performing pset: R-squared, sum of squared error (SSE), intercept, and slope. In addition, another measure was developed, here named the *extended Root Mean Square Error (eRMSE)*, for each pset across all sites to generate a single metric for ranking psets. The eRMSE is derived from the sum of square errors of predictions against observed values (SSE) and the sum of square errors of predictions from the line of best fit against observed values (SSF). The SSE is the squared difference between the observed values and the R3PG predictions (which corresponds to residuals from the identity line (slope = 1, intercept = 0) (Equation 3.7). The SSF is the squared difference between the observed values and the line of best fit of observed values vs. predictions (Equation 3.8). The values from the line of best fit are calculated by inserting the observed values into the (observed-vs-R3PG prediction) regression equation (Equation 3.9).

$$SSE = \sum_{i=1}^{n}\left(X_{obs,i} - y_{R3PG,i}\right)^2$$

Equation 3. 7

$$SSF = \sum_{i=1}^{n}\left(X_{obs,i} - y'_{fit,i}\right)^2$$

Equation 3. 8

$$y'_{fit} = aX_{obs} + b$$

Equation 3. 9

Where, *a* is the slope from the regression equation, *b* is the intercept from the regression equation, X_{obs} is the observed value (including all the variables included in the evaluation – Dq, basal area, and mean stem height), y_{R3PG} is the corresponding 3PG-predicted value, y'_{fit} is the corresponding fitted value on the line of best fit between observed and predicted values.

Then eRMSE was calculated as

$$eRMSE = \sqrt{\frac{SSE + SSF}{2n}}$$

Equation 3. 10

Where, SSE is the sum of square error, SSF is the sum of square fit, and n is the number of observed values.

The sum of square fit in the evaluation metric helps to avoid choosing psets that might have a low sum of square error while at the same time having a bias (with systematic over/under prediction at low vs. high values). Overall, the eRMSE aims to aggregate the prediction and fit errors into a single measure of predictive power, thereby enabling the selection of psets with minimized residuals, low bias, and line of best fit close to the identity line (slope = 1, intercept = 0). The graphical representation of this concept is shown in Figure 3.5. The complete R scripts and template file for this algorithm are available on GitHub at https://github.com/EucXylo/R3PG_pset_refining

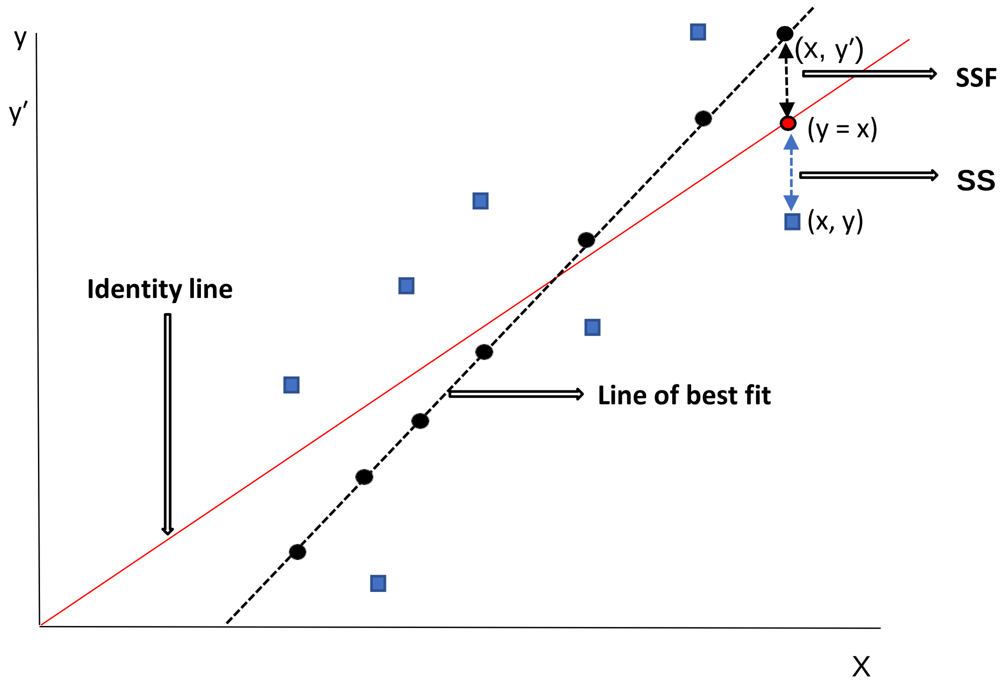


Figure 3.5 A hypothetical graph explaining the eRMSE concept, where SSE is the sum of square error (R3PG-predicted vs. observed values), SSF is the sum of square error fit (predictions from the line of best fit vs. observed values), the solid red line is the identity line (predicted = observed), and the black dashed line is the line of best fit for R3PG-predictions vs. observed values. The blue squares show the original model-predicted and observed value data, black circles show the regression fit values, and the red circle represents a perfect model.

3.4.5 Validation of the 3PG model

The predictive accuracy of the 3PG model was further tested by validating the model against data from 15 independent sites in the same region managed by the two forest companies. Stand growth data at a specific age were made available. Summary of the site and stand information used is presented in (Table 3.8). Weather data were obtained using the interpolation technique described in Chapter 2. Plant available soil water was estimated from the South African soil classification map (Soil Classification Working Group, 1991). However, ASW obtained from this map were overestimated for sandy soil (99 – 105mm) compared to the typical value (±80mm) for the region's soil form (Fernwood) specified by (Olivier, 2017) and those derived from soil texture (34.7 – 60.7mm) used in model calibration (section 3.4.3.1).

For this reason, the initial ASW was set as the mean ASW of the calibration sites. A constant FR value of 0.5 was also used during validation. Parameter values obtained from section 3.4.3 and three Brazilian parameter sets by Borges et al. (2012) and Almeida et al. (2004) were used to run the 3PG model. Similar to the calibration sites, there was mortality at post-planting, which became stable afterward. Therefore, we set the initial stem number for these sites at the stable value. Using these parameter sets, four sets of model predictions (stand basal area and height) were used to estimate stand volume using Equation 3.2, and this was compared with observed stand volume. The following statistical error and indices from the Agricultural and Meteorological software (AgriMetSoft, 2019) were used to evaluate the performance of 3PG model: root mean square error (RMSE), coefficient of determination (R^2), and Nash Sutcliffe efficiency index (E) as described in section 2.4.

Table 3.6 Test parameters and their values considered as bound during parameter estimation

Parameter	**Egxu** (Borges et al., 2012)	**Clone 15** (Almeida et al., 2004)	**Clone 22** (Almeida et al., 2004)	**E.grandis** (Esprey, 2006)
pFS2	1.64	0.7	0.7	-
pFS20	0.15	0.1	0.11	-
pRx	0.5	0.6	0.6	-
pRn	0.1	0.07	0.12	-
gammaF1	0.07	0.13	0.13	-
Tmin	8	8	8	3
Topt	25	25	25	23
Tmax	40	36	36	25
alphaCx	0.08	0.068	0.068	-
wSx1000	300	180	180	-
CoeffCond	0.0324	0.045	0.05	-

Table 3.7 Summary of the site and stand information of the sites used for validation

Company	Compartment name	Longitude	Latitude	Plant Date	Enumeration Date	Altitude	Age	Spacing	SI	TPH0	TPH1	Soil depth
Sappi	FutululuE4b	32.20881	-28.38448	2010/07/19	2017/07/19	58.2	7.70	3.0 x 2.2	13.7	1515	1449	1.2
Sappi	FutululuE4c	32.2138	-28.38134	2010/07/19	2017/07/19	54.5	7.70	3.0 x 2.2	12.9	1515	1480	1.2
Sappi	FutululuE4i	32.20957	-28.38366	2010/07/19	2017/07/19	56.2	7.70	3.0 x 2.2	12.7	1515	1512	1.2
Mondi Forest	Kwambonambi_F011A	32.07278	-28.58583	2010/05/10	2018/04/12	84.1	7.92	2.5 x 3.0	21.0	1390	1368	1.5
Mondi Forest	Nseleni_C002	31.97126	-28.72292	2013/05/09	2018/04/18	55.1	4.92	2.5 x 3.0	17.8	1450	1428	1.5
Mondi Forest	Nyalazi_D006	32.26713	-28.27855	2011/04/21	2018/03/10	40.0	6.92	2.5 x 3.0	19.6	1330	1206	1.5
Sappi	Palm RidgeA18	32.26028	-28.30064	2008/08/15	2017/09/25	37.8	9.70	2.7 x 2.4	14.8	1515	1441	1.2
Sappi	Palm RidgeA21a	32.26401	-28.30437	2010/07/21	2017/07/25	39.1	7.70	3.0 x 2.2	14.4	1515	1401	1.2
Sappi	Palm RidgeA21b	32.26339	-28.30393	2010/07/21	2017/07/25	41.2	7.70	3.0 x 2.2	13.4	1515	1361	1.2
Sappi	Palm RidgeA21i	32.26377	-28.30591	2010/07/21	2017/07/25	41.0	7.70	3.0 x 2.2	13.7	1515	1393	1.2
Mondi Forest	Siyaqhubeka_B060	32.28511	-28.32640	2012/04/10	2018/03/15	60.3	5.92	2.5 x 3.0	19.7	1833	1833	1.5
Mondi Forest	Siyaqhubeka_C001	32.30249	-28.30861	2012/05/09	2018/03/15	57.7	5.83	2.5 x 3.0	16.5	1700	1593	1.5
Sappi	TrustL11b	32.0878	-28.55849	2010/07/20	2017/07/21	89.1	7.70	3.0 x 2.2	15.0	1515	1369	1.2
Sappi	TrustL11e	32.08544	-28.55538	2010/07/20	2017/07/21	92.7	7.70	3.0 x 2.2	12.6	1515	1348	1.2
Sappi	TrustL11j	32.08386	-28.55765	2010/07/20	2017/07/21	93.5	7.70	3.0 x 2.2	13.1	1515	1433	1.2

3.5 Results and Discussion

3.5.1 Allometric parameters a_s and n_s

The relationship between w_s and D_q measured at the "Marie Curie" sites used in fitting the allometric parameters (a_S and n_S) is presented in Figure 3.6. The allometric parameters, where, $a_S = 0.099$ and $n_S = 2.51$, gave an excellent fit ($R^2 = 0.99$; $p < 0.001$). The standard errors for this parameter calibration are $a_S = 0.477$, $n_S = 0.005$. Notably, the value of a_S (0.099) is higher than that obtained by Almeida et al. (2004) and Borges et al. (2012). However, n_S falls within the range of values obtained by both authors (see Table 3.9). These parameters determine the prediction of stem diameter and basal area by the 3PG model.

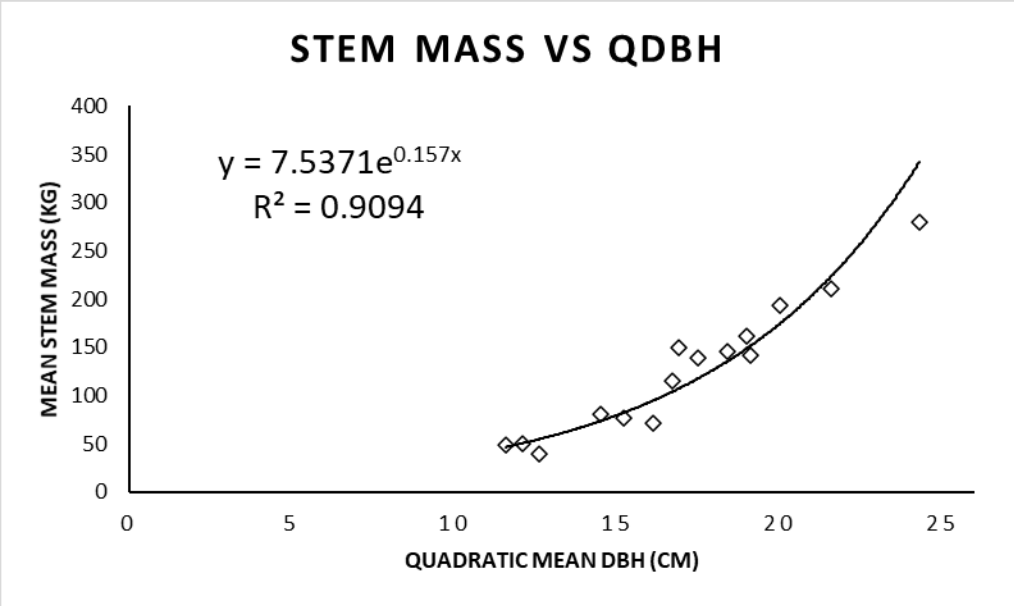


Figure 3.6 Allometric relationship between mean single-tree stem biomass (w_S) and Dq obtained from five "Marie Curie" plots. The line (———) is the fitted relationship used in 3PG ($a_S = 0.099$, $n_S = 2.51$).

3.5.2 Parameter Estimation

For this study, the parameter set with the lowest *eRMSE* was selected as the optimized parameter values for *Eucalyptus grandis x urophylla* in the Zululand region of South Africa. Using this parameter set resulted in the best predictions of the three variables (Mean height, basal area, and stem diameter at breast height) and stand volume calculated from basal area and height (Figure 3.7). For the 17 calibration sites, the 3PG predictions accounted for more than 80% of the variance in the observed values for all output variables considered (Table 3.9). For all output variables, the linear regression slopes between observed and predicted values were significantly different from zero ($p < 0.001$ in all cases). All the output variables used in the parameter estimation process resulted in a negative bias (Table 3.9). As illustrated in Figures 3.8, 3.9, and 3.10, 3PG underestimated growth in the early stages. The possible cause of this systematic error is discussed in the next section. The Nash Sutcliffe model efficiency index (E), which indicates how well the line of best fit from observed vs. predicted data fits the identity line, shows that 3PG prediction produced a very good match to the observed data (E > 0.7, where E = 1 indicates a perfect match). Overall, the good agreement between the observed and predicted output variables indicates adequate calibration of the 3PG parameters to predict forest growth in the study area. The list of parameter values from this study, by Almeida et al. (2004) and Borges et al. (2012) are presented in Table 3.10.

Table 3.8 Statistics describing the relationship between observed and predicted variables. Statistics include mean bias error (MBE), Nash Sutcliffe efficiency index (E), intercept, slope, p-value, and R^2

Output	MBE	E	Intercept	Slope	p-value	R^2	N
Basal Area	-1.65	0.78	-3.00	1.07	<0.001	0.87	105
Dq	-0.80	0.73	-1.625	1.06	<0.001	0.85	105
Mean Height	-0.16	0.78	-0.934	1.05	<0.001	0.84	105
Volume	-5.83	0.85	-18.624	1.10	<0.001	0.90	105

3.5.3 Model behavior at calibration sites

During the drought period, there was a decrease in Dq and mean height growth rate at most sites, but not at A017, B044, F011A, B35b, and C55 (Figures 3.8 and 3.9). All of these sites showed continuous growth during the dry period (Figure 3.8 and 3.9). Xulu et al. (2018) reported on the major effect of this drought in the region. However, it is an important insight that not all sites were affected to the same extent. Another factor to consider is genotype. Crous et al. (2018) highlighted significant intra-hybrid variation in the hydraulic traits of the *Eucalyptus* clones planted in this region. Nevertheless, in this region, broad-scale differences can be seen. Visually inspecting the sites for the drought year from satellite Imagery shows distinct differences between northern and southern plantations (See figure 3.10).

Another interesting finding from this visual analysis is that the five sites mentioned above, all southern sites, were established adjacent to indigenous forest conservation zones, which almost invariably grow along perennial watercourses. Accordingly, it would seem very likely that these managed blocks had higher-than-normal access to groundwater. However, clones planted at B032, which is also close to a natural forest located further North, appeared to have reduced growth during the drought and may have experienced die-back (See figure 3.10). This observed pattern is consistent with previous research, which predicted that tree response to this environmental fluctuation would be site and species-specific (Warburton & Schulze, 2008; Almeida et al., 2009; Pinkard et al., 2010; Naumberg et al., 2001; Booth, 2013).

In 3PG, radiation interception is mainly driven by the leaf area index (LAI), which determines the amount of dry mass production (Landsberg & Waring, 1997). We observed realistic LAI at some sites, particularly sites at the drier region, which are consistent with LAI values (Peak LAI at between 2.5 and 3 years and to vary between 4.0 and 5.3 m^2/m^2) shown by Campion et al. (2005). However, the higher peak LAI values at four (A017, B044, C55, B35) of the five sites previously mentioned in the KwaMbonambi region are the most striking (Figure 3.11). Higher LAI values were also discovered by Dye et al. (2004) for *Eucalyptus* plantations in the KwaMbonambi region. A qualitative comparison of the predicted LAI with the Landsat 8 NDVI values showed that there was a general decline in the NDVI values during the drought period, and the 3PG predicted LAI also showed the same pattern. Generally, the decline rate and recovery rate differ per region (Figure 3.11). The fact that the 3PG model showed response to the drought events demonstrates its usefulness in simulating growth patterns in response to the environment.

As illustrated in Figure 3.8 and 3.9, 3PG under-predicted early growth (Dq and mean stem height) from age zero to about five years at some, but not all, sites. According to Landsberg & Waring (1997), some systematic errors are expected due to the limitation of using Beer's law to calculate absorbed photosynthetically active radiation. The model assumes a closed canopy which is not necessarily true for young trees, coupled with the fact that some of the sites experienced mortality at post-planting. This explains the biasness observed in Table 3.9. However, as the stand age, 3PG prediction tends to match with observed values (Table 3.8 and 3.9). This pattern was also reported by Esprey (2006) and Miehle et al. (2009).

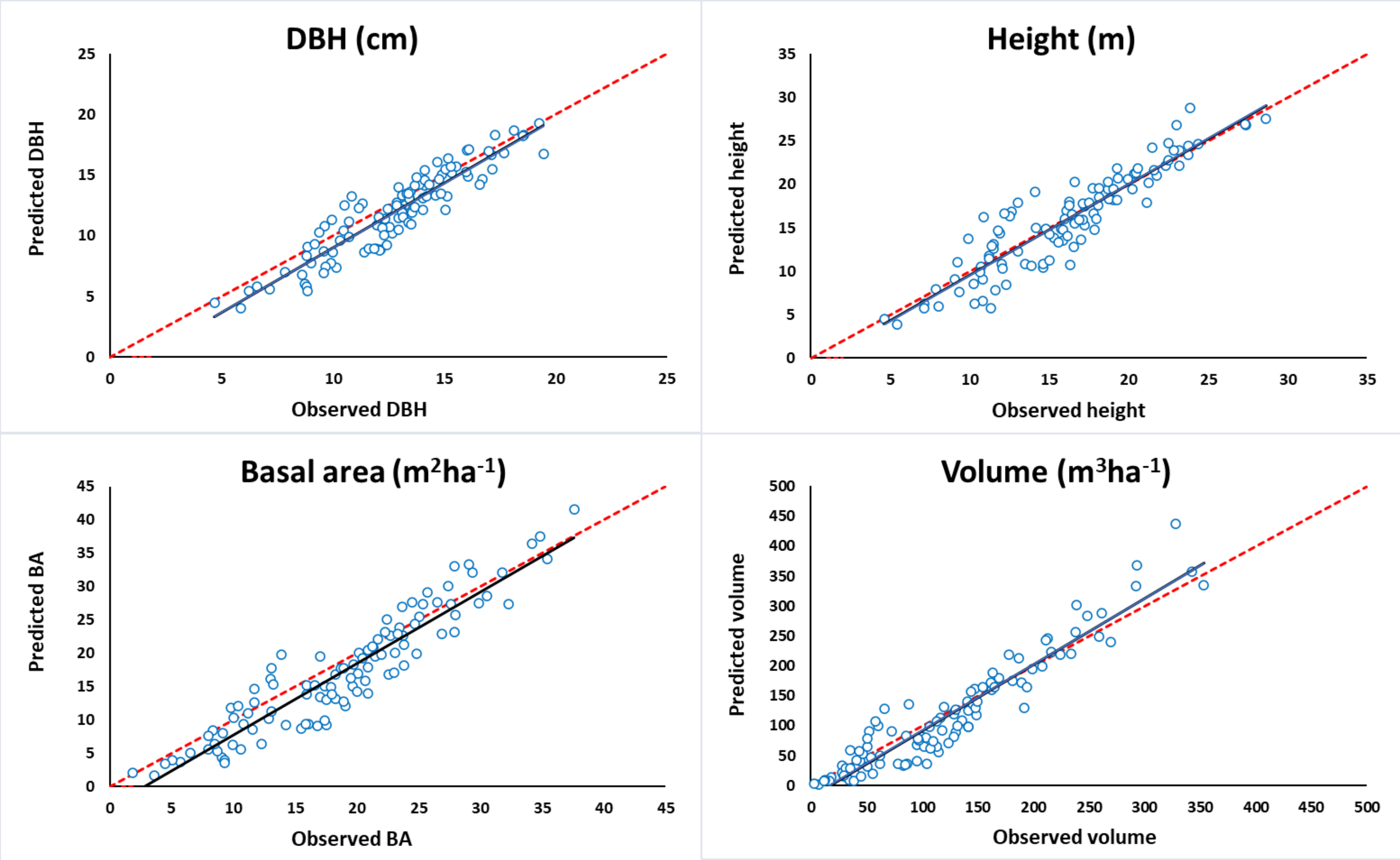


Figure 3.7 Comparison of observed and 3PG predicted values for (A) quadratic mean diameter (Dq, cm), (B) mean stem height (H, m), (C) basal area (BA, m^2ha^{-1}), and (D) volume (V, m^3ha^{-1}). Red dashed lines are identity lines (1:1), solid black lines are fitted lines from the regression.

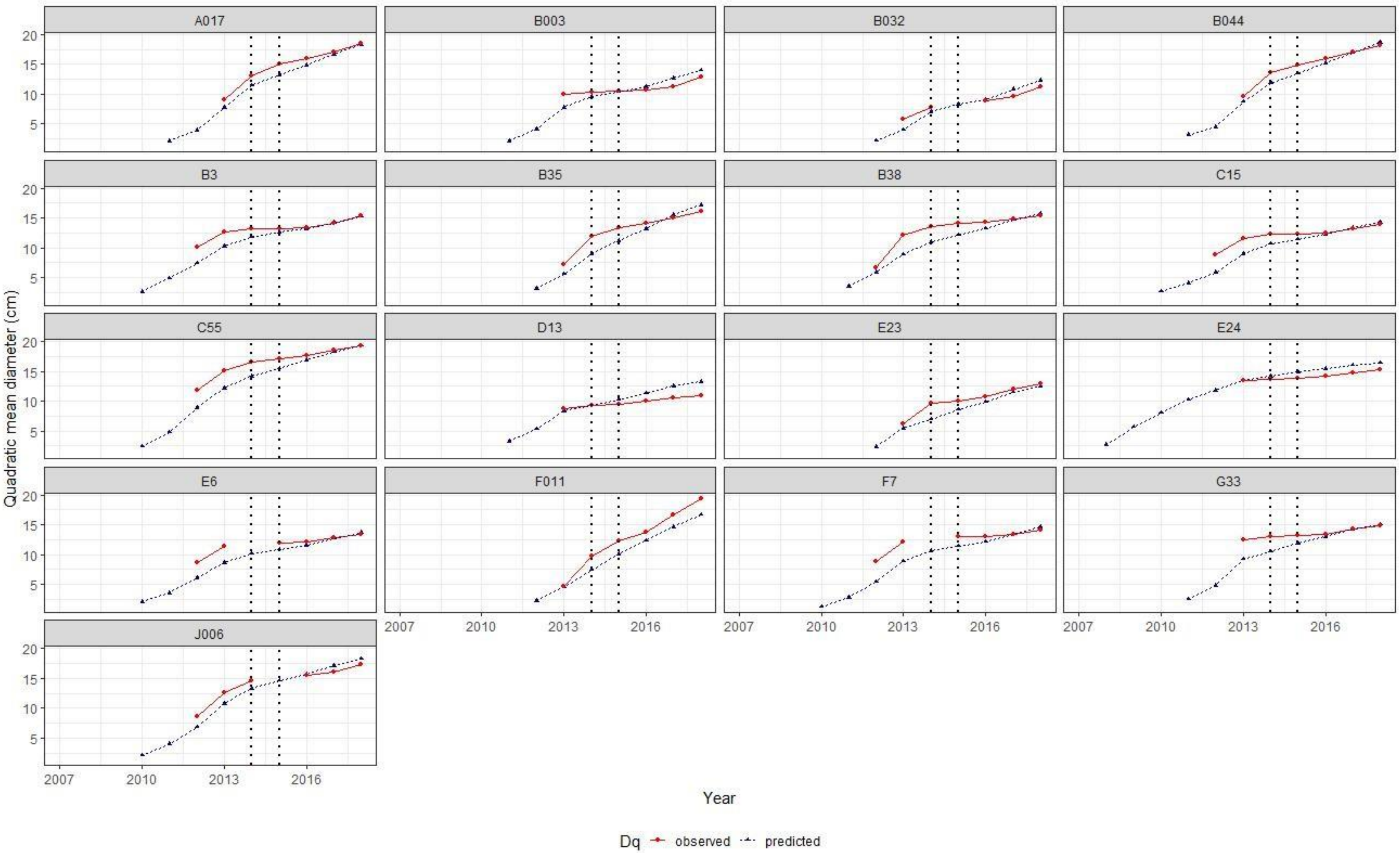


Figure 3. 8 Comparison of observed (red lines) and predicted (dark blue lines) time series quadratic mean diameter (cm) for the calibration plots**.** The two black vertical lines represent drought years (2014 – 2015)

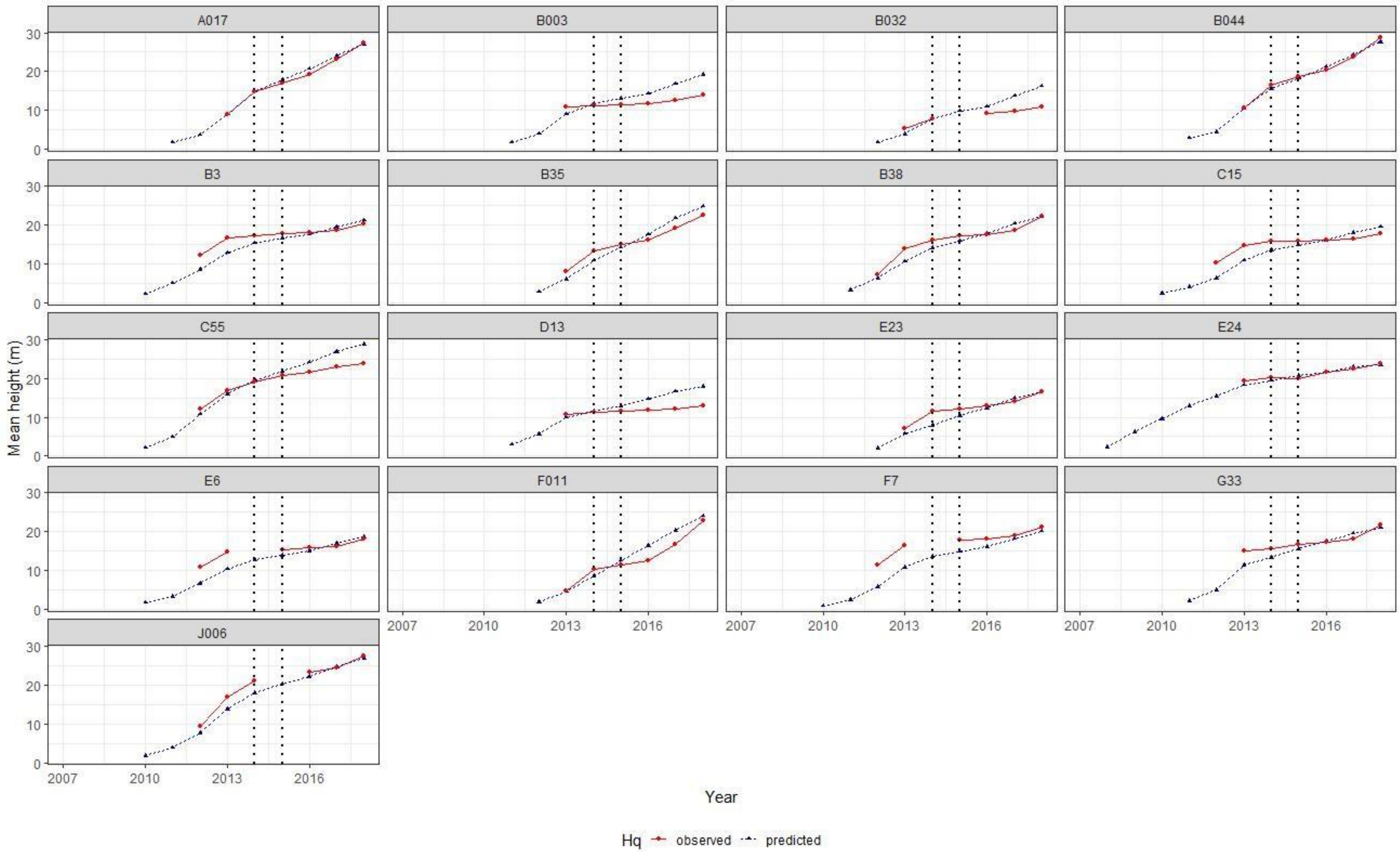


Figure 3. 9 Comparison of observed (red line) and predicted (dark blue line) time series mean stem height (m) for the calibration plots. The two black vertical lines represent drought years (2014 – 2015)

Figure 3. 10 Images of PSP and site condition during the drought year retrieved from Google Earth ©. Image A shows PSP in the Southern region compared to image B in the Northern region. Both are of the same clone type and near-natural forest. Image C in the Southern region compared to image D, planted in the Northern region.

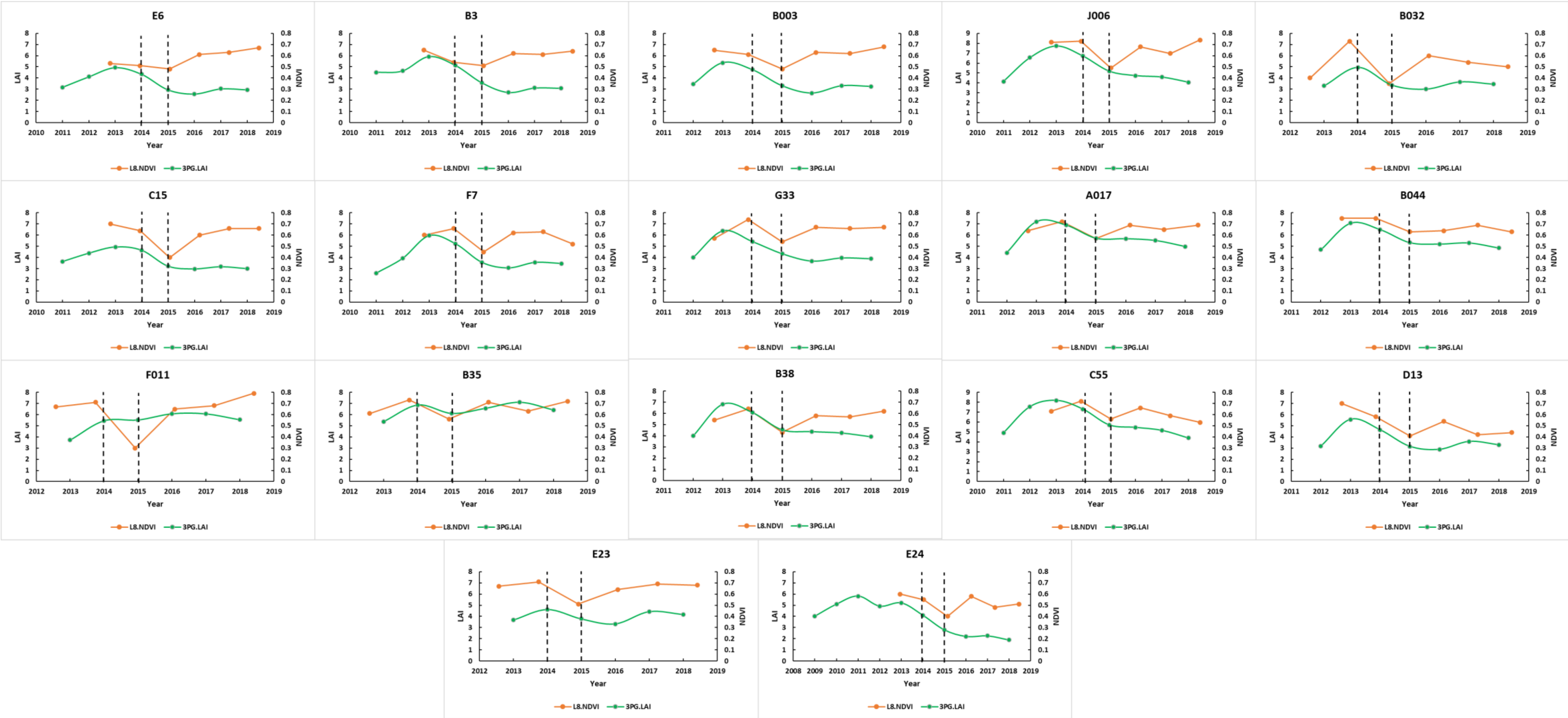
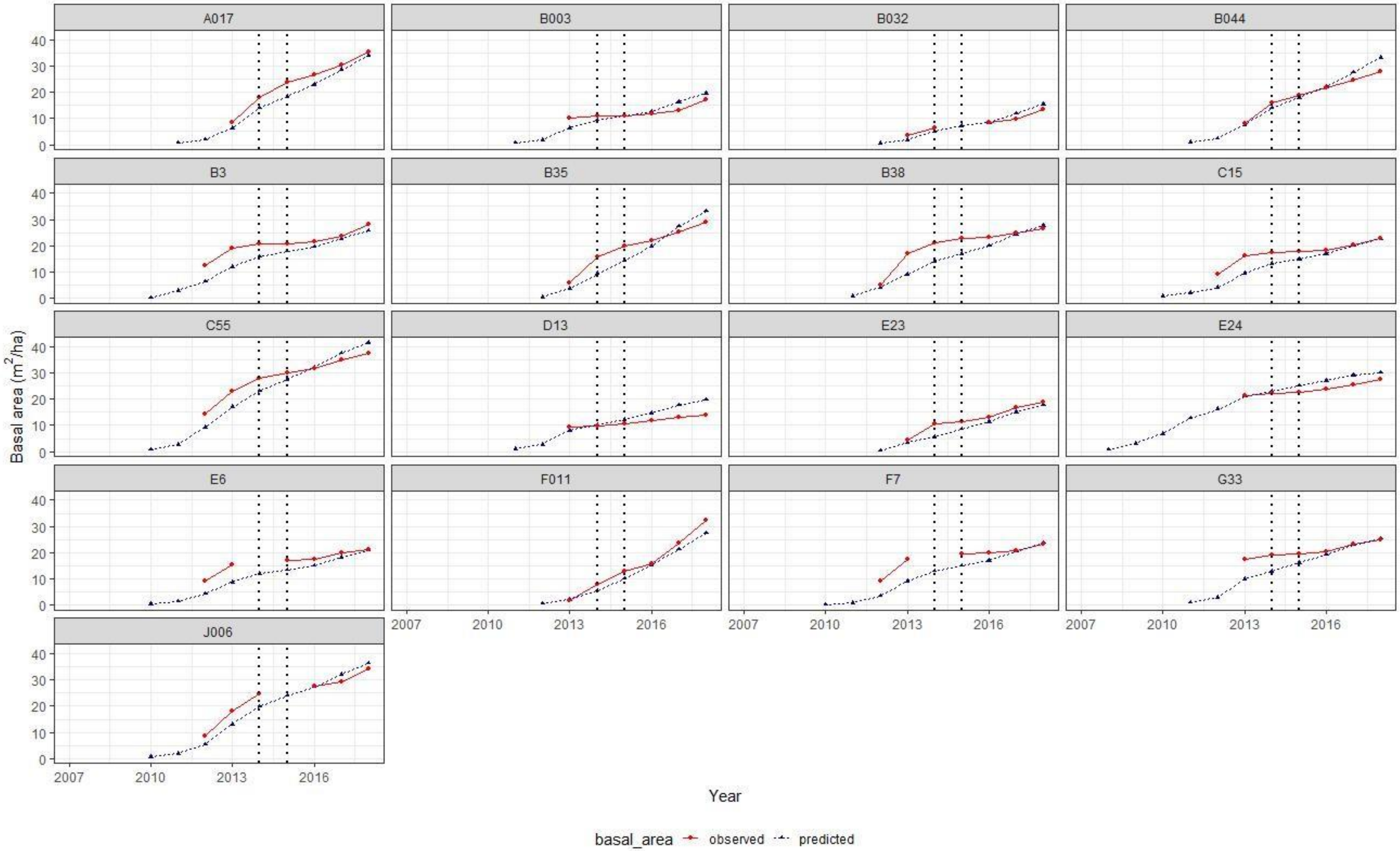


Figure 3.11 Comparison of 3PG predicted LAI values (green lines) with Landsat 8 NDVI values (orange lines) across the 17 calibration sites. The two black vertical lines represent drought years (2014 – 2015)



0Figure 3.12 Comparison of observed (red lines) and predicted (dark blue lines) time series basal area (m^2/ha) for the calibration plots. The two black vertical lines represent drought years (2014 – 2015)

Table 3.9 List and source of parameters used in the calibration of 3PG, and the result of 3PG calibration in this study

Meaning/comments	**Symbol**	**Units**	**Egxu** (Borges et al., 2012)	**Clone_15** (Almeida et al., 2004)	**Clone_22** (Almeida et al., 2004)	**This study**	**Source**
Foliage:stem partitioning ratio @ D= 2cm	pFS2	-	1.64	0.7	0.7	1	E
Foliage:stem partitioning ratio @ D= 20cm	pFS20	-	0.15	0.1	0.11	0.15	E
Constant in the stem mass vs diam. relationship	a_S	-	0.02	0.049	0.033	0.099	F
Power in the stem mass vs diam. relationship	n_S	-	3.11	2.822	2.912	2.506	F
Maximum fraction of NPP to roots	pRx	-	0.5	0.6	0.6	0.6	E
Minimum fraction of NPP to roots	pRn	-	0.1	0.07	0.12	0.1	E
Maximum litterfall rate	gammaF1	1/month	0.07	0.13	0.13	0.07	B
Litterfall rate at t = 0	gammaF0	1/month	0.001	0.00169	0.00169	0.001	B
Age at which litterfall rate has median value	tgammaF	Months	4	13	13	4	B
Average monthly root turnover rate	gammaR	1/month	0.025	0.025	0.025	0.025	B
Minimum temperature for growth	Tmin	°C	8	8	8	5	E
Optimum temperature for growth	Topt	°C	25	25	25	23	E
Maximum temperature for growth	Tmax	°C	40	36	36	40	E
Moisture ratio deficit for $f_\theta = 0.5$	SWconst	-	0.5	0.5	0.5	0.7	D
Power of moisture ratio deficit	SWpower	-	5	5	5	9	D
Value of 'm' when FR = 0	m0	-	0	0	0	0	D
Value of 'fNutr' when FR = 0	fN0	-	0.5	0.6	0.6	0.6	D
Power of (1-FR) in 'fNutr'	fNn	-	1	1	1	1	D
Maximum stand age used in age modifier	MaxAge	Years	9	9	9	9	B
Power of relative age in function for fAge	nAge	-	4	4	4	4	D
Relative age to give fAge = 0.5	rAge	-	0.95	0.95	0.95	0.95	D
Mortality rate for large t	gammaNx	%/year	0	0	0	0.6	F
Seedling mortality rate (t = 0)	gammaN0	%/year	0	0	0	1.01	F
Age at which mortality rate has median value	tgammaN	years	0	0	0	3.36	F
Shape of mortality response	ngammaN	-	1	1	1	1	F

Values source: B - Base parameter (Borges et al., 2012), D – Default (Sands & Landsberg, 2002), E - Estimated, F- fitted

List and source of parameters used in the calibration of 3PG, and the result of 3PG calibration in this study (Continues)

Meaning/comments	**Symbol**	**Units**	**Egxu** (Borges et al., 2012)	**Clone_15** (Almeida et al., 2004)	**Clone_22** (Almeida et al., 2004)	**This study**	**Source**
Max. stem mass per tree @ 1000 trees/ha	wSx1000	Kg/tree	300	180	180	300	E
Specific leaf area at age 0	SLA0	m^2/kg	13.74	11	9	13.74	B
Specific leave area for mature leaves	SLA1	m^2/kg	7.56	8	7.3	7.56	B
Age at which specific leaf area = (SLA0+SLA1/2)	tSLA	years	1.23	2.5	2.5	1.23	B
Extinction coefficient for absorption of PAR by canopy	k	-	0.5	0.5	0.5	0.5	D
Age at canopy cover	fullCanAge	Years	2	2	2	2	D
Maximum proportion of rainfall evaporated from canopy	MaxIntcptn	-	0.15	0.15	0.15	0.15	D
LAI for maximum canopy conductance	LAImaxIntcptn	-	3.33	3	3	3	D
Alpha	alphaCx	molC/molPAR	0.08	0.068	0.068	0.08	E
Ratio NPP/GPP	Y	-	0.5	0.47	0.47	0.5	B
Maximum canopy conductance	MaxCond	m/s	0.02	0.02	0.022	0.02	D
LAI for maximum canopy conductance	LAIgcx	-	3.33	3	3	3.33	D
Defines stomatal response to VPD	CoeffCond	1/mBar	0.0324	0.045	0.05	0.0324	B
Canopy boundary layer conductance	BLcond	m/s	0.2	0.2	0.2	0.2	D
Branch and bark fraction at age 0	fracBB0	-	0.59	0.3	0.3	0.59	B
Branch and bark fraction for mature stands	fracBB1	-	0.19	0.12	0.12	0.19	B
Age at which fracBB = (fracBB0+fracBB1)/2	tBB	years	2.17	2	2	2.17	B
Minimum basic density for young trees	rhoMin	t/m^3	0.382	0.48	0.4	0.382	B
Maximum basic density for older trees	rhoMax	t/m^3	0.505	0.52	0.48	0.505	B
Age at which rho = (rhoMin+rhoMax)/2	tRho	years	2.264	3	3	2.264	B
Constant in stem height relationship	aH		0.67	0	0	0.67	B
Power of DBH in stem height relationship	nHB		1.27	0	0	1.27	B

Values source: B - Base parameter (Borges et al., 2012), D – Default (Sands & Landsberg, 2002), E - Estimated, F- fitted

3.5.4 Model performance at independent validation sites

Basal area, DBH, and height predicted from running the 3PG model using the four parameter sets (Table 3.10) were used to estimate stand volume (Equation 3.4) at the fifteen validation sites described in section 3.4.4. All four parameter sets accounted for more than 60% of the variance in the observed data (Table 3.11), but the two parameter sets from Almeida et al. (2004) severely under-predicted fast-growing sites. The parameter set developed by Borges et al. (2012) had slightly greater precision ($R^2 = 0.68$) compared to this study ($R^2 = 0.65$) but performed poorly in terms of slope. (Figure 3.13, Table 3.11). Although predictions from the new parameter set simulated the observed volume reasonably well, there was notable variability (Figure 3.13; Table 3.11). Stand volume at one site, L11b, was severely underestimated by the model. However, the Brazilian "clone-level" parameter sets showed strong trend bias (Figure 3.13). The poor modelling efficiency index was because of the distance between the observed vs. 3PG prediction regression line and the identity line (1:1) (Figure 3.11). The overprediction by the 3PG model may be attributed to the available soil water values used during validation. Another possibility might be from the observational data. Esprey (2006) also observed poor agreement between observed and 3PG predictions using final observation data and recommends that the biological reasonableness of the model should be evaluated by testing model prediction against time-series data.

Table 3.10 Statistics describing the relationship between observed and predicted volumes. Statistics include root mean square error (RMSE), Nash Sutcliffe efficiency index (E), intercept, slope, p-value, and R^2.

Parameter Source	RMSE	E	Intercept	Slope	p-value	R^2	N
Egxu (Borges et al., 2012)	23.91	0.31	59.121	0.621	<0.001	0.68	15
Clone 15 (Almeida et al., 2004)	26.26	0.16	61.99	0.297	<0.001	0.64	15
Clone 22 (Almeida et al., 2004)	34.16	-0.41	65.34	0.185	<0.001	0.68	15
This study	33.02	-0.32	43.26	0.856	<0.001	0.65	15

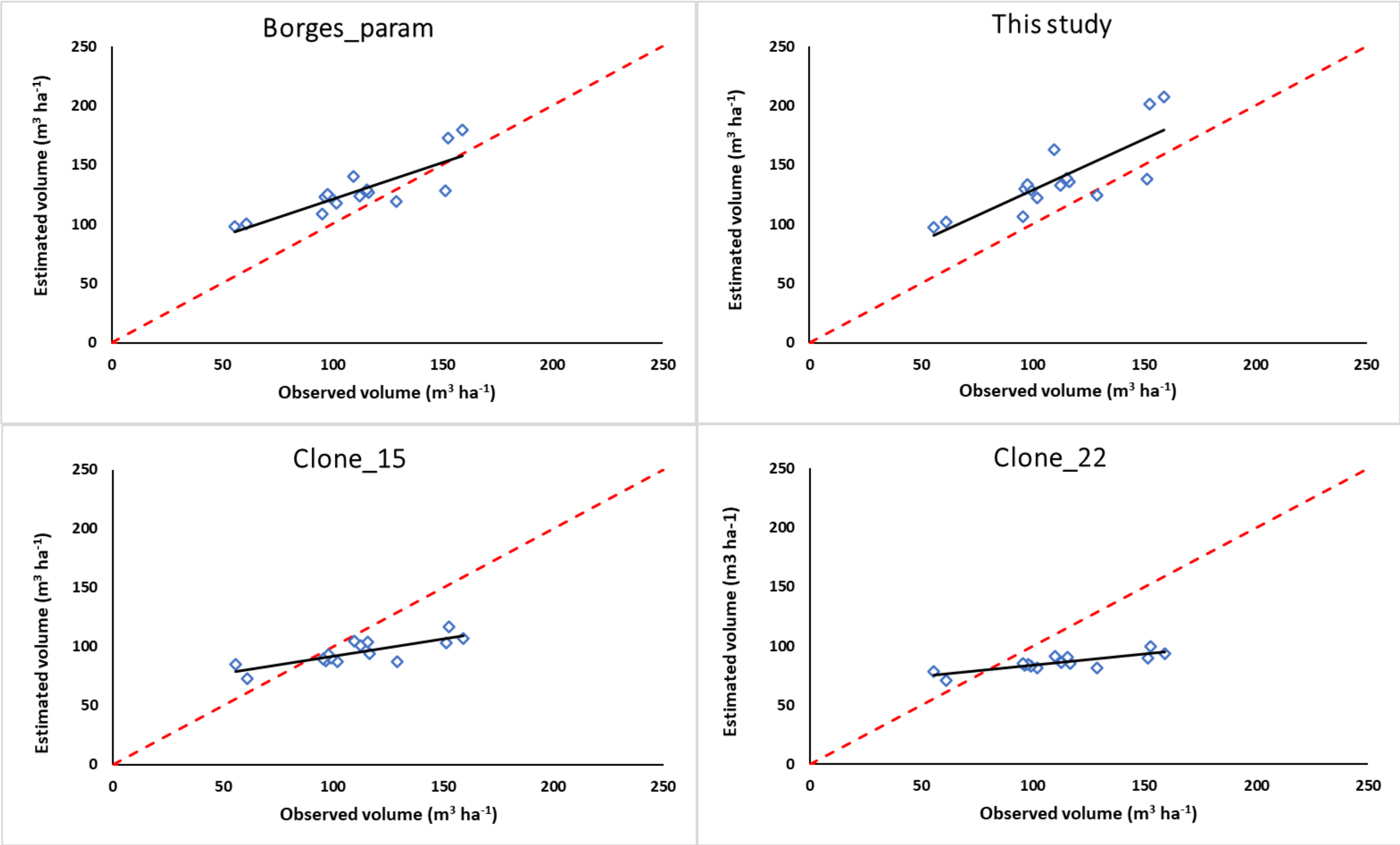


Figure 3.13 Comparison of observed and 3PG predicted volume (V, m^3ha^{-1}). Red dashed lines are identity lines (1:1), solid black lines are fitted lines from the regression equation

Chapter 4 Performance of four modelling approaches in predicting the productivity of *Eucalyptus grandis x urophylla* in coastal Zululand, South Africa

4.1 Introduction

The focus of the statistical growth and yield models in forest management has been the development of prediction tools to aid in decision making (Burkhart & Tomé, 2012). Forestry managers are practical people (Landsberg, 2003b). These models' relative simplicity and practicability have made them a default operational tool (Burkhart & Tomé, 2012). Relatively, simple models have proven useful in providing quantitative information for management and planning, predicting growth and yield, and providing product profile information (Landsberg, 2003a).

Forests are dynamic ecosystems that are constantly changing (Peng & Wen, 2006). As a result of the steady gains in the understanding of forest biology and ecology (Johnsen et al., 2001), there has been a growing awareness of the forest ecosystem complexity, which involves interactions between environmental variables, growth, and the developmental processes in trees (Gupta & Sharma, 2019). Therefore, the assumption of the traditional growth and yield model that the environment is "static" and forecasting growth and yield from historical data is no longer reliable. Consequently, these issues present forest managers with difficult questions concerning forest ecosystem management (Peng & Wen, 2006). Among others, droughts, the impact of pests and diseases, and the physiological diversity among hybrid clones are major challenges limiting the application of the conventional growth and yield model in South African short-rotation forestry (Dye, 2001; Kotze, 2018).

Additionally, social and political pressure demands that forest management take environmental and social aspects of sustainable development more explicitly (Peng & Wen, 2006; Dyer, 2007). This has created debate on the utility of simple growth and yield models versus more complex ecophysiological models. Several authors have discussed the benefits and limitations of both modelling approaches; however, the value of both modelling approaches and how they can be applied in forest ecosystem management is well captured in detail by Korzukhin et al. (1996).

In that context, the 3PG model is an interesting case. It has managed to find a niche in this continuum so that it is more considered a "hybrid model" (incorporating elements of both process-based and empirical models) (Landsberg & Sands, 2011; Weiskittel et al., 2011). Nevertheless, the question arises as to whether this leaves the model unable to provide the level of accuracy required for robust decision-making at a scale relevant to forest managers. Given that it is still more complex and data-hungry than some other approaches, a second question is if the parameterization and data acquisition requirements lead to sufficient gains in model skill? Therefore, the main objective of this chapter is to parameterize and validate three simple models for the same region. The result will be used to compare the performance of the 3PG model.

4.2 Materials and Methods

4.2.1 Site index-based model for Egxu

For decades, most commercial forestry companies in South Africa have used the empirical, stand-level growth and yield modelling system to forecast the growth and yield of *Eucalyptus grandis x urophylla* (which is mainly used as gum pulpwood without thinning) (Kotze & Fletcher, 2013). This model served as the conventional growth and yield model for planning, volume estimation, and forecast supply and was developed using different statistical regression models and functions (Kotze, 2018)**.** Stand growth and yield are projected by calibrating the model with inventory data (when available) or with site index-based defaults (when no inventory is available) (Kotze, 2018). For the study reported in this thesis, the site index-based model was used. The model requires the default site index (SI) value, SI reference age (SIRefAge), initial stem number, the survival rate at SIRefAge, and stem number at SIRefAge. Survival percentage at SIRefAge was set at ninety-five percentage (95%) for all stands. Basal area and mean height predicted by the model were used to estimate stand volume using Equation (3.4). The model has been previously calibrated for Egxu by Kotze (2018). Therefore, functions, equations, and derived parameters (coefficients are presented in Table 4.1) for modelling Egxu growth are summarized.

- **The Hossfeld function to model dominant height (Palahí et al., 2004)**

$$HD_2 = \frac{AGE_2^2}{b_1 + AGE_2 [(AGE_1/HD_1) - b_2 \times AGE_1 - (b_1/AGE_1) + b_2 \times AGE_2]}$$

Equation 4. 1

Where HD_1 = Dominant height in m at AGE_1; HD_2 = projected dominant height at AGE_2; AGE_i = stand age in years

- **The Clutter-Jones function to model tree survival (Clutter & Jones, 1980)**

$$TPH_2 = \left[TPH_1^{b_1} + b_2 \cdot \left[\left(\frac{AGE_2}{100}\right)^{b_3} - \left(\frac{AGE_1}{100}\right)^{b_3}\right]\right]^{\frac{1}{b_1}}$$

Equation 4. 2

Where TPH_1 = stems per hectare at AGE_1; TPH_2 = projected stems per hectare at AGE_2.

- **The Schumacher-type stand-level basal area function for untinned stands to model basal area (Pienaar & Harrison, 1989)**

$$BA_2 = exp\left[ln(BA_1) + b_1 \cdot \left(\frac{1}{AGE_2} - \frac{1}{AGE_1}\right) + b_2 \cdot \left(ln(TPH_2) - ln(TPH_1)\right) + b_3 \cdot \left(ln(HD_2) - ln(HD_1)\right) + b_4 \cdot \left(\frac{ln(TPH_2)}{AGE_2} - \frac{ln(TPH_1)}{AGE_1}\right) + b_5 \cdot \left(\frac{ln(HD_2)}{AGE_2} - \frac{ln(HD_1)}{AGE_1}\right)\right]$$

Equation 4. 3

Where BA_1 = basal area per hectare at AGE_1; BA_2 = basal area per hectare at AGE_2; b_1, b_2, b_3, b_4, and b_5 are estimated parameters.

- **Multiple linear regression models to model standard deviation of DBH and minimum DBH (Kassier, 1993)**

$$Dsdev = b_1 + b_2 \cdot AGE + b_3 \cdot BA$$

Equation 4. 4

Where Dsdev = standard deviation of Dmean; BA = basal area per hectare; AGE = stand age in years

$$\boldsymbol{Dmin = (b_1 - b_2 \cdot (Dsdev/Dmean) - b_3 \cdot (TPH/SPH) + b_4 \cdot AGE) \cdot Dmean}$$

Equation 4. 5

Where Dmin = minimum DBH; TPH = number of trees per ha; PSPH = Stems Per Hectare; Dmean = arithmetic mean dbh.

- **Linear model to estimate mean height from dominant height (Kassier, 1993)**

$$HM = b_1 + b_2 \cdot HD - b_3 \cdot (Dsdev/Dmean)$$

Equation 4. 6

Where HM = mean height in m; HD = dominant height in m; Dmean = arithmetic mean dbh; Dsdev = standard deviation of Dmean.

Table 4.1 Coefficients derived for the site index-based model by Kotze (2018)

Function	**Coefficients**				
	b₁	***b₂***	***b₃***	***b₄***	***b₅***
Hossfeld function (Equation 4.1)	0.074177	0.023642			
Clutter Jones function (Equation 4.2)	-0.0477040	0.0206090	0.8280200		
Schumacher-type stand-level basal area function (Equation 4.3)	6.75324	0.33312	0.85559	0.75451	0.072155
Standard deviation of DBH (Equation 4.4)	0.87858	0.17123	0.042825		
Minimum DBH (Equation 4.5)	0.80906	1.97080	0.011475		
Linear model to estimate mean height from dominant height (Equation 4.6)	1.91604	0.89072	6.83548		

4.2.2 Cumulative Aridity Index

Aridity indices are quantitative indicators of the degree of water deficiency (in relation to evaporative demand vs. supply) at a given location (Stadler, 2005). Numerous aridity indexes have been proposed. However, the aridity index widely used considers rainfall, temperature, and evaporation in their formulation (see Stadler (2005) for a list of selected aridity indices). The majority of approaches to quantitatively calculate aridity are either constrained by the data requirements of the parameters defined or ineffective representation of the concept (Stadler, 2005). Therefore, Thornthwaite (1948) approach, moisture index (Im), was widely accepted due to the simplicity of the data requirements and general agreement with world vegetation patterns (Stadler, 2005). This approach classified a location's climate (moist or dry) based on the balance of water supply and demand and was considered in this study. The equation is given as

$$I_m = 100\left[\frac{P}{PE} - 1\right]$$

Equation 4. 7

Where P = Precipitation (mm), PE = Potential Evapotranspiration (mm).

For analyses undertaken in this thesis, the R package ClimClass 2.1.0 (Eccel et al., 2016), freely available at *R* CRAN repository (https://CRAN.R-project.org/package=ClimClass), was used to calculate the aridity index for each site. Monthly weather variables (precipitation, minimum and maximum temperature) from 2008 – 2018 generated from the Random Forest model (See Chapter 2) were used. The *R* package calculates the *climate normals* (mean monthly values from the monthly series of temperature and precipitation) using the function *climate.* The mean monthly extra-atmospheric radiation is calculated using latitude and day of the year (this contains a vector with middle days for every month in a year) by employing the *ExAtRa* function. The Hargreaves' formula was used to calculate Potential evapotranspiration (PE), and a vector of twelve (12) coefficients was used to adjust the Hargreaves' estimation of PE by including *coeff_Hargr* in the code.

Furthermore, variable *monthly* was set to *FALSE* to focus on cumulative aridity of the site for the period specified (2008 – 2018). The output contains a single line data frame with the desired aridity index (the package calculates six aridity indices according to different authors, see Eccel et al. (2016) for full list). The calibration sites were used to parameterize the model. Aridity index from planting date to enumeration date was calculated for each calibration and validation site. The values obtained from the calibration sites were correlated with the measured volume in 2018 using simple linear regression analysis. The linear regression equation and coefficients derived were used to estimate volume for the validation site (Equation 4.8).

Table 4.2 Aridity classification according to Thornthwaite (1948)

Classification	Value
Arid (E)	-60 to -40
Semiarid (D)	-40 to -20
Dry sub-humid (C1)	-20 to 0
Moist sub-humid (C2)	0 to 20
Humid (B1)	20 to 40
Humid (B2)	40 to 60
Humid (B3)	60 to 80
Humid (B4)	80 to 100
Perhumid (A)	> 100

4.2.3 Cumulative Rainfall

Water availability has been identified as the critical limiting factor to forest growth in South Africa (Dye, 2000; Edwards & Roberts, 2006). Anecdotal evidence for the Zululand region (where our study is focused) suggests that a very simple cumulative model based on rainfall alone may be quite adequate for predicting volume variability in eucalypt plantations. To that end, we also evaluated the performance of a simple cumulative rainfall received over the stand age and for each PSP in predicting the volume of *Eucalyptus* hybrid.

For the simple cumulative rainfall model, the sum of rainfall received for each stand age in the calibration site was calculated from the weather data obtained in chapter two. A simple linear regression analysis was used to correlate these values with observed volume data measured in 2018. Weather data were generated for the validation sites using the same spatial interpolation method described in chapter two. The coefficients derived from the calibration sites were used to estimate volume for the validation sites (Equation 4.8)

$$Y = a + bX$$

Equation 4. 8

Where Y is the predicted volume (m^3ha^{-1}), X is the explanatory variable, a is the intercept, and b is the slope of the line. See Table 4.3 for values of a and b.

4.3 Results and Discussion

4.3.1 Parameterization of cumulative rainfall and aridity index

The simple linear regression calculated to predict stand volume based on cumulative rainfall received over the stand age at the calibration sites resulted in a significant regression equation ($F(1, 15) = 23.44$, $p < 0.001$) (Table 4.3), accounting for 61% of the variance in the actual volume (Figure 4.1). The simple linear regression calculated to predict stand volume based on cumulative aridity over the stand age also results in a significant regression equation ($F(1, 15) = 35.21$, $p < 0.001$) (Table 4.3), explaining 70% of the variance in actual volume (Figure 4.1). Both models produced good site ranking, but more clearly with the aridity index model (Figure 4.1). However, this ranking is more of a “site condition type” (wet, medium, and dry), not necessarily yield. Classification of the PSPs aridity, according to Thornthwaite (1948), is presented in Table 4.4. From the classification, all sites in the Northern region were classified as “Semiarid”, three sites at the southern region, which are also close to the coast, were classified as “moist sub-humid”, and sites around the center of the study area were classified as “dry sub-humid” (Table 4.4).

Table 4.3 Regression analysis of the cumulative rainfall and aridity index model at the calibration sites

Model		Coefficients	Standard error	p-value	R^2	N
Cumulative rainfall	Intercept	-227.443	90.44	<0.05	0.61	17
	Slope	0.0619	0.013	<0.001		
Cumulative aridity index	Intercept	296.613	20.16	<0.001	0.70	17
	Slope	6.59	1.111	<0.001		

Table 4.4 Classification of the calibration sites' aridity according to Thornthwaite (1948)

Compartment name	Age	Im	Classification
Futululu_E6a	8.0	-26.65	Semiarid
Mavuya_B3a	8.3	-18.06	Dry sub-humid
Mtubatuba_B003	7.1	-26.95	Semiarid
Nseleni_J006	8.1	-10.18	Dry sub-humid
Nyalazi_B032	6.1	-30.86	Semiarid
Palm_Ridge_C15a	8.2	-23.16	Semiarid
Salpine_F7	7.7	-15.5	Dry sub-humid
Salpine_G33b	7.4	-10.48	Dry sub-humid
Siyaqhubeka_A017	7.1	2.87	Moist sub-humid
Siyaqhubeka_B044	7.2	5.34	Moist sub-humid
Siyaqhubeka_F011A	6.0	-5.78	Dry sub-humid
South_Areas_B35b	6.3	-3.84	Dry sub-humid
Terranera_B38	7.4	-0.6	Dry sub-humid
Terranera_C55	8.2	0.11	Moist sub-humid
Trust_D13	7.2	-26.39	Semiarid
Trust_E23	6.1	-23.18	Semiarid
Trust_E24	10.4	-23.98	semiarid

4.3.2 Performance of the simple models at the independent validation sites

The stand volume predicted by the three simple models (Cumulative rainfall, Cumulative aridity index, and the site index-based model) was compared to the actual volume estimated at the validation sites. The site index-based model had the highest precision (R^2 =0.85), followed by the cumulative rainfall (R^2 =0.80), while the cumulative aridity performed poorly (R^2 =0.003). Out of all four models (including 3PG), the site index-based model was the best performing model in terms of all metrics used (Table 4.5). Overall, all the models showed good site ranking except for the cumulative aridity index model (Figure 4.2). Although the site index-based simulated the observed volume reasonably well, the fast-growing sites were under-predicted. In contrast, these sites were severely over-predicted by the cumulative rainfall model and 3PG (Figure 4.2). This is surprising given that the trees were growing during a drought cycle. In principle, the site index-based model is based on SI values derived from historical data; as a result, we would have expected the model to overestimate because we assume they were developed for typical conditions.

Table 4.5 Statistics describing the relationship between observed and predicted stand volumes by the different modelling approaches. Statistics include root mean square error (RMSE), Nash Sutcliffe efficiency index (E), intercept, slope, p-value, R^2

Model	RMSE	E	Intercept	Slope	p-value	R^2	N
Cumulative rainfall	30.89	-0.152	-24.19	1.405	<0.001	0.80	15
Cumulative aridity index	50.06	-2.02	147.26	-0.049	>0.05	0.003	15
Site index-based model	14.44	0.75	-8.40	1.005	<0.001	0.85	15
3PG model	33.02	-0.32	43.26	0.856	<0.001	0.65	15

The simple cumulative rainfall model and the site index-based model performed well in site ranking, making them valuable tools for forest management in the Zululand region. As a result, the question of why we need complex process-based models when simple models work well arises. One answer to this question is that process-based models such as the 3PG model can be used to formulate hypotheses and exploration of these scenarios, quantify the influence of external factors on forest productivity, and also estimate the potential productivity of a site that has no previous field data (Almeida et al., 2004). It is important to reiterate that models are nothing more than abstract representations of the system being modelled (Landsberg & Sands, 2011). Therefore, if the system to be managed or the question to be addressed is simple, simple statistical models can be applied. If the system deals with processes, structures, and cause-effect questions, complex models such as 3PG can be applied (Pretzsch, 2009). No single model is *"omnipotent"*. The strength and weakness of both empirical and process-based models have been argued by Korzukhin et al. (1996).

However, several authors have asserted that the future of forest modelling lies in models that combine the strength of both modelling approaches (hybrid models) (Battaglia & Sands, 1998; Mäkelä et al., 2000; Landsberg, 2003b). A hybrid approach based on the 3PG model and an empirical model was developed by Almeida et al. (2003) to manage fast-growing *Eucalyptus grandis* hybrid plantations in Brazil. Additionally, 3PG has also been hybridized with other models such as 3PGH – combining 3PG with TOPMODEL to improve water balance sub-model (Almeida & Sands, 2016; Almeida et al., 2016), 3PGN – combining 3PG with a carbon balance model, ICBM/2N to predict FR (Xenakis et al., 2008), 3PG+/CAT – integrating 3PG+ into the Catchment Analysis Tool (CAT) framework to predict impacts of plantation on water balances of catchments, (Nolè et al., 2009) coupled 3PGS with a modified soil respiration model to estimate net ecosystem fluxes.

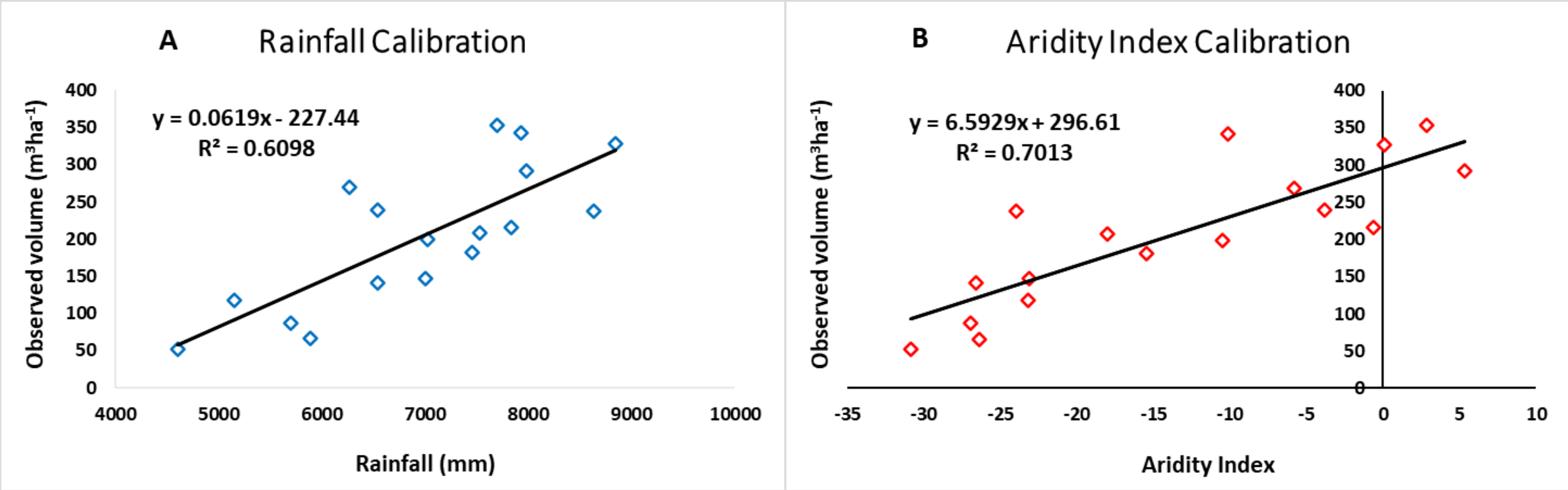


Figure 4.1 Relationship between observed volume in 2018 and (A) cumulative rainfall over the stand age (B) cumulative aridity index over the stand age

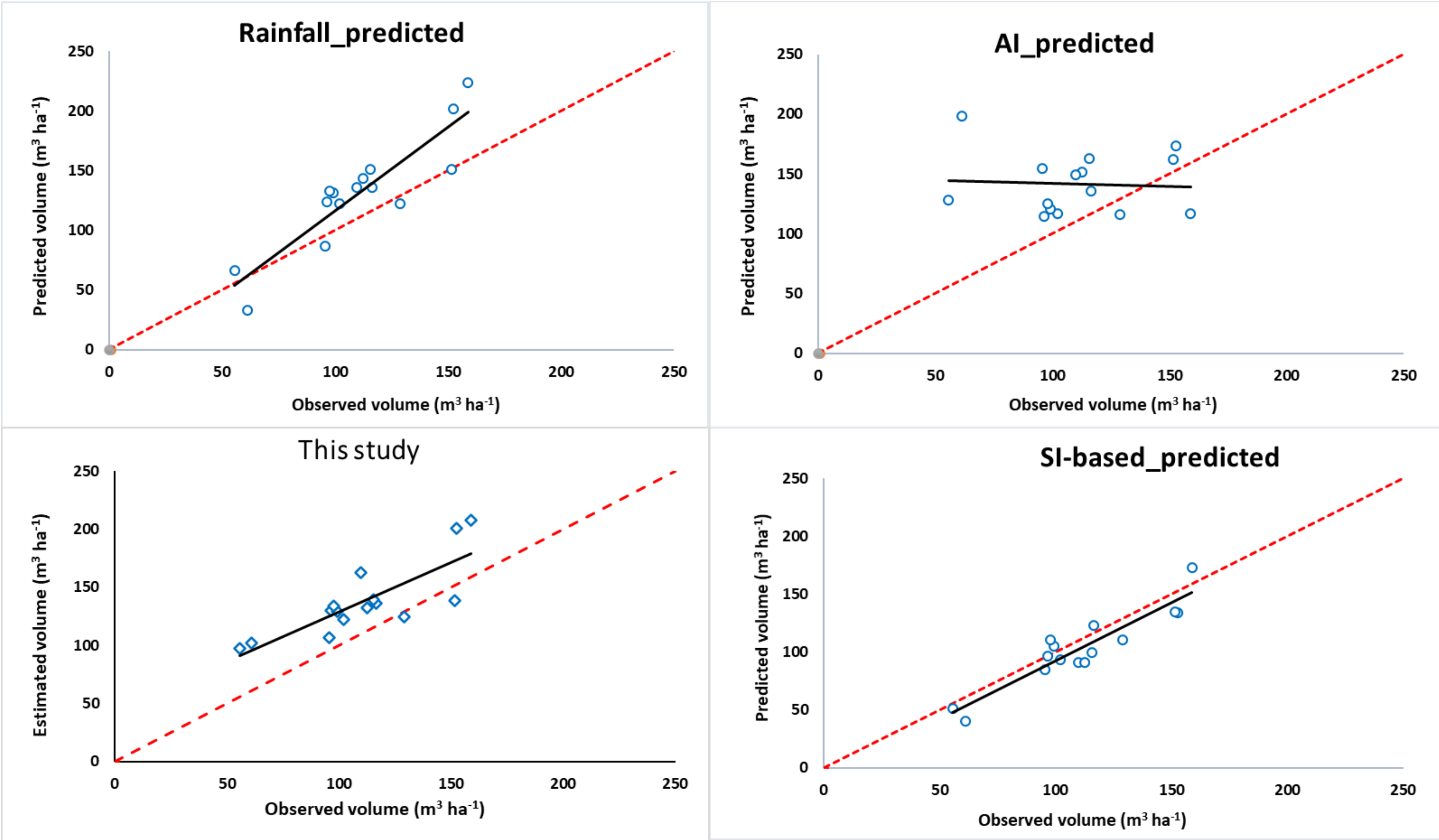


Figure 4.2 Comparison of observed and predicted stand volume (V, m^3ha^{-1}) using (A) Cumulative rainfall model (B) Cumulative aridity index model C) the 3PG model (D) Site index-based model. Red dashed lines are identity lines (1:1), solid black lines are fitted lines from the regression.

Chapter 5 General Conclusions and Comments

The primary aim of the research reported in this thesis was to calibrate and test the 3PG model for stands of *Eucalyptus grandis x urophylla* hybrid trees growing under South African conditions and to compare its performance with other models, including a statistical growth and yield model used by the two forestry companies who were partners in the research. To achieve this, some key research questions were raised (section 1.6), and specific objectives were set (section 1.5) to address these questions.

Spatial Interpolation of weather data for the ungauged plantations (Objective 1)

Based on this objective, we conclude that both of the tested spatial interpolation techniques can be calibrated to predict long-term weather data in the study area. Depending on the temporal scale desired, Meteoland can be used to predict daily weather data. However, in terms of reduced errors and higher precision, the RF algorithm outperformed the Meteoland approach, which underscores the significance of machine learning in spatial interpolation. Additionally, the RF captured the pattern of rainfall variation previously reported for this region (du Plessis & Zwonlinski, 2003). One possible explanation could be the inclusion of the 'distance to coast' variable in the RF model, which was not part of the input variable for Meteoland. This variable has been reported to correlate with weather patterns (van Niekerk & Joubert, 2011). Therefore the Meteoland package can be improved by including this variable in its routine.

We acknowledged the possibility of human error when handling large datasets such as those used in this study. This led to the development of algorithms in R to handle, process, and prepare the input files for both models. These algorithms are a significant outcome of this study and can further be developed as an R package. Apart from the dataset generated in this study, another valuable outcome of this study was adopting a simple modelling technique that uses readily available data, is available in R software, and is easily integrated into the forest simulation model used in this study. This makes it reproducible for future research work.

This study generated monthly weather data for the Zulualnd region for ten years (2008 – 2018), and we acknowledge that it was time-consuming. To generate weather data for a longer time, we propose that future research consider merging these spatial interpolation techniques with the GIS software to create a weather surface map for this region. This will facilitate the extraction of point weather data across the entire region.

Calibration and validation of the 3PG model (Objective 2)

The *R3PG* package, a Fortran implementation of the 3PG model, embedded into an *R* package developed by Trotsiuk et al. (2020), was used to perform simulation runs and parameter optimization. The package was designed to run simulations on a single site; as a result, we created a loop function (Figure 3.2) to run the *R3PG* for multiple sites. The integration of algorithms developed (section 3.4.3) into *R3PG* facilitated model parameterization compared to the previous implementation of 3PG in visual basic. Another innovation in this study is the simple goodness-of-fit approach *(eRMSE)*, developed to select the parameter set with minimized residuals, low bias, and line of best fit close to the identity line (slope = 1, intercept = 1).

The PSPs used for calibration represents a broad range of site and climatic conditions; therefore, the parameter set developed in this study is expected to be applicable for E. gxu hybrids grown in South Africa. The simple parameter calibration from a base parameter (previously published for the same species but in a different region) and fitting allometric parameters produced a parameter set that characterized E. gxu hybrids' growth in an unthinned, short rotation stands. While there were apparent performance gains at the calibration stage (Table 3.9), there were also some drawbacks.

- Lack of time-series stem, foliage, and root biomass data to test 3PG performance during parameter estimation. This may result in a good fit of outputs to observed data for the wrong reason (Sands, 2004).
- There was inadequate information regarding the effective rooting depth of the soil.
- Limited information on available soil water in this region. This was estimated using soil physical properties. However, soil physical properties data were not available for the validation sites; hence the average ASW for the calibration sites was used.
- Lack of time-series litterfall and ground-truthed SLA or LAI data.

The lack of this information made it difficult to check if the parameter values were biologically reasonable. Additionally, we acknowledge the limitation of assigning values to soil fertility indices in the 3PG model. A fixed value of 0.5 was used for all sites because there was no beneficial empirical relationship between variables (soil physical and chemical properties, site index, and ASW) and optimized fertility rating. This variable affects the canopy quantum efficiency and root biomass allocation. As a result, other parameters related to biomass production and allocation will be forced to account for variability in biomass during parameter estimation (Forrester et al., 2021).

Parameters values that differ from the base parameter during parameter estimation were majorly parameters related to the allocation of biomass produced - *foliage:stem partitioning ratio (D = 2cm) (pFS2)*, *maximum fraction of NPP to root (pRx)*, *minimum and optimum temperature for growth (Tmin and Topt)*. Almeida et al. (2004) also reported these intra-specific parameter differences between clones (Clone 15 and Clone 22). We found that 3PG captured the decline in growth rate during the dry years. Although no tree death was recorded due to the drought, mortality occurred at post-planting in most sites, and this was accounted for by fitting the density-independent mortality parameters. We conclude that the 3PG model provides a useful tool for modelling the growth pattern in trees in response to environmental changes.

Performance of four modelling approaches in predicting the productivity of *Egxu* in coastal Zululand (Objective 3&4)

The performance of the site index-based model can be attributed to the use of improved SI values calculated after inventory. This modelling approach can still be regarded as a valuable tool for forest management in the Zululand region. At the calibration stage, the cumulative rainfall and cumulative aridity index model produced a good site ranking. However, it was more related to "site condition" (wet, medium, and arid dry) not necessary yield. At the validation stage, the cumulative aridity index had a poor performance. The cumulative rainfall model's performance underscores the importance of water availability as a critical determinant in forest growth in the coastal Zululand region. This study showed that a simple cumulative rainfall might be adequate for forecasting volume variability in eucalypt plantations in coastal Zululand. However, we suggest that the age of the stand should be included in the simple model. This is due to the spatiotemporal variability of rainfall in this region; it is possible to have the same cumulative rainfall in two different sites, one over four years and the other over eight years. This will result in the same estimated volume even though the growth rate over time might be completely different.

The 3PG model also performed well in site ranking and gave a realistic prediction of tree growth in response to environmental changes. However, the over-predictions observed can be attributed to the fact that key input variables (such ASW, FR, weather data, soil depth) were estimated rather than directly observed or measured in this study. The 3PG model demonstrated its potential in exploring scenarios ("*what if" questions*), such as the case of the five sites suspected to have higher-than-normal access to groundwater. The model provided additional estimates of productivity such as net primary production (NPP) and leaf area index (LAI) and was able to identify site factors limiting growth. These findings revealed that the 3PG model resulted in significant improvements to simulate forest growth.

The following can be considered for future research:

- A measurement routine to collect time-series biomass and litterfall data from the PSPs is recommended. These variables are strongly tied to the internal dynamics of the 3PG model. They can be used to re-calibrate the 3PG model using the experience gained in this study. This will help us better understand the hybrid's growth pattern and boost our confidence in the 3PG modelling capability.
- The Depth-to-water (DTW) index has been shown to correlate with soil water availability and tree height (Oltean et al. 2016). We recommend that this index be tested and validated across many forestry regions. It has the potential of estimating minimum and maximum available soil water.
- There is still a need to objectively characterize the fertility of a site. It will be beneficial to couple the 3PG model with soil nutrition models that explain the dynamics of nutrient fluxes such as ICBM/2N (Xenakis et al., 2008) or SNAP (Paul et al., 2002). Although this will add to the complexity of the model, it will improve our understanding of soil fertility dynamics throughout a stand rotation.
- The developers of the R3PG package added a spatial simulation feature to the model, which they used to simulate stand biomass on a 1 x 1 grid. This enables the prediction of forest productivity on a large scale or for sites that have no previous data, and explores scenarios such as the impact of climate change, pest and disease infestation using remotely-sensed data.
- Consider developing a linkage between the 3PG model and the statistical growth and yield model.

References

AgriMetSoft. (2019). *Agricultural and Meteorological Software. Online Calculator. Available at https://www.agrimetsoft.com/calculators/*.

AgriSA. (2016). *A Raindrop in the Drought. Report to the Multi-Stakeholder Task Team on the Drought - Agri SA's status report on the current drought crisis*. http://www.nstf.org.za/wp-content/uploads/2016/06/Agri-SA-Drought-Report_CS4.pdf

Albaugh, J. M., Dye, P. J., & King, J. S. (2013). Eucalyptus and Water Use in South Africa. *International Journal of Forestry Research*, *2013*, 11. https://doi.org/10.1155/2013/852540

Alig, R., Adams, D., Joyce, L., & Sohngen, B. (2004). Climate Change Impacts and Adaptation in Forestry: Responses by Trees and Markets. *Choices. The Magazine of Food,Farm, and Resources Issues*, *19*(3), 1–6. www.choicesmaeazine.ore/scriPts/~rintVersion.h?ID=OO4-3-07

Almeida, A. C., Sands, P. J., Bruce, J., Siggins, A. W., Leriche, A., Battaglia, M., & Batista, T. R. (2009). Use of a spatial process-based model to quantify forest plantation productivity and water use efficiency under climate change scenarios. *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings*, *January 2014*, 1816–1822.

Almeida, A.C., Maestri, R., Landsberg, J. J., & Scolforo, J. R. S. (2003). Linking process-based and empirical forest models in Eucalyptus plantations in Brazil. In A. Amaro & M. Tome (Eds.), *Modelling Forest Systems.* (pp. 63–74). CAB International.

Almeida, Auro C. (2003). *Application of a process-based model for predicting and explaining growth in Eucalyptus Plantations*. The Australian National University.

Almeida, Auro C. (2018). Forest growth modelling for decision making: practical applications and perspectives. *New Frontiers in Forecasting Forests, Stellenbosch South Africa, 25 - 28/9/2018*, 60–62.

Almeida, Auro C., Landsberg, J. J., & Sands, P. J. (2004). Parameterisation of 3-PG model for fast-growing Eucalyptus grandis plantations. *Forest Ecology and Management*, *193*(1–2), 179–195. https://doi.org/10.1016/j.foreco.2004.01.029

Almeida, Auro C., Landsberg, J. J., Sands, P. J., Ambrogi, M. S., Fonseca, S., Barddal, S. M., & Bertolucci, F. L. (2004). Needs and opportunities for using a process-based productivity model as a practical tool in Eucalyptus plantations. *Forest Ecology and Management*, *193*(1–2), 167–177. https://doi.org/10.1016/j.foreco.2004.01.044

Almeida, Auro C., & Sands, P. J. (2016). Improving the ability of 3-PG to model the water balance of forest plantations in contrasting environments. *Ecohydrology*, *9*(4), 610–630. https://doi.org/10.1002/eco.1661

Almeida, Auro C., Smethurst, P. J., Siggins, A., Cavalcante, R. B. L., & Borges, N. (2016). Quantifying the effects of Eucalyptus plantations and management on water resources at plot and catchment scales. *Hydrological Processes*, *30*(25), 4687–4703. https://doi.org/10.1002/hyp.10992

Antoine, F. (2021). View and Browse Code Using Flow Diagrams. R package version 0.02. *Https://CRAN.R-Project.Org/Package=flow*. https://cran.r-project.org/web/packages/flow/index.html

Apaydin, H., Kemal Sonmez, F., & Yildirim, Y. E. (2004). Spatial interpolation techniques for climate data in the GAP region in Turkey. *Climate Research*, *28*(1), 31–40. https://doi.org/10.3354/CR028031

Battaglia, M., & Sands, P. J. (1998). Process-based forest productivity models and their application in forest management. *Forest Ecology and Management*, *102*(1), 13–32. https://doi.org/10.1016/S0378-1127(97)00112-6

Baudoin, M. A., Vogel, C., Nortje, K., & Naik, M. (2017). Living with drought in South Africa: lessons learnt from the recent El Niño drought period. *International Journal of Disaster Risk Reduction*, *23*, 128–137. https://doi.org/10.1016/j.ijdrr.2017.05.005

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST 2016 25:2*, *25*(2), 197–227. https://doi.org/10.1007/S11749-016-0481-7

Binkley, D., Campoe, O. C., Alvares, C., Carneiro, R. L., Cegatta, Í., & Stape, J. L. (2017). The interactions of climate, spacing and genetics on clonal Eucalyptus plantations across Brazil and Uruguay. *Forest Ecology and Management*, *405*, 271–283. https://doi.org/10.1016/j.foreco.2017.09.050

Binkley, D., & Fisher, R. F. (2013). Ecology and Management of Forest Soils: Fourth Edition. In *Ecology and Management of Forest Soils: Fourth Edition*. https://doi.org/10.1002/9781118422342

Booth, T. H. (2013). Eucalypt plantations and climate change. *Forest Ecology and Management*, *301*, 28–34. https://doi.org/10.1016/j.foreco.2012.04.004

Borges, J. S., Neves, J. C. L., Lourenço, H. M., de Barros, N. F., & Dias, S. C. M. (2012). Parameterization of the 3PG model for Eucalypt in the region of Cerrado in Minas Gerais State. *Ciência Florestal, Santa Maria*, *22*(3), 567–578. https://doi.org/10.5902/198050986623

Breiman, L. (2001). Random Forests. *Machine Learning 2001 45:1*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Burengengwa, G. (2020). *Comparison of approaches for spatial interpolation of weather data on a specific date by*. *April*.

Burkhart, H. E., & Tomé, M. (2012). Modeling forest trees and stands. In *Springer* (1st ed.). Springer Netherlands. https://doi.org/10.1007/978-90-481-3170-9

Campion, J. M. (2005). *Climatic and Nutritional Controls on the Growth of Eucalyptus Grandis in South Africa*. University of the Witwatersrand, Johannesburg.

Campion, J. M., Esprey, L. J., & Scholes, M. C. (2005). Application of the 3-pg model to a eucalyptus grandis stand subjected to varying levels of water and nutritional constraints in kwazulu-natal, south africa. *Southern African Forestry Journal*, *203*(1), 3–13. https://doi.org/10.2989/10295920509505213

Chauke, M. (2018). Stand Height Growth Model Conditioned to Changes in Rainfall for Eucalyptus Pulpwood in Mondi, South Africa. *New Frontiers in Forecasting Forests, Stellenbosch South Africa, 25-28/9/2018*, 18–20.

Chen, F.-W., & Liu, C.-W. (2012). Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy and Water Environment*, *10*(3), 209–222. https://doi.org/10.1007/s10333-012-0319-1

Chen, S., & Guo, J. (2017). Spatial interpolation techniques: their applications in regionalizing climate-change series and associated accuracy evaluation in Northeast China. *Geomatics, Natural Hazards and Risk*, *8*(2), 689–705. https://doi.org/10.1080/19475705.2016.1255669

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., & Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *CATENA*, *151*, 147–160. https://doi.org/10.1016/J.CATENA.2016.11.032

Clutter, J. L., & Jones, E. P. (1980). *Prediction of Growth After Thinning in Old-field Slash Pine Plantations*. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station. https://books.google.co.za/books?id=nqPcjuYOKQkC

Coops, N. C., Hember, R. A., & Waring, R. H. (2010). Assessing the impact of current and projected climates on Douglas-Fir productivity in British Columbia, Canada, using a process-based model (3-PG). *Canadian Journal of Forest Research*, *40*(3), 511–524. https://doi.org/10.1139/x09-201

Crous, C. J., Greyling, I., & Wingfield, M. J. (2018). Dissimilar stem and leaf hydraulic traits suggest varying drought tolerance among co-occurring Eucalyptus grandis × E. urophylla clones. *Https://Doi.Org/10.2989/20702620.2017.1315546*, *80*(2), 175–184. https://doi.org/10.2989/20702620.2017.1315546

DAFF (DEPARTMENT OF AGRICULTURE, F. A. F. (2019). *Report on commercial timber resources and primary roundwood processing in South Africa* (Vol. 18th). https://doi.org/10.1163/9789004272224_029

Dawson, C. W., Abrahart, R. J., & See, L. M. (2007). HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling and Software*, *22*(7), 1034–1052. https://doi.org/10.1016/j.envsoft.2006.06.008

De Cáceres, M., Martin-StPaul, N., Turco, M., Cabon, A., & Granda, V. (2018). Estimating daily meteorological data and downscaling climate models over landscapes. *Environmental Modelling and Software*, *108*, 186–196. https://doi.org/10.1016/j.envsoft.2018.08.003

Drew, T. J., & Flewelling, J. W. (1977). Some Recent Japanese Theories of Yield-Density Relationships and their Application to Monterey Pine Plantations. *Forest Science*, *23*(4), 517–534. https://doi.org/10.1093/FORESTSCIENCE/23.4.517

du Plessis, M., & Zwolinski, J. (2003). Site and Stand analysis for growth prediction of Eucalyptus grandis on the Zululand Coastal plain. *Southern African Forestry Journal*, *198*, 23–33.

du Toit, B., Smith, C. W., Little, K. M., Boreham, G., & Pallett, R. N. (2010). Intensive, site-specific silviculture: Manipulating resource availability at establishment for improved stand productivity. A review of South African research. *Forest Ecology and Management*, *259*(9), 1836–1845. https://doi.org/10.1016/j.foreco.2009.07.015

Dye, P. J., Jacobs, S., & Drew, D. (2004). Verification of 3-PG growth and water-use predictions in twelve Eucalyptus plantation stands in Zululand, South Africa. *Forest Ecology and Management*, *193*(1–2), 197–218. https://doi.org/10.1016/j.foreco.2004.01.030

Dye, P., Megown, R., Jacobs, S., Drew, D., Megown, K., Dicks, M., Mthembu, S., & Pretorius, C. (2002). *Determining the water use and growth of forest plantations through GIS-based integration of remote sensing and field data in the 3-PG model* (Issue 1194/1/02).

Dye, Peter J. (2001). Modelling growth and water use in four pinus patula stands with the 3-pg model. *Southern African Forestry Journal*, *191*(1), 53–63. https://doi.org/10.1080/20702620.2001.10434151

Dye, Peter J. (2000). Water use efficiency in South African Eucalyptus plantations: A review. *Southern African Forestry Journal*, *189*(1), 17–26. https://doi.org/10.1080/10295925.2000.9631276

Dyer, C. (2007). Forestry faces big issues to remain sustainable - A role for forestry research. *Southern Hemisphere Forestry Journal*, *69*(1). https://doi.org/10.2989/19919310709505185

Eccel, E., Cordano, E., & Toller, G. (2016). *ClimClass: Climate Classification According to Several Indices. R package version 2.1.0.* Https://CRAN.R-Project.Org/Package=ClimClass. http://www.globalbioclimatics.org/

Edwards, M. B. P., & Roberts, P. J. T. (2006). Managing forests for water: the South African experience. In *Source: The International Forestry Review* (Vol. 8, Issue 1). https://about.jstor.org/terms

Elli, E. F. (2020). *Eucalyptus simulation models : understanding and mitigating the impacts of climate variability and change on forest productivity across Brazil* (Issue August) [University of Sao Paulo]. https://doi.org/10.11606/T.11.2020.tde-13082020-180005

Esprey, L. J. (2006). Assessment of a Process-Based Model to Predict the Growth and Yield of Eucalyptus grandis Plantations in South Africa. *ICFR Bulletin Series*, *05/2006*, 1–221.

Esprey, L. J., Sands, P. J., & Smith, C. W. (2004). Understanding 3-PG using a sensitivity analysis. *Forest Ecology and Management*, *193*(1–2), 235–250. https://doi.org/10.1016/j.foreco.2004.01.032

Forestry | South African Government. (n.d.). Retrieved September 8, 2021, from https://www.gov.za/about-sa/forestry

Forrester, D. I., Hobi, M. L., Mathys, A. S., Stadelmann, G., & Trotsiuk, V. (2021). Calibration of the process-based model 3-PG for major central European tree species. *European Journal of Forest Research*, *3*, 1–22. https://doi.org/10.1007/S10342-021-01370-3

Forrester, D. I., & Tang, X. (2016). Analysing the spatial and temporal dynamics of species interactions in mixed-species forests and the effects of stand density using the 3-PG model. *Ecological Modelling*, *319*, 233–254. https://doi.org/10.1016/j.ecolmodel.2015.07.010

Gardner, R. A. W. (2012). Alternative eucalypt species for Zululand: Seven year results of site: species interaction trials in the region. *The Southern African Forestry Journal*, *190*(1), 79–88. https://doi.org/10.1080/20702620.2001.10434119

Genuer, R., & Poggi, J.-M. (2020). *Random Forests with R* (R. Gentleman, K. Hornik, & G. Parmigiani (Eds.); 1st ed.). Springer. https://doi.org/10.1007/978-3-030-56485-8

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014

Germishuizen, I. (2018). Mapping Risk at Different Spatial and Temporal Scales for Short-and-

Long Term Risk Evaluation: The Case of the Eucalypt Gall Wasp Leptocybe Invasa. *New Frontiers in Forecasting Forests, Stellenbosch South Africa, 25-28/9/2018*, 74–76.

Gibberd, V., Rook, J., Sear, C. B., & Williams, J. B. (1996). *Drought Risk Management In Southern Africa: The Potential of Long Lead Climate Forecasts for Improved Drought Management.* https://core.ac.uk/download/pdf/42390419.pdf

GISCOE. (2001). *GISCOE 20m GISCOE DTM Data*. GISCOE Pty Ltd.

Gonzalez-Benecke, C. A., Jokela, E. J., Cropper, W. P., Bracho, R., & Leduc, D. J. (2014). Parameterization of the 3-PG model for Pinus elliottii stands using alternative methods to estimate fertility rating, biomass partitioning and canopy closure. *Forest Ecology and Management*, *327*, 55–75. https://doi.org/10.1016/j.foreco.2014.04.030

Gonzalez-Benecke, C. A., Teskey, R. O., Martin, T. A., Jokela, E. J., Fox, T. R., Kane, M. B., & Noormets, A. (2016). Regional validation and improved parameterization of the 3-PG model for Pinus taeda stands. *Forest Ecology and Management*, *361*, 237–256. https://doi.org/10.1016/j.foreco.2015.11.025

Goovaerts, P. (1999). Performance comparison of geostatistical algorithms for incorporating elevation into the mapping of precipitation. *IV International Conference on GeoComputation, Mary Washington College, Fredericksburg*, 1–16.

Goulding, C. J. (1994). Development of growth models for Pinus radiata in New Zealand - experience with management and process models. *Forest Ecology and Management*, *69*(1–3), 331–343. https://doi.org/10.1016/0378-1127(94)90239-9

Gupta, R., & Sharma, L. K. (2019). The process-based forest growth model 3-PG for use in forest management: A review. *Ecological Modelling*, *397*(December 2018), 55–73. https://doi.org/10.1016/j.ecolmodel.2019.01.007

Hijmans, R., & Parra, J. L. (2005). Very high resolution interpolated climate surfaces of global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/joc.1276

Home | Sappi Global. (n.d.). Retrieved September 8, 2021, from https://www.sappi.com/

Hu, Q., Li, Z., Wang, L., Huang, Y., Wang, Y., & Li, L. (2019). Rainfall spatial estimations: A review from spatial interpolation to multi-source data merging. *Water (Switzerland)*, *11*(3), 1–30. https://doi.org/10.3390/w11030579

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Leraning: with Applications in R* (G. Casella, S. Fienberg, & I. Olkin (Eds.); 1st ed.). Springer.

Jeffrey, S. J., Carter, J. O., Moodie, K. B., & Beswick, A. R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling & Software*, *16*(4), 309–330. https://doi.org/10.1016/S1364-8152(01)00008-1

Johnsen, K., Samuelson, L., Teskey, R., Mcnulty, S., & Fox, T. (2001). Process Models as Tools in Forestry Research and Management. *Forest Science*, *47*(1), 2–8.

Karavani, A., De Cáceres, M., Martínez de Aragón, J., Bonet, J. A., & de-Miguel, S. (2018). Effect of climatic and soil moisture conditions on mushroom productivity and related ecosystem services in Mediterranean pine stands facing climate change. *Agricultural and Forest Meteorology*, *248*, 432–440. https://doi.org/10.1016/J.AGRFORMET.2017.10.024

Kassier, H. W. (1993). *Dynamics of Diameter and Height Distributions in Even-Aged Pine Plantations. Unpublished Ph.D. dissertation.* Stellenbosch University.

Kassier, H. W. (2005). *Angle count sampling as an alternative inventory method. Unpublished document.* (p. 41).

Kimmins, J. P. (1990). Modelling the sustainability of forest production and yield for a changing and uncertain future. *Forestry Chronicle*, *66*(3), 271–280. https://doi.org/10.5558/tfc66271-3

Korzukhin, M. D., Ter-Mikaelian, M. T., & Wagner, R. G. (1996). Process versus empirical models: Which approach for forest ecosystem management? *Canadian Journal of Forest Research*, *26*(5), 879–887. https://doi.org/10.1139/x26-096

Kotze, H. (2018). Forecasting with empirical stand-level growth and yield models and drought modifiers for short rotation Eucalyptus Pulpwood in Mondi, South Africa. *New Frontiers in Forecasting Forests, Stellenbosch South Africa, 25-28/9/2018*, 165.

Kotze, H., & Fletcher, Y. (2013). *The 2013 Growth and Yield Model for Egxu. Unpublished Report.*

Landsberg, J. (2003a). Modelling forest ecosystems: State of the art, challenges, and future directions. *Canadian Journal of Forest Research*, *33*(3), 385–397. https://doi.org/10.1139/x02-129

Landsberg, J. (2003b). Physiology in forest models: history and the future. *Fbmis*, *1*(January 2003), 49–63.

Landsberg, J. J., & Gower, S. T. (1997). *Applications of Physiological Ecology to Forest Management.* Academic Press, Elsevier. https://doi.org/10.2307/2405244

Landsberg, J. J., & Waring, R. H. (1997). A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest Ecology and Management*, *95*(3), 209–228. https://doi.org/10.1016/S0378-1127(97)00026-1

Landsberg, J. J., Waring, R. H., & Coops, N. C. (2003). Performance of the forest productivity model 3-PG applied to a wide range of forest types. *Forest Ecology and Management*, *172*(2–3), 199–214. https://doi.org/10.1016/S0378-1127(01)00804-0

Landsberg, J. J, & Sands, P. (2011). Physiological Ecology of Forest Production: Principles, Processes and Models. In R. E. James, M. James, & G. T. Monica (Eds.), *Terrestrial Ecology Series* (First edit, Vol. 4, p. 352). Academic Press, Elsevier. https://doi.org/10.1016/B978-0-12-374460-9.00016-0

Landsberg, Joe J., & Coops, N. C. (1999). MODELING FOREST PRODUCTIVITY ACROSS LARGE AREAS AND LONG PERIODS. *Natural Resources Modelling*, *12*(4), 383–411.

Landsberg, Joe J, Johnsen, K. H., Albaugh, T. J., & Allen, H Lee and Mckean, S. E. (2001). Applying 3-PG, a simple process-based model designed to produce practical results from Loblolly Pine experiments. *Forest Science*, *47*(1), 43–51.

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241. https://doi.org/10.1029/1998WR900018

Leirvik, T., & Yuan, M. (2021). A Machine Learning Technique for Spatial Interpolation of Solar Radiation Observations. *Earth and Space Science*, *8*(4), 1–22. https://doi.org/10.1029/2020EA001527

Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, *53*, 173–189. https://doi.org/10.1016/j.envsoft.2013.12.008

Li, J., Heap, A. D., Potter, A., & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, *26*(12), 1647–1659. https://doi.org/10.1016/J.ENVSOFT.2011.07.004

Louw, J. H., Germishuizen, I., & Smith, C. W. (2011). A stratification of the South African forestry landscape based on climatic parameters. *Southern Forests*, *73*(1), 51–62. https://doi.org/10.2989/20702620.2011.574825

Lynch, S. D. (2004). *Development of a Raster Database of Annual, Monthly and Daily Rainfall for Southern Africa*.

Mäkelä, A., Landsberg, J., Ek, A. R., Burk, T. E., Ter-Mikaelian, M., Ågren, G. I., Oliver, C. D., & Puttonen, P. (2000a). Process-based models for forest ecosystem management: Current state of the art and challenges for practical implementation. *Tree Physiology*, *20*(5–6), 289–298. https://doi.org/10.1093/treephys/20.5-6.289

Mäkelä, A., Landsberg, J., Ek, A. R., Burk, T. E., Ter-Mikaelian, M., Ågren, G. I., Oliver, C. D., & Puttonen, P. (2000b). Process-based models for forest ecosystem management: Current state of the art and challenges for practical implementation. *Tree Physiology*, *20*(5–6), 289–298. https://doi.org/10.1093/treephys/20.5-6.289

McGrath, J. M., Spargo, J., & Penn, C. J. (2014). Soil Fertility and Plant Nutrition. *Encyclopedia of Agriculture and Food Systems*, *5*, 166–184. https://doi.org/10.1016/B978-0-444-52512-3.00249-7

Melesse, S. F., & Zewotir, T. (2015). Fitting three parameter growth curves using a nonlinear mixed effects modelling approach. *South African Statistical Journal*, *49*(2), 233–240.

Melesse, S. F., & Zewotir, T. (2017). Variation in growth potential between hybrid clones of Eucalyptus trees in eastern South Africa. *Journal of Forestry Research*, *28*(6), 1157–1167. https://doi.org/10.1007/s11676-017-0400-0

Miehle, P., Battaglia, M., Sands, P. J., Forrester, D. I., Feikema, P. M., Livesley, S. J., Morris, J. D., & Arndt, S. K. (2009). A comparison of four process-based models and a statistical regression model to predict growth of Eucalyptus globulus plantations. *Ecological Modelling*, *220*(5), 734–746. https://doi.org/10.1016/j.ecolmodel.2008.12.010

Mondi Group - Global leader in packaging and paper. (n.d.). Retrieved September 8, 2021, from https://www.mondigroup.com/en/home/

Monteith, J. L. (1965). Evaporation and environment. *Symposia of the Society for Experimental Biology*, *19*, 205–234.

Moriasi, D. N., Arnold, J. G., Van Liew, M. ., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2008). Model Evaluation Guildlines For Systematic Quantification of Accuracy in Watershed Simulations. *Colombia Medica*, *39*(3), 227–234. https://doi.org/10.1234/590

Naidoo, S., Davis, C., & Archer Van Garderen, E. (2013). *Forests, Rangelands and Climate Change in Southern Africa* (Forests and Climate Change Working Paper No. 12.). www.fao.org/publications

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Ndlovu, M., Clulow, A. D., Savage, M. J., Nhamo, L., Magidi, J., & Mabhaudhi, T. (2021). An assessment of the impacts of climate variability and change in Kwazulu-Natal province, South Africa. *Atmosphere*, *12*(4). https://doi.org/10.3390/ATMOS12040427

Ndlovu, M. S., & Demlie, M. (2020). Assessment of Meteorological Drought and Wet Conditions Using Two Drought Indices Across KwaZulu-Natal Province, South Africa. *Atmosphere 2020, Vol. 11, Page 623*, *11*(6), 623. https://doi.org/10.3390/ATMOS11060623

Nolè, A., Law, B. E., Magnani, F., Matteucci, G., Ferrara, A., Ripullone, F., & Borghetti, M. (2009). Application of the 3-PGS model to assess carbon accumulation in forest ecosystems at a regional level. *Canadian Journal of Forest Research*, *39*(9), 1647–1661. https://doi.org/10.1139/X09-077

Olivier, F. (2017). *Irrigation: Basics of Irrigation Scheduling* (No. 5; Information Sheets). www.sugar.org.za/sasri

Oltean, G. S., Comeau, P. G., & White, B. (2016). Carbon isotope discrimination by Picea glauca and Populus tremuloides is related to the topographic depth to water index and rainfall. *Canadian Journal of Forest Research*, *46*(10), 1225–1233. https://doi.org/10.1139/cjfr-2015-0491

Palahí, M., Tomé, M., Pukkala, T., Trasobares, A., & Montero, G. (2004). Site index model for Pinus sylvestris in north-east Spain. *Forest Ecology and Management*, *187*(1), 35–47. https://doi.org/10.1016/S0378-1127(03)00312-8

Paul, K. I., Polglase, P. J., O'Connell, A. M., Carlyle, J. C., Smethurst, P. J., & Khanna, P. K. (2002). Soil nitrogen availability predictor (SNAP): a simple model for predicting mineralisation of nitrogen in forest soils. *Australian Journal of Soil Research*, *40*(6), 1011–1026. https://doi.org/10.1071/SR01114

Peng, C. (2000). Growth and yield models for uneven-aged stands: Past, present and future. *Forest Ecology and Management*, *132*(2–3), 259–279. https://doi.org/10.1016/S0378-1127(99)00229-7

Peng, C., & Wen, X. (2006). Forest Simulation Models. In G. Shao & K. M. Reynolds (Eds.), *Computer Applications in Sustainable Forest Managemnet: Including Perspectives on Collaboration and Integration* (First Edit, pp. 101–125). Springer, Dordrecht. https://doi.org/https://doi.org/10.1007/978-1-4020-4387-1_6

Pienaar, L. V, & Harrison, W. M. (1989). Simultaneous Growth and Yield Prediction Equations for Pinusï¿½elliottii Plantations in Zululand. *South African Forestry Journal*, *149*(1), 48–53. https://doi.org/10.1080/00382167.1989.9628992

Pinkard, E. A., Battaglia, M., Bruce, J., Leriche, A., & Kriticos, D. J. (2010). Process-based modelling of the severity and impact of foliar pest attack on eucalypt plantation productivity under current and future climates. *Forest Ecology and Management*, *259*(4), 839–847. https://doi.org/10.1016/j.foreco.2009.06.027

Pretzsch, H. (2009). Forest Dynamics, Growth and Yield: From Measurement to Model. In *Springer Nature* (1st ed.). Springer-Verlag Berlin Heidelberg.

Retief, E., & Stanger, T. (2009). Genetic parameters of pure and hybrid populations of Eucalyptus grandis and E. urophylla and implications for hybrid breeding strategy. *Southern Forests: A Journal of Forest Science*, *71*(2), 133–140.

https://doi.org/10.2989/SF.2009.71.2.8.823

Rezende, G. D. S. P., de Resende, M. D. V., & de Assis, T. F. (2014). Eucalyptus Breeding for Clonal Forestry. In F. T (Ed.), *Challenges and Opportunities for the World's Forests in the 21st Century. Forest Sciences* (Vol. 81). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-7076-8_16

Sánchez-Pinillos, M., Ameztegui, A., Kitzberger, T., & Coll, L. (2018). Relative size to resprouters determines post-fire recruitment of non-serotinous pines. *Forest Ecology and Management*, *429*, 300–307. https://doi.org/10.1016/J.FORECO.2018.07.009

Sands, P. (1988). RESOURCE MODELLING: ITS NATURE AND USE. *The Memoirs of the Entomological Society of Canada*, *120*(S143), 5–10. https://doi.org/10.4039/ENTM120143005-1

Sands, P. (2004). *Adaptation of 3-PG to novel species : guidelines for data collection and parameter assignment* (Issue Technical Report 141). https://3pg.sites.olt.ubc.ca/files/2014/04/3-PG-guidelines.TR141.pdf

Sands, P. J., & Landsberg, J. J. (2002). Parameterisation of 3-PG for plantation grown Eucalyptus globulus. *Forest Ecology and Management*, *163*(1–3), 273–292. https://doi.org/10.1016/S0378-1127(01)00586-2

Sands, Peter J. (2010). *3PG PJS User Manual*. *September*, 1–27.

Saxton, K. E., & Rawls, W. J. (2006). Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal*, *70*(5), 1569–1578. https://doi.org/10.2136/sssaj2005.0117

Schulze, R. E. (2007). Soils: Agrohydrological Information Needs, Information Sources and Decision Support. In R. E. Schulze (Ed.), *South African Atlas of Climatology and Agrohydrology* (WRC Report, pp. 1–13). Water Research Commision.

Scott, D. F., & Gush, M. B. (2017). Forest management and water in the Republic of South Africa. In P. A. Garcia-Chevesich, D. G. Neary, D. F. Scott, R. G. Benyon, & T. Reyna (Eds.), *Forest Management and the impact on water resources: a review of 13 countries* (p. 17). United Nations Educational, Scientific and Cultural Organization and UNESCO Regional Office for Sciences for Latin America and the Caribbean. http://www.unesco.org/

Sekulić, A., Kilibarda, M., Heuvelink, G. B. M., Nikolić, M., & Bajat, B. (2020). Random forest spatial interpolation. *Remote Sensing*, *12*(10), 1–29. https://doi.org/10.3390/RS12101687

Sekulić, A., Kilibarda, M., Protić, D., & Bajat, B. (2021). A high-resolution daily gridded meteorological dataset for Serbia made by Random Forest Spatial Interpolation. *Scientific Data*, *8*(1), 1–12. https://doi.org/10.1038/s41597-021-00901-2

Singh, J., Knapp, H. V., & Demissie, M. (2004). *Hydrologic Modeling of the Iroquois River Watershed Using HSPF and SWAT. Illlinois State Water Survey Contract Report 2004-08*.

Soil Classification Working Group. (1991). Soil classification : a taxonomic system for South Africa. In *Memoirs on the agricultural natural resources of South Africa ; no. 15* (2nd rev. e). Dept. of Agricultural Development. https://sun.on.worldcat.org/oclc/38349167

Soil water availability · Issue #51 · trotsiuk/r3PG. (n.d.). Retrieved September 20, 2021, from https://github.com/trotsiuk/r3PG/issues/51

Stadler, S. J. (2005). Aridity Indexes. In J. E. Oliver (Ed.), *Encyclopedia of world climatology* (pp. 89–94). Springer. https://doi.org/10.5860/choice.43-3766

Stanger, T. K., Galloway, G. M., & Retief, E. C. L. (2011). Final results from a trial to test the effect of plot size on Eucalyptus hybrid clonal ranking in coastal Zululand, South Africa. *Southern Forests*, *73*(3–4), 131–135. https://doi.org/10.2989/20702620.2011.639492

Stape, J. L., Ryan, M. G., & Binkley, D. (2004). Testing the utility of the 3-PG model for growth of Eucalyptus grandis x urophylla with natural and manipulated supplies of water and nutrients. *Forest Ecology and Management*, *193*(1–2), 219–234. https://doi.org/10.1016/j.foreco.2004.01.031

Subedi, S., Fox, T. R., & Wynne, R. H. (2015). Determination of fertility rating (FR) in the 3-PG model for loblolly pine plantations in the southeastern United States based on site index. *Forests*, *6*(9), 3002–3027. https://doi.org/10.3390/f6093002

Tesfamichael, S. G., Ahmed, F. B., & Aardt, J. A. N. Van. (2010). Investigating the impact of discrete-return lidar point density on estimations of mean and dominant plot-level tree height in Eucalyptus grandis plantations. *Http://Dx.Doi.Org/10.1080/01431160903144086*, *31*(11), 2925–2940. https://doi.org/10.1080/01431160903144086

Thornthwaite, C. W. (1948). An Approach toward a Rational Classification of Climate. *Geographical Review*, *38*(1), 55–94. https://doi.org/10.2307/210739

Trotsiuk, V., Hartig, F., & Forrester, D. I. (2020). r3PG – An r package for simulating forest growth using the 3-PG process-based model . *Methods in Ecology and Evolution*, *2020*(August), 1–6. https://doi.org/10.1111/2041-210x.13474

Van Den Berg, G. J. (2017). *A Comparative Study of Two Eucalyptus Hybrid Breeding Strategies and the Genetic Gains of these Strategies* (Issue August). University of Pretoria.

van Niekerk, A., & Joubert, S. J. (2011). Input variable selection for interpolating high-resolution climate surfaces for the Western Cape. *Water SA*, *37*(3), 271–279. http://www.wrc.org.za

Vega-Nieva, D. J., Tomé, M., Tomé, J., Fontes, L., Soares, P., Ortiz, L., Basurco, F., & Rodrígez-Soalleiro, R. (2013). Developing a general method for the estimation of the fertility rating parameter of the 3-PG model: Application in Eucalyptus globulus plantations in northwestern Spain. *Canadian Journal of Forest Research*, *43*(7), 627–636. https://doi.org/10.1139/cjfr-2012-0491

Warburton, M. L., & Schulze, R. E. (2008). Potential impacts of climate change on the climatically suitable growth areas of Pinus and Eucalyptus: Results from a sensitivity study in South Africa. *Southern Forests*, *70*(1), 27–36. https://doi.org/10.2989/SOUTH.FOR.2008.70.1.5.515

Warburton, M., & Schulze, R. (2006). *CLIMATE CHANGE AND THE SOUTH AFRICAN COMMERCIAL FORESTRY SECTOR: An Initial Study. ACRUcons Report 54.*

Waring, R. H., Landsberg, J. J., & Williams, M. (1998). Net primary production of forests: a constant fraction of gross primary production? *Tree Physiology*, *18*(2), 129–134. https://doi.org/10.1093/TREEPHYS/18.2.129

Weiskittel, A. R., Hann, D. W., Kershaw, J. A., & Vanclay, J. K. (2011). Forest Growth and Yield Modeling. In *Forest Growth and Yield Modeling* (1st ed.). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119998518

Willmott, C. J. (1981). ON THE VALIDATION OF MODELS. *Physical Geography*, *2*(2), 184–194. https://doi.org/10.1080/02723646.1981.10642213

Xenakis, G., Ray, D., & Mencuccini, M. (2008). Sensitivity and uncertainty analysis from a coupled 3-PG and soil organic matter decomposition model. *Ecological Modelling*, *219*(1–2), 1–16. https://doi.org/10.1016/j.ecolmodel.2008.07.020

Xulu, S., Peerbhay, K., Gebreslasie, M., & Ismail, R. (2018). Drought influence on forest plantations in Zululand, South Africa, using MODIS time series and climate data. *Forests*, *9*(9). https://doi.org/10.3390/f9090528

Youssef, A. M., Pourghasemi, H. R., Pourtaghi, Z. S., & Al-Katheeri, M. M. (2015). Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides 2015 13:5*, *13*(5), 839–856. https://doi.org/10.1007/S10346-015-0614-1