Mapping *Begonia* Hotspots on Borneo

A. A. Piyali Dias

August 2019

Thesis submitted in partial fulfilment for the MSc in
Biodiversity of Taxonomy of Plants
University of Edinburgh

Royal
Botanic Garden
Edinburgh

Declaration

I declare that this thesis composed of work carried out by myself except where specifically acknowledged. It has not been previously accepted for any degree and is not being currently considered for any other degree at this or any other university.

Abstract

Borneo, a major evolutionary hotspot for southeast Asian biodiversity, is also a hotspot for the pantropical mega-diverse genus *Begonia* and currently harbours 254 species at nearly 100% endemism. Different collection densities make it difficult to identify *Begonia* hotspots in Borneo. Sarawak has the highest *Begonia* richness (149 species) while Kalimantan has the lowest (24) though it occupies the largest land area of the island. To identify true *Begonia* hotspots in Borneo, we modelled *Begonia* richness based on species occurrence data from herbarium specimens using MaxEnt and Single Point Extrapolation (SPE) approaches and produced a final *Begonia* richness map through stacking 154 individual distribution models. The results revealed that *Begonia* diversity hotspots are distributed in all regions in Borneo; Sabah is home to the richest *Begonia* hotspots of Mt. Kinabalu and the Tawau hills. Some of the predicted *Begonia* hotspots are located in limestone rich habitats. However, we found that limestone is not the only factor for begonia distribution, but also warm, humid environments, nutrient rich volcanic soils and a variety of geological formations also strongly influence the distribution and diversity of begonias in Borneo. MaxEnt itself couldn’t identify all *Begonia* hotspots, and this gap was overcome by developing the SPE approach demonstrating that a Maxent-SPE combined approach is a promising method to model species distributions for narrow endemics.

Acknowledgement

My supervisor, Mark Hughes deserves a big thank, for writing the initial project proposal and giving me this wonderful opportunity to embrace the field of biogeography. Without his enthusiastically sharing expertise, understanding, never-ending patience, reading through this entire thesis and guiding me for necessary corrections, none of these would have been possible. A big thank also to my subsidiary supervisor Antje Ahrends, for her outstanding expertise, valuable time and for the friendly support and guidance extended throughout this project and expanding my horizon on “GIS”.

This project would have not been possible without the hard works of the people who have been collected *Begonia* specimens throughout a long period of time. I’m indebted to all who contributed to *Begonia* database at Royal Botanical Garden Edinburgh and specially, Martin Pullan for updating “Padme” into a user-friendly interface within a short period of time.

Assistance during project time came up from numerous sources including Niels Raes for providing external gazetteers, Peter Moonlight for helping me in “R”, all ICT staff for solving IT issues and providing me access to GIS database and all the library staff for being very helpful in finding material.

The project would have not been easy without continuous encouragements of especially, Isuru Kariyawasam, Yu-Hsin Tseng, Zhifang Liu and all my MSc batchmates (Yutong, Phillipa, Mayra, Sarah, Rebecca, Prima, Pim, Matthew and Pietro) who have become good friends throughout this process and created a wonderful international blend among us.

My study in UK would not have been possible without the financial assistance of ‘Chevening Scholarship, the UK government’s global scholarship programme, funded by the Foreign and Commonwealth Office (FCO) and partner organisations’ to whom I’m very much grateful.

Thanks also, of course to my family who were always behind me in all ups and downs!

Table of Contents

List of Tables and Figures

CHAPTER 1: INTRODUCTION

1.1 The island of Borneo

The region at the junction of where the Australian, Eurasian, Indian, Pacific and Philippine plates interact holds a certain fascination and a great historical significance, being where Alfred Russel Wallace developed his greatest insights into the geography of life. The area is a composite domain of smaller micro-continental fragments and comprises both mainland and numerous archipelagos of oceanic and continental islands (Lohman *et al.*, 2011) which has resulted from more than 300 million years of continent-continent and continent-arc collisions (Bruyn *et al.*, 2014).

The island of Borneo is located in the midst of this extremely complex oceanic region in South East Asia (Figure 1), where three major plates are interacting and the volcanism has been common throughout the Tertiary (Wilson and Moss, 1999). It is a giant, rugged island, in third place among the largest islands in the world after Greenland and New Guinea, and is surrounded by Sumatra, Java, Sulawesi and Philippines. The island is located between 4 ^{0} to 7 ^{0} N latitudes and 108 ^{0} to 119 ^{0} E longitudes and shared with the Malaysian states of Sabah and

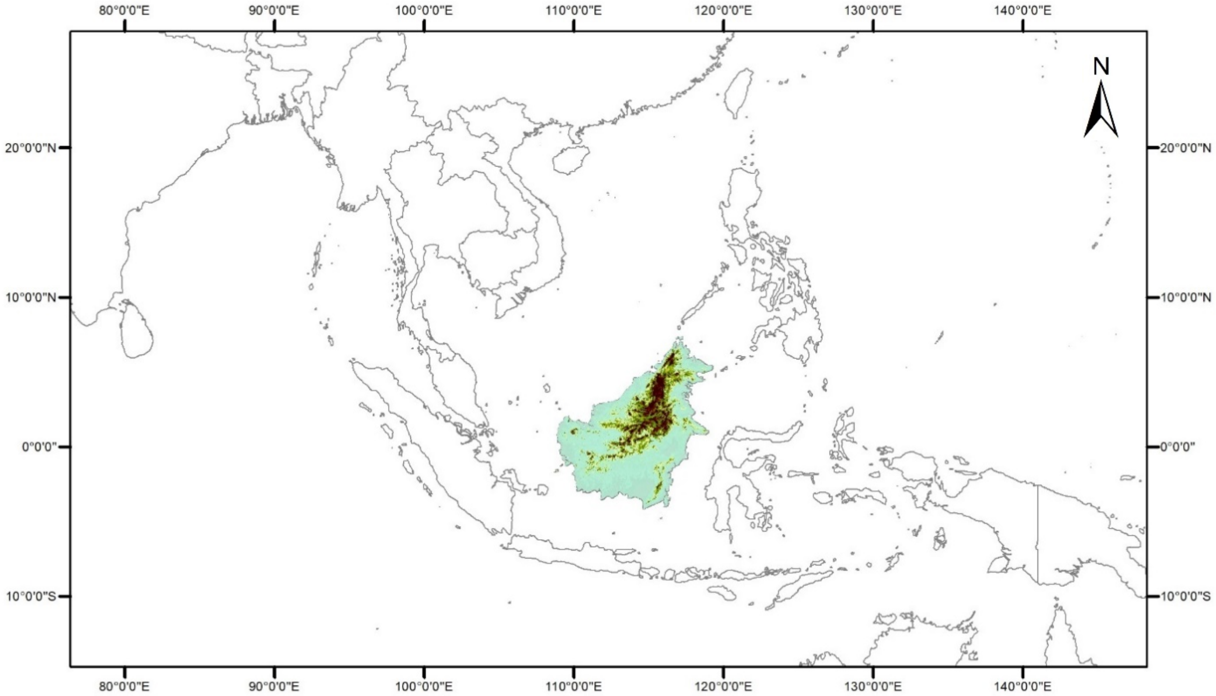

Figure 1: Location of Borneo island in the Indo-Australian archipelago.

Sarawak, Indonesian Kalimantan and the tiny little independent nation, Brunei Darussalam (Figure 2). Together these political states form one geographic unit and share a wealth of biological resources and tropical habitats rich in biodiversity.

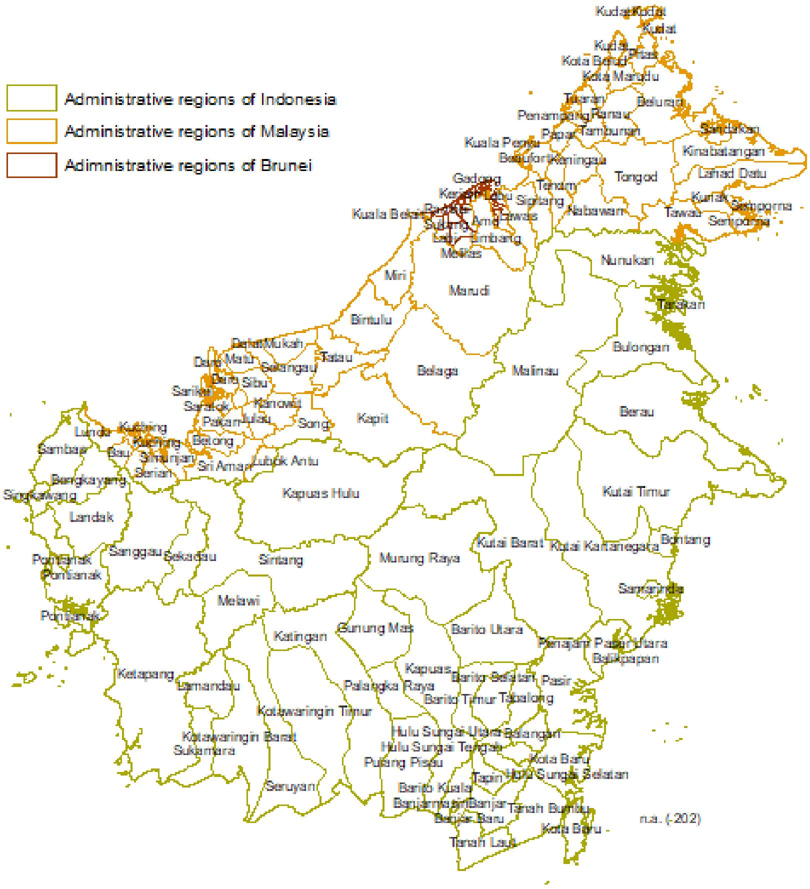

Figure 2: Administrative boundaries of Sabah, Sarawak, Kalimantan and Brunei of Borneo.

1.1.1 Origin

During the Pleistocene, the low sea levels increase the connectivity of terrestrial habitats across Sundaland and resulted in extensive islands and archipelagos (Beck *et al.*, 2017). As recently as 25 million years ago in the geological time scale, the Indonesian archipelago did not exist, as we know it today. The core of the Southeast Asia, Sundaland, was initially assembled from continental blocks that separated from Gondwanaland in the Palaeozoic and united with Asian blocks in Triassic (Hall, 2012a). By the end of Triassic (ca. 200 Ma), core Sundaland comprised Sumatra, Thai-Malay Peninsula and most of the present-day Sunda Shelf. It is generally accepted that three major collisions have been significantly influenced the current archipelago configuration and probably substantially affect both climate and assembly of regional biota (Hall, 2011; Metcalfe, 2013).

Major changes began in the Eocene (Hall, 2002; Lohman *et al.*, 2011). Importantly, the Asia-Australia collision (ca 23-25 Ma) had a major impact on the climate and vegetation of the region by disrupting the Indonesian throughflow. Specially, the Southeast Asian gateway, the connection from the Pacific to the Indian Ocean, diminished from a wide ocean to a narrow passage with a deep topographic barrier to biotic movements (Bruyn *et al.*, 2014). Further, this collision caused widespread volcanism at the active margin and produced chains of islands in Indonesian region (Lohman *et al.*, 2011).

Hamilton (1973), states that Borneo includes the largest of continental blocks with rocks exposed that are older than Mesozoic. He further emphasized the surrounding region, from Sarawak, Sabah, East Kalimantan and East Java to South Sumatra has been interpreted as Cretaceous and Tertiary subduction complexes and often reflected changing subduction patterns. It is often assumed that the Southwest Borneo continental core was attached to Sundaland before the Cretaceous, while Hall (2012b) further discussed Borneo is the first and the largest fragment rifted from Northwest Australia in the Late Jurassic at about 160 Ma. Further, Southern Kalimantan is supposed to be rifted from Australia in the early cretaceous and northern Borneo had probably rifted from east Asia even earlier and added to Sundaland (MacKinnon *et al.*, 1997; Hall, 2012a). Afterwards, in the early Miocene, as the Australian plate had made contact with submerged Sundaland, the Sundaland began to rotate counter clockwise hence, producing a “Deep Regional Unconformity” (Hall, 1995; Lohman *et al.*, 2011). Richter *et al.* (1999) emphasize that in mesozoic rocks in Kalimantan and Sarawak, a stronger rotation declination rotation of 90^0 counter-clockwise rotation has been recorded. At about the same time, mountains rose from central Borneo and the rivers shed large amount of sediments at onshore and offshore basins around Borneo which resulted from the generally wetter climate in Sundaland (Morley, 2000). Moreover, crustal melting over the past 7 million years caused the rise of Mount Kinabalu to its present height of more than 4000 m and widening the area of lower hills surrounding in Borneo (Lohman *et al.*, 2011) .

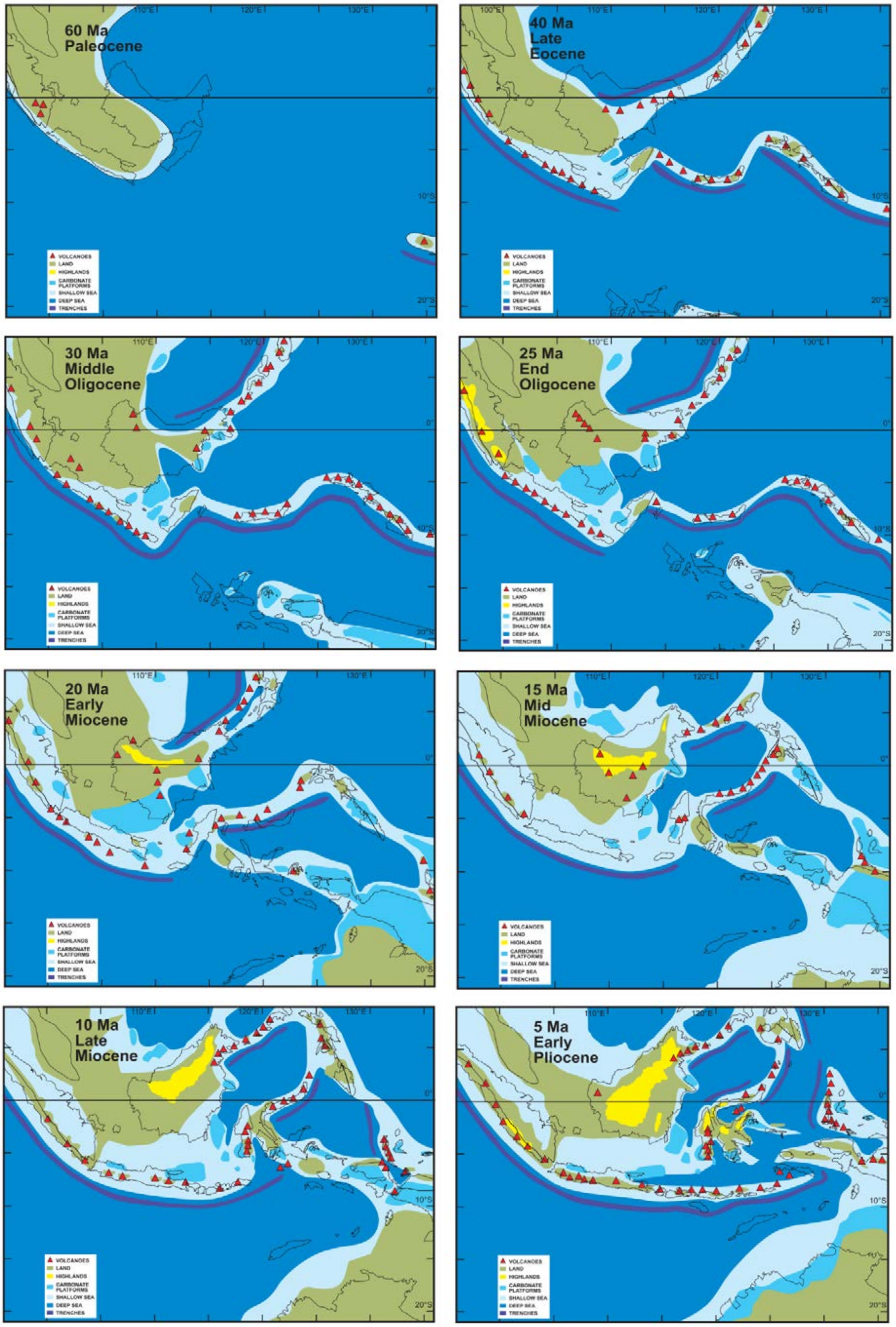

Figure 3: Postulated distribution of the land and sea in the Walacea region (Hall, 2009).

1.1.2 Physical geography of Brunei, Sarawak, Sabah and Kalimantan

Geology

Borneo consists of various geological formations originated in different time scales throughout millions of years. Northern parts of Sabah consisting of sedimentary and igneous rocks (Tongkul, 1994). The oldest rocks in Sabah are those of the Crystalline Basement Formation. The western half formations are mainly hard sandstone mixed with other sedimentary rocks, the area dominated by the Crocker Formation, geosynclinals flysch deposits of sandstone and shales (*CAIMS*, 2005). According to Hall *et al.* (2013), the oldest rocks from Sabah are high grade metamorphic rocks assigned to the crystalline basement and most of the exposed basement rocks are ophiolites and associated sediments. In the area close to Kinabalu ophiolites are represented by partly serpentinised peridotites. Sedimentary and ultrabasic rocks and forms the core of the Kinabalu massif. The core, granodiorite, is acid intrusive rock somewhat similar to granite which consist essentially of feldspar, hornblende and small amount of free quarts (Collenette, 1964). The youngest rocks in the Kinabalu area are terrestrial sediments, contain

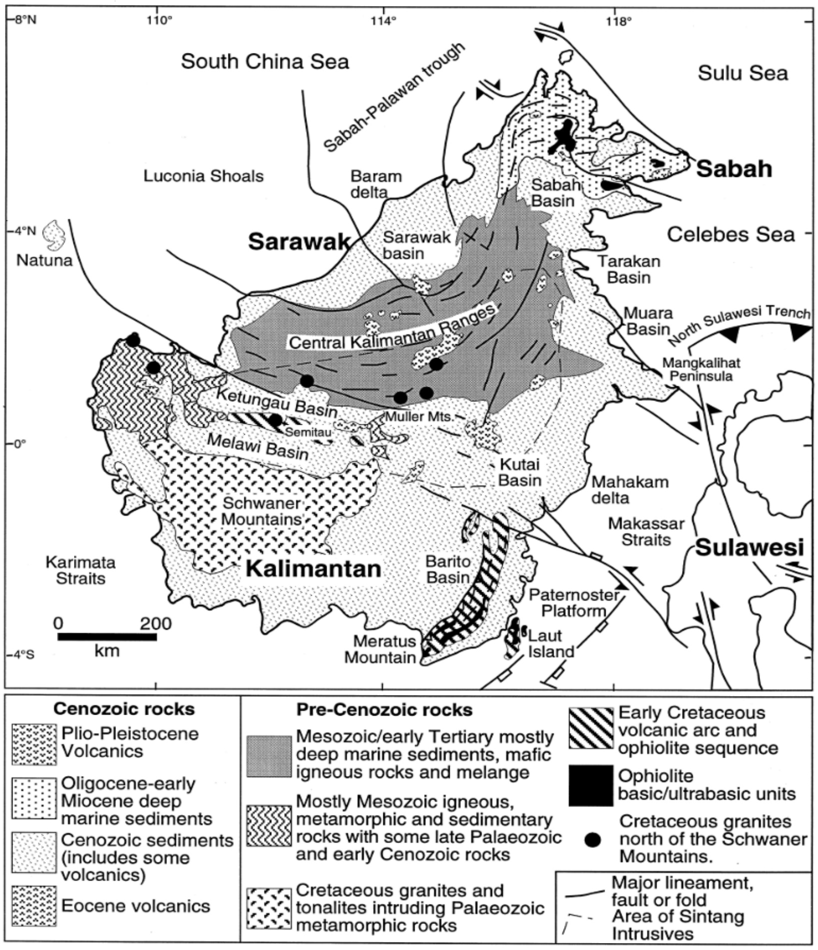

Figure 4: Generalized geological map of Borneo (Wilson and Moss, 1999).

a variety of rock types that include the granite and most of the older rocks surrounding Mt Kinabalu. The igneous rocks of Kinabalu have been described as quartz monzonite, adamellite, granodiorite and granite. Sarawak consists of a variety of geological formations such as limestones, sandstones, granites, mudstones, lignites, and siltstones.(Lin, Chung and Peng, 2014). Limestones occurs in upper Sarawak as the southwest extremity of a long line of limestone outcrops extending throughout the whole of Sarawak characterizing two well-marked types (Scrivenor, 1927). Further, in upper Sarawak most striking feature is shale, specially the mountain Matang, near Kuching, consists, in great part, of shale, sandstone, and a little conglomerate. The Schwaner mountains in south western, Meratus mountains in eastern and the central Kalimantan range from northeast to southwest comprise pre-tertiary rocks. The Schwner mountains intruded by numerous cretaceous granites and tonalites. To the north of the Schwaner Mountains and within the Central Kalimantan ranges are with smaller Cretaceous granites. Much of the Kalimantan consist of consolidated and semi-consolidated rocks including quaternary limestones and volcanic rocks. Large areas of Central, East and South Kalimantan are composed of sedimentary rocks such as sandstones and shales (MacKinnon *et al.*, 1997). Although Borneo has no active volcanos, it does have substantial old volcanic rocks in southwest and eastern Kalimantan. The geology of Brunei is primarily made out of Cenozoic rocks while the oldest rocks are part of the Meligan Formation and Temburong Formation in the east and are the deformed remnants of floodplain sediments.

Soils

Soils are a major factor in affecting vegetation distribution. Since the weathering is strong in the humid tropics because of high temperature and moisture, soils are constantly leaching. MacKinnon *et al.* (1997), mention that, the high levels of weathering, leaching and biological activity makes Bornean soils generally less fertile compared to rich volcanic soils in Java islands. Strongly weathered soils (ultisols) forms the high proportion of red-yellow podsolic soils in central and north-eastern Borneo. Inceptisols are the commonest soils found in Kalimantan which generally more fertile of the Bornean soils. Fluvents and the alluvial soils are generally found on larger river valleys, flood plains and on the coastal plains across the island. Histosols or the peats covers a vast area of the lowland while ultisols are the widely distributed ones on the hilly lands (Sakurai, 1999). Commonly, the mountain soils are changing with increasing altitudes and generally becoming more acidic and nutrient poor, which also characteristics of Kinabalu soils. Soils on limestone karsts are often richer in bases, especially in calcium and magnesium and higher cation-exchange capacity than other lowland soils MacKinnon *et al.* (1997).

Climate

Since Borneo lies on the equator, it has a moist, tropical climate all year around where main climatic variable is rainfall. Mean annual temperature ranges from 25 °C to 35 °C in low land areas whereas between 10 °C to 20 °C across highlands. Distribution of rainfall across the island is paramount importance in determining the type of vegetation in each region. Pattern of rainfall in Borneo is mainly determined by southeast and northwest monsoons while usually the northwest monsoon is wetter than southeast monsoon. Most of the hilly areas annually receive between 2000-4000mm rain. While the rainfall is relatively heavy in wet season, climate is relatively drier from June to August but not less than 100 mm rainfall in Kalimantan (MacKinnon *et al.*, 1997).

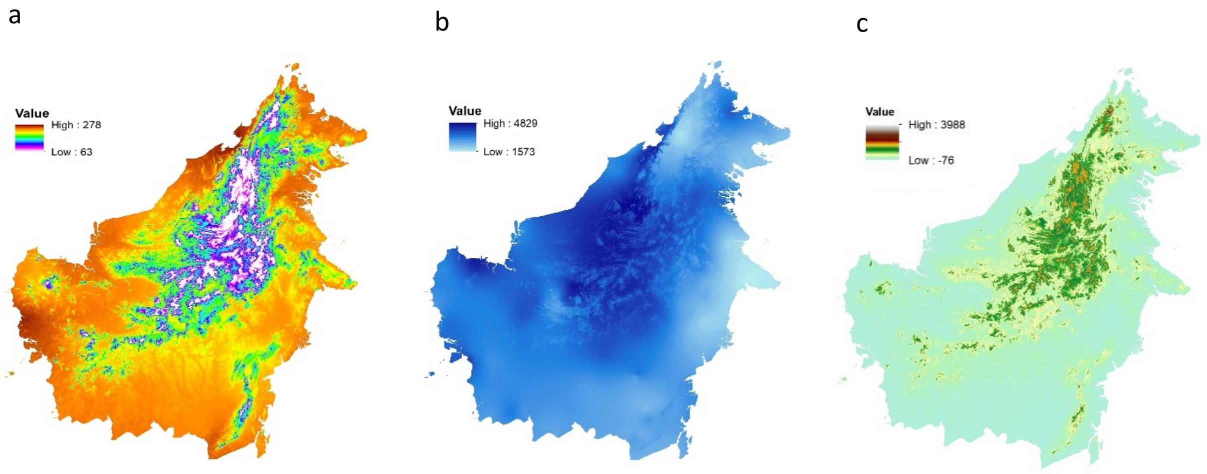

Figure 5: Variation of (a) Annual Mean Temperature, (b) Annual Mean Precipitation and (c) Altitude in Borneo (Fick and Hijmans, 2017).

Elevation and topography

The island is highly varying in elevation and topography from lowland to highlands up to ca. 4000 m high at the peak of Mt. Kinabalu in northern Borneo. The northern and central part of the island has the larger mountain ranges with varying topography such as; mountains, hills, valleys, plains, steep slopes, etc. the major mountain range running from southwest to northeast. In the south, elevations are low only rarely reaching 1200 m but further northern specially the central mountain range- the Indonesian-Malaysian boundary reaches even passing 2400 m of height where Kinabalu deserves the top (MacKinnon *et al.*, 1997).

1.1.3 Historical biogeography and biodiversity distribution

The historical assembly of the flora in the Malay Archipelago has been documented by numerous studies. Recent studies found that the Sundaland rainforest previously covered a substantially larger area than it occupied today. This is supposed to be due to sea level changes which modified the total land area while climate change modified the geographical distribution of forest types (Cannon, Morley and Bush, 2009). It is estimated that lowland forests were probably approximately reduced from 1.3 to 0.8 x 10^6 km^2 while upland forests were probably reduced by half, from approximately 2.0 to 1.0 x 10^5 km^2 and coastal mangrove and swamp forests too experienced dramatic change during deglaciation.

It is supposed that continental Asia and the Sunda shelf would have been directly connected by land and forest-covered throughout Miocene (Hall, 2009; Thomas *et al.*, 2012) specifically, a large part of Sunda shelf (currently between Sumatra and Borneo) was covered by lowland forests (Cannon, Morley and Bush, 2009). They further explained that Sundaland forests have been noticeably different in composition and structure. It is supposed that the complex geological history, limited volcanic activity and climate change over past millennia are the reasons for creation of mixed tree communities in Borneo (Cannon and Leighton, 2019). Early Cretaceous vegetation of Sundaland was dominated by gymnosperms with thick cuticles (Bruyn *et al.*, 2014). According to Morley (2000), angiosperms had begun to evolved in late Cretaceous and is evident that the southeast Borneo today, had been predominated with palm-dominated peat swamps during the middle Eocene (Harley and Morley, 1995).

Biotic dispersal from continental Asia to the Sunda shelf would have potentially been over land dispersal (Thomas *et al.*, 2012) and the earliest Sunda–Sahul disjunctions involved long-distance dispersal over distances up to 450 km (Hall, 2013). Borneo particularly hosted extensive rainforest distribution throughout all periods of time (Bruyn *et al.*, 2014). However, Cannon and Leighton (2019) found that the rain forest trees have limited dispersal capacity across Borneo, suggesting that distant communities would have evolved independently leading a high level of endemism across the island.

A series of tectonic events has influenced biotic diversification in the Southeast Asian region, and it has found that the biotic dispersal across Walacea's line has mostly occurred from west (Sunda shelf) to east (Sahul shelf) during the Miocene. This historical biotic exchange is evidenced by numerous studies carried out using floristic and fossil data specifically, molecular phylogenies and methodological developments in molecular clock dating methods which allows timing of the evolution of lineages (Crayn, Costion and Harrington, 2015). The hypothesis of west to east dispersal is supported by Richardson *et al.* (2014) in his study about Sapotaceae and also added that Indo-Sri Lankan and Australasian linages would have been derived from Sunda region. Specially, the predominant trends of *Begonia* dispersals between continental Asia and Malesia has also been found to be from west to east, where continental Asian area is the origin for Malesian begonias (Thomas *et al.*, 2012). This pattern is not limited to floral dispersal but also found a clear west-east dispersal followed by a rapid radiation in Southeast Asia of fresh/brackish water killifish within the region (Beck *et al.*, 2017). However, few studies have been found extensive biotic exchange in the reverse direction from east to west (Richardson, Costion and Muellner, 2012).

Tropical South East Asia harbors an extraordinary species richness comprising four of earth's 34 biodiversity hotspots and high degree of local and regional endemism (Bruyn *et al.*, 2014). Borneo is one of the major evolutionary hotspots for South East Asian biodiversity hosting the highest levels of vascular plant species and terrestrial mammalian richness in SE Asia, whereas the top 10% of mammalian species richness is hosted in the northeast of the island (Bruyn *et al.*, 2014). Importantly, Borneo is one of the top five centres of vascular plants with more than 5,000 species per 10,000 km^2 whilst the island covers only 0.2% of the terrestrial surface (Bathlott and Rafiqpoor, 2005; Mutke and Barthlott, 2005; Zachos and Habel, 2011). Additionally, Bornean forests are home to much rare and important species like orangutan, rhinoceros hornbill, giant-flowered Rafflesia and the durian fruit (Primack and Hall, 1992) while several Bornean plants and animals are already considered as threatened species in the IUCN red list (*IUCN 2019*).

Borneo mainly consists of lowland forests, mangrove forests, peat swamp forests, limestone forests, heath forests and various montane vegetations which supports high species richness as well as the great endemism across the island (MacKinnon *et al.*, 1997). Besides generalized inland flora, two other notable floristic subprovinces; "Rio Pocket" and "East coast Sabah" with rather different plant composition has been identified (Wong, 1998). Recent study of

Roos *et al.* (2004), found that Borneo has estimated total of 14,423 plant species with 28.4% endemism. The most endemic plant species can be found in smaller areas in the north, the central mountain chain and in the south eastern Meratus mountains (Slik *et al.*, 2003; Raes *et al.,* 2009) as well as the north-western Sarawak (Slik *et al.*, 2003) while areas like Kinabalu and Crocker Range are the hotspots with special plant diversity (Ashton, 1992 in Wong, 1998).

Northern Borneo is reportedly known as a biodiversity hotspot (Zachos and Habel, 2011). Raes *et al.* (2009), found that Northwest and northern Borneo; Sabah, Sarawak and Brunei, as the places where highest botanical richness predicted whereas the lowest richness has been found in southern and southwestern parts of the island. This is similar to the results of Wong (1998), even though Wong suggested it could have been due to different collection densities however, Raes later agreed with the comparatively low diversity of flora in Kalimantan. The results support the fact that northern Borneo is a centre of vascular plant diversity (Bathlott and Rafiqpoor, 2005; Mutke and Barthlott, 2005; Kreft and Jetz, 2007).

Borneo climate and soil properties are important determinants of diversity and richness of plant species, are markedly change with mountain slopes. The elevational patterns of the vascular plant species richness has been quantified by numerous studies where most of them found the species richness is decreased exponentially with elevation and also depends with the plant study group (Kessler, 2001; Grytnes *et al.*, 2006; Brambach *et al.*, 2017). Findings of Slik *et al.* (2009) about environmental correlates of the tree diversity patterns of Borneo, also explained that elevation together with soil depth, soil texture and annual rainfall is significantly correlate with tree diversity across Borneo. Many Bornean taxa are more species rich than Malay peninsula, suggesting that this is because of larger land area, median position in Malay archipelago and greater geographical complexity (Wong, 1998).

1.2 *Begonia*

1.2.1 Mega diversity and morphology

Begonia L. (Linnaeus, 1753), a megadiverse genus of family Begoniaceae, is the sixth largest angiosperm genera with 1947 species and is classified to 70 accepted sections (Moonlight *et al.*, 2018). The genus named in honour of Michel Begon during 1638-1710, who was French

Governor of Santo Domingo in the West Indies from 1682-1685 and was a patron of botany (Kiew *et al.*, 2015).

Begonia's are hydrophilous herbs or shrubs, prefer strongly sheltered environments (Hughes, Girmansyah and Ardi, 2015) and commonly found in wet, shaded areas like stream banks, steep slopes and mist zone around waterfalls in the forest (Hughes, 2002; Kiew *et al.*, 2015) and commonly known as an economically important ornamental plant (Goodall-Copestake *et al.*, 2010).

Figure 6: Different growth forms of Begonia species growing in shady, moist habitats. (a) B. magnicarpa, (b) B. renek, (c) B. calcarea and (d) B. paoana (Photos-Mark Hughes).

The genus is easily recognizable through asymmetric leaves, unisexual-monoecious flowers with twisted-papillose stigmas, three- winged capsules though there are numerous deviations leading the mega-diversity of the genus (Dewitte *et al.*, 2011). Interestingly, this morphological diversity showing some interesting adaptations towards a successful establishment on the land. Hughes (2002) explained that asymmetric *Begonia* leaves showing many adaptations such as; bullate leaves, glistening/satiny leaves, anthocyanin rich lower surfaces and iridescence leaves which thought to be for coping with shade in the forests for increase the efficiency of photosynthetically useful light capturing.

The further, majority of begonias are monoecious, bearing male and female flowers in the same inflorescence and many of them are insect pollinated, mostly by bees while with few exceptions. Only the male flowers of *Begonia* offer pollen for pollination but interestingly, the yellow stylodia in female flowers mimic androecium of male flowers to attract pollinators to complete the life cycle.

Fruits are commonly three winged dehiscent capsules while some shows variations, one of the wings may be considerably enlarged. Seeds are small and some bearing adaptations for more efficient wind dispersal. However, according to the Hughes and Hollingsworth (2008), very high number of *Begonia* species are narrowly endemic suggesting a poor dispersal while relatively small number of species are wide spread.

1.2.2 Global distribution of *Begonia*

The origin of *Begonia* is thought to be in Africa during the Oligocene (Rajbhandary, Hughes and Shrestha, 2011b) and substantially distributed into Asian and American regions over last 20-50 million years during the Miocene (Goodall-Copestake et al., 2010; Dewitte et al., 2011; Thomas et al., 2012) through either long-distance dispersal or over land dispersal (Thomas et al., 2012). The genus has a pantropical distribution and is only absent in Australian forests and the pacific region from east of Fiji to Galapagos islands (Thomas, 2010). The distribution of begonias uneven throughout tropics with the highest diversity find in America (ca. > 600 species) and Asia (ca. 750 species) where the bulk is in southeast Asia (Rajbhandary, Hughes and Shrestha, 2011b). A relatively poor *Begonia* richness (ca. 160 species) is in African region (Dewitte et al., 2011).

1.2.3 *Begonia* in Borneo

Borneo is also a hotspot of diversity for the mega-diverse genus *Begonia*. To date 254 *Begonia* species have described from Borneo which comprises Brunei, Sabah, Sarawak and Kalimantan (Hughes *et al.*, 2015). The further, data shows that Sarawak has the highest *Begonia* diversity of the island with 149 species following that Sabah- 82 species, Kalimantan- 24 species and Brunei with 23 species which known as the only well-studied region in Borneo. However, it is also supposed to be more *Begonia* species in Kalimantan considering its largest land area compare to other regions.

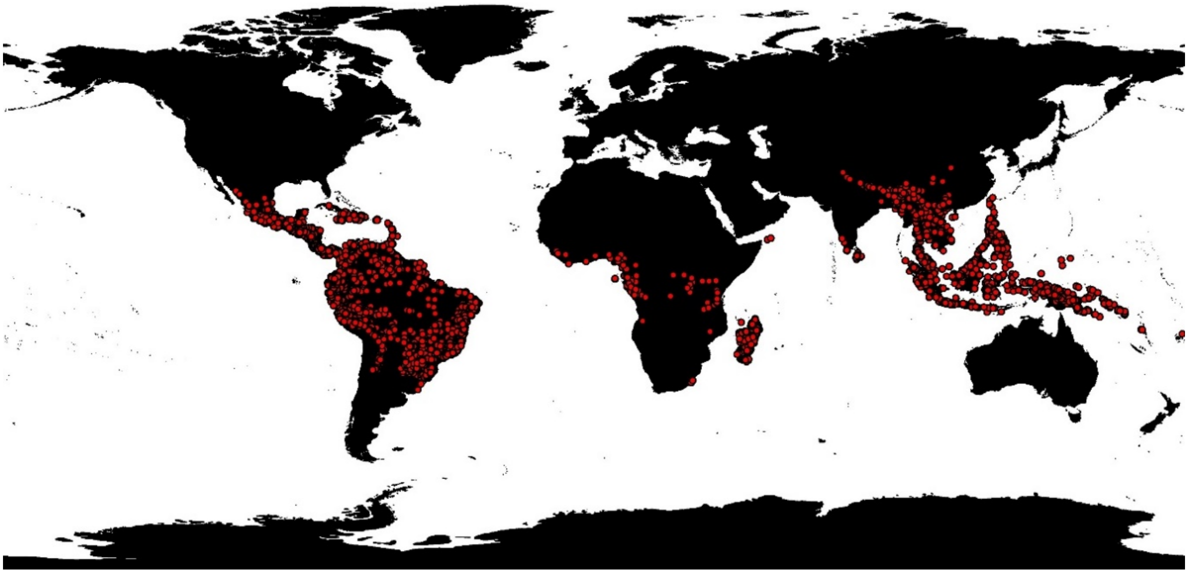

Figure 7: Global distribution of Begonia (Hughes et al., 2015).

According to Sang and Kiew (2014), all begonias are in Borneo are endemic to the island while the level of local endemism at the province level is also high more than 90%. Only six species (*B. baramensis, B. borneensis, B. chlorocarpa, B. cyanescens, B. fuscisetosa* and *B. stenogyna*) are cross the political boundaries of Brunei, Sabah, Sarawak and Kalimantan where only *B. baramensis* and *B. fuscisetosa* can find in all four regions. She further mentioned that about 80% *Begonia* species have only single collection record and many others also collected only from single locality, confirming the narrow endemism of begonias in the island. This is also confirmed by the data used for this study where 41% of species have only single collection record while 71% species have less than five.

The tropical climate of Borneo provides ideal conditions to *Begonia* growth whereas warm temperature throughout the year with no dry seasons (Kiew *et al.*, 2015). Rainfall ranging from the lowest 2850 mm in Sabah to the highest 4000-6500 mm in the Sarawak highlands offer ideal moist conditions in shady forests. Further, the habitats rich in limestones, mountains and rocky streams are preferably support for the *Begonia* dispersal whilst the lowland forests, peat swamps and heath forests are relatively poor in for the purpose.

1.3 Species distribution analysis

1.3.1 Species distribution modelling

Information related to distribution of species is fundamental in application of a broad set of conservation, ecological and evolutionary questions. Species distribution models (SDM) also known as ecological niche models are a suite of methods that predict species' distribution through relating their presence records to environmental factors, have become more popular over the past decade (Thuiller *et al.*, 2009; Zimmermann *et al.*, 2010; Syfert *et al.*, 2014; Moonlight, 2017).

SDMs are extensively used to investigate distributional patterns and narrow endemism of numerous species over wide range of geographical areas. For examples, De Klerk *et al.* (2002), studied the patterns of species richness and narrow endemism of bird species in Afrotropical region, and the species with a range-size of 4 or few grid cells have been defined as range restricted species, and used as a discontinuous measure of narrow endemism. Raes *et al.* (2009), predicted the botanical richness and endemicity patterns of Bornean plant species using SDM approach. In his findings, the hotspots of endemic species have been identified by mapping the relative residuals of the species richness-weighted endemism relationship. Further, Rajbhandary, Hughes and Shrestha (2011a), has also been used SDM to assess the distributional patterns of *Begonia* species in Nepal against climatic and topographical variations. The resulted SDMs predictions identified the two main localities of the Nepalese *Begonia* can find and the differences of ecology and phenology of two large *Begonia* sections occurring in Nepal.

In addition, the use of SDMs for identifying conservation priority areas has been demonstrated by Syfert *et al.* (2014). The study investigated how effective the SDM derived predictions of the extent of occurrence of a species with small number of collections from point-based predictions when it comes to a red list assessment. Furthermore, Zhang *et al.* (2012) used SDMs to assess priority areas of conservation, species diversity hotspots the level of endangerment in Yunnan province in China using the same approach.

SDM predicted range for American bullfrog, *Rana catesbeiana*, a major invasive species has successfully been evaluated the establishment and invasion of the species where monitoring and management efforts need to be focused (Ficetola, Thuiller and Miaud, 2007).

Interestingly, some studies have been focused to predict range shifts of species due to climatic change using species distribution models. Findings of Zhang *et al.* (2014) has found that, up to 1400 species (of the 1996 tested) are expected to lose more than 30% of their current range under the most extreme climate change scenarios, mainly due to increasing temperature variability and declining precipitation during the dry season in Yunnan, southwest China.

1.3.2 Species Distribution Modelling vs Minimum Convex Polygon method

According to Syfert *et al* (2014), the SDM derived extent of occurrence of certain species may be better representative of full extent of occurrence than a minimum convex polygon (MCP) drawn around the known localities. The extent of occurrence is the widely accepted method in IUCN red list assessment in estimating the extinction risk of species; it does not intend to be an estimate of amount of area occupied or potential habitat of a species. Further, this is relatively well interpreted only with “well sampled” species where the species represent their full geographical extent while poorly sampled species represent a proportion of its geographic range (Syfer*t et al*, 2014). Unlike MCP, Species distribution models predicts the potential distribution of a species by interpolating identified relationships between presence/absence or presence only data with environmental predictors across an area of interest (Elith *et al.*, 2006).

However, there are numerous challenges discussed in determining the quality of distribution models since the model outputs are highly depended on the existent techniques and the quality of data. Radosavljevic and Anderson (2014) pointed out several methodological issues unresolved yet in model building such as; selecting appropriate evaluation data, detecting overfitting, and tuning program settings to approximate optimal model complexity. Wide range of studies have been assessed species occurrence data in terms of sample size and environmental factors (Pearson *et al.*, 2007), sampling bias (Pearson *et al.*, 2007) and thresholds (Escalante *et al.*, 2013; Liu, White and Newell, 2013; Liu, Newell and White, 2016) towards model quality.

“Threshold” is the transforming value, which transform the model output (non-binary) to a binary prediction therefore; selecting the threshold is with paramount importance. Numerous studies have shown that many methods of selecting thresholds for presence/absence data while maximizing the sum of sensitivity and specificity is one of the best threshold selection method for presence/absence data (Liu *et al.*, 2005). Liu, White and Newell (2013), had investigated thirteen existing methods of threshold selection, which can be used with presence only data of species. They had further investigated whether the same threshold can be produced using either presence/absence data or presence-only data. The method based on maximizing the sum of sensitivity and specificity while confirming it as a promising method for threshold selection when only presence data are available. On the delimitation of the areas of endemism for North American mammals, Escalante *et al.* (2013) has been evaluated the effect of threshold using four thresholds in MaxEnt; minimum training presence, tenth percentile training presence, equal training sensitivity and specificity, and 0.5 logistic probability, showing the results that tenth percentil training presence threshold were giving the best identification of areas of endemism in north American region. Norris (2014), has also been analysed the suitability of different threhsold on model performance during his study about the distribution of lowland tapir (Tapirus terrestris) in a continuous Atlantic forest of southeast Brazil and the resulted models have been substantially different with different thresholds. Here, importantly the identified results were contrast with the Liu, White and Newell (2013) highlighting the importance of understanding the species natural history to determine not only appropriate model parameters, but also the biological relevance of SDMs.

1.3.3 MaxEnt for Species Distribution Modelling

Maxent is a programme for modelling species distributions. It is a general-purpose machine learning method with a simple and precise mathematical formulation (Phillips, Anderson and Schapire, 2006). Unlike many other modelling programmes which uses both presence and absence data, MaxEnt uses presence only data for predicting species distributions (Elith *et al.*, 2011; Philips *et al.*, 2017). Since most of the species records are available in the form of presence-only records in herbariums or databases, the desire to use Maxent has widely increased to maximize the usage of such hugely important data sources. Accordingly, hundreds of biological studies have been used and are still using MaxEnt in predicting distribution of species of interest across the globe since becoming available in 2004. Further, MaxEnt has been described as it is especially efficient in handling complex interactions between predictor and response variables (Elith *et al.*, 2006, 2011), performs with great accuracy even with a small and biased data set (Rebelo and Jones, 2010) and is extremely easy to use which, made MaxEnt up to the most widely used SDM algorithm (Fourcade *et al.*, 2014).

Published examples of studies which used MaxEnt cover diverse aims of analyzing species distribution data such as; predict species richness, diversity and endemism (Raes *et al.*, 2009; Rajbhandary, Hughes and Shrestha, 2011a; Vollering *et al.*, 2016; Syfert *et al.*, 2018), forecast distributions to understand changes with climate change (Qin *et al.*, 2017; Abolmaali, Tarkesh and Bashari, 2018), predict current distribution as a input of conservation planning (Tinoco *et al.*, 2009; Tittensor *et al.*, 2009) and for analyzing model performance against other available modelling methods (Elith *et al.*, 2006).

1.3.4 SDMs from small sample sizes

Generally, the predictions based on few records of occurrence are not good as those based on a larger number of occurrence records (Hernandez, Master and Albert, 2006; Papeş and Gaubert, 2007; Wisz *et al.*, 2008) and the minimum acceptable sample size varies between different studies. However, MaxEnt can perform well with as lower as 10 sample size (Wisz *et al.*, 2008; Fourcade *et al.*, 2014) with excellent predictability. This is supported by Stockwell and Peterson (2002), obtaining the 90% of average success of predicting occurrence of a species within ten sample points. In 2007, Pearson and his team found a high success rate with statistical significance with as few as five sample points to produce a biologically meaningful model while van Proosdij *et al.*, (2016) assessed the minimum sample sizes as low as 3 for narrow ranged and 13 for widespread species for a virtual study site which represents an ideal, balanced, orthogonal world while the minimum was 14 for narrow-ranged while for 25 for widespread species in real conditions. Therefore, five is the lowest sample size described which giving a biologically realistic prediction up to date.

1.4 Aims of the study

Borneo, a major evolutionary hotspot for southeast Asian biodiversity, is also a hotspot for pantropical mega-diverse genus *Begonia* which currently harbors 254 known *Begonia* species across the island (Hughes *et al*., 2015). According to the available collection records, Sarawak home to the highest *Begonia* richness, holding 149 species where Kalimantan, the Indonesian part of Borneo holds only 24 species. Sabah and Brunei are home to 82 and 23 species respectively. On one hand this could be because northern Borneo is exceptionally rich in biodiversity compare to southern parts where much area is covered with lowland forests, peat swamps and heath forests. However, by looking at *Begonia* collection densities, Kalimantan, which occupies the largest land area of 743 330 km^2, has the lowest collection density (0.03 species/1000 km^2) of the island seemingly underexplored for *Begonia* richness. This could be due to collection bias, though much of the land area in Kalimantan suitability for *Begonia* is unknown. However, inevitably, many collections have made in easily accessible areas in Sabah, Sarawak and Brunei than hardly accessible Kalimantan habitats, but it is sure that undisturbed forest patches which ideally suits for *Begonia* growth are still available in Kalimantan.

Therefore, this study aims,

i. To map hotspots of predicted *Begonia* species richness on Borneo,
ii. To identify data gaps where no collections have been made to date and future collections should be done,
iii. To identify biogeographical correlates of *Begonia* richness in Borneo and explain the distribution of species richness,

through species distribution modelling of *Begonia* distribution data recorded in *Begonia* Resource Centre, the online *Begonia* database of Royal Botanical Garden Edinburgh.

CHAPTER 2: MATERIALS AND METHODS

2.1 Data preparation

2.1.1 Databasing

Our occurrence dataset comprised presence-only data from the *Begonia* database of the Royal Botanical Garden Edinburgh (Hughes *et al.*, 2015). In addition, new collection records for Brunei, Sarawak and Sabah in (Ahmad *et al.*, 2015; Repin, Karim and Mujih, 2015; Repin, Sampang and Limbawang, 2015; Sang and Yiing, 2015; Sang *et al.*, 2015; Sang, Kiew and Yiing, 2015; Yun *et al.*, 2015; Yun, Guanih and Repin, 2015) were added to the database. All specimens from Kalimantan in the E herbarium were also databased.

2.1.2 Morphogrouping

Unidentified *Begonia* specimens from Kalimantan were grouped together into morphospecies considering their morphological characters. The characters such as; leaf shape, leaf margin, leaf base, length of petiole, length of internode, hairs on stems and leaves and some other useful characters were carefully observed. In addition, some useful information in the labels such as name of the collector, collection date and locality details were also considered in grouping. Each group of specimens were considered as one species and named as Species A, Species B, Species C and etc.

2.1.3 Geo-referencing

The dataset originally comprised with both geo-referenced and non-geo-referenced specimens. Approximate coordinates were searched for the specimen records which weren't geo-referenced at the collection time but had clear locality details, with the aids of external gazetteers, Google Earth, other web-based digital repositories such as Global Biodiversity information Facility (GBIF) and specific other literature sources. Significant effort has made into ensuring the validity of such localities, through expert verification and precise geo-referencing. Duplicates and less reliable geo-references were removed with expert judgement prior to the analysis, through checking matches with altitude, coastline and provincial boundaries.

2.2 Maxent distribution modelling

The distribution of *Begonia* species with more than five geo-referenced data were modelled in Maxent version 3.4.1.

2.2.1 Obtaining environmental data for SDM

Nineteen bioclimatic variables were obtained from WORLDCLIM database version 2. with a 30 arc second (~1 km at the Equator) spatial resolution (http://www.worldclim.org/, Fick and Hijmans, 2017)

2.2.2 Check the effect of 'Regulization multiplier' on model building

Before conducting the main experiment, a preliminary analysis was carried out to test how closely the output distribution fitted to the given presence records and expert opinion. The 'Regulization Multiplier' is a parameter that adds additional constraints to the model to prevent over-fitting by controlling the intensity of chosen feature class used to building the model (Morales, Fernández and Baca-González, 2017). Several values of the regularization multiplier were tried, changing the default setting '1' to smaller values and larger values, to get a more localized output distribution for the given presence records of narrowly distributed species. The models were run using the following values of the regularization multiplier: 0.01, 0.1, 0.25, 0.5, 0.75, 1, 1.5, 2 & 3. Then the models were compared to expert opinion for the best fit for a genus which has so many narrow endemics, to make the final map more biologically realistic.

2.2.3 Model *Begonia* distribution in MaxEnt

Maxent works with presence only data has proved successful with numerous distribution modelling studies including many with small sample sizes. We used Maxent version 3.4.1 to make the distribution models to estimate probability of presence of *Begonia* species, assuming that typical presence localities have 0.5 of probability of presence (Philips *et al*, 2017). For the species with more than 5 collections, we fitted using "10-fold-cross validation approach" which involves randomly dividing the set of observations into 10 groups of approximately equal size. The first group is treated as a validation set, and the model is fit on the remaining nine groups. In other words, 90% of our data is using to build the model while randomly set 10% is using for model validation (Syfert *et al.*, 2018).

2.2.4 Evaluation of the model outputs obtained from the Maxent

The performance of the SDMs was evaluated using the model outputs. The first assessment was done using the area under the curve (AUC) of the receiver operating characteristic (ROC) plot. The ROC plot shows the "fit" of the model to training and test data where the blue line indicates the fit of the model to the testing data. This is a good indication of the models' predictive power. When the blue line is towards the top left corner of the graph, the model is better at predicting the presences in the test sample. When the AUC value is one, it considers as the model is with perfect sensitivity and specificity whilst the models with AUC value greater than 0.7 are usually consider as good for further analysis.

Jackknife plots were assessed which show the importance of each of the environmental variables when only that variable is included and also when only that variable is excluded but all the other variables are present. Though the model output gave three Jacknife important curves, the priority was given for the AUC jackknife plot instead of the test and training jackknife plots and was only checked if they show something significantly different.

2.2.5 Threshold selection and binary mapping

Since the MaxEnt gives the probability of presence of the species which is a continuous variable ranging from 0 to 1, we needed to make a binary decision to see whether the habitat is suitable or unsuitable for a particular species. The parameter called "Logistic threshold" in the omission threshold table in the MaxEnt output is often used to make the decision that gives some sensible measurement about the probability of suitability of a habitat.

However, the effect of the logistic threshold for different individual species can give different presence-absence results which are not realistic with their actual occurrence. The effect was tested by changing the thresholds for selected species, one with a very narrow distribution (*B. paoana*) and one with a fairly wide distribution (*B. fuscicetosa*), for which the distributions are well known. The threshold producing the most biologically realistic presence-absence map was chosen.

Secondly, the thresholds were individually selected for each species considering the p-value, training omission value and the natural occurrence of the species. The priority was given for the thresholds with p-values lower than 0.05 and lowest possible training omission. Along with those two values, the model’s predictive power was also considered. The thresholds which were either equally good at predicting presences and absences for the test data (“values with Equal test sensitivity and specificity) or which maximized sensitivity of the model with test data (“values with Maximum test sensitivity plus specificity”) were selected. For the species with higher training omission rates at these two options, other thresholds values from the training outputs were considered.

2.3 Single Point Extrapolation

Theoretically, MaxEnt needs a minimum of ten for the sample size to give reliable predictions, however Pearso*n et al* (2007) found high success rates and statistical significance in Jackknife tests even with sample sizes as low as five. Since many of the Bornean begonias are micro endemics and have only one or two collection records, those species cannot be modelled using MaxEnt. Therefore, 131 species with less than five occurrences were manually modelled in R version 3.6.0 and ArcGIS, using a method designed for this thesis called ‘Single Point Extrapolation (SPE)’. In this method, we looked for pixels of very similar climate in the immediate area to model the distribution of these micro endemics.

2.3.1 Selection of mostly correlated bioclimatic variables for *Begonia* distribution

There are 19 layers in the Bioclim dataset and using all layers may have given inflated values for some cells due to environmental similarity for layers which are not important for predicting *Begonia* distributions. To choose a subset of layers with high relevance for *Begonia*, I examined species with a high number of geo-referenced datapoints which were modelled in Maxent, and the bioclimatic variables with highest permutation importance along with highest percent contribution for model building were selected. These layers were chosen to build the SPE distributions. To check the effect of using this subset of layers, selected species with a single occurrence point were first modelled with all 19 bioclimatic variables, then compared with the model obtained when using only the subset of variables.

2.3.2 Model species distribution using SPE method

This is a novel approach to make species distribution predictions for species with smaller sample sizes of less than five occurrences. Initially we needed to decide a suitable range of similarity to the point sample for the target species for each layer. For that, raster maps of no further than 5%, 10% and 15% away from the optimum were produced. According to the results, 5% were taken as the most biologically realistic distance for micro endemics through expert judgement.

Step 1: Calculate the absolute difference

A raster of distances to the value of the occurrence sample point are calculated for selected bioclimatic variables. First, the absolute difference was calculated between the value of the selected occurrence pixel to each other pixel of the particular bioclimatic layer. Then the resulting value is divided by the range of the selected bioclimatic layer to calculate the standardized distance for the selected target value (Equation 1), giving a raster of values between 0 and 1 for each layer.

$$D = \text{abs}(R1 - T) / R2$$

Equation 1: Mathematical formula used to calculate standardize absolute difference from the optimum. D= Standardized absolute difference, R1=raster, T= target value, R2= the range of the raster

Step 2: Select the 5% range

In the second step, the areas no further than 5% away from the optimum were identified on a raster by raster basis (Equation 2). Then the resulting rasters were summed, giving a raster of overall environmental similarity to the sample point.

$$A = D * 0.05$$

Equation 2: Mathemiatical formula to calculate five percent distance from the optimum. A= Area within standerdize distance from the target point, D= Standerdized absolute difference from equation 1.

Step 3: Produce presence absence maps

Presence-Absence maps were produced from the environmental similarity maps from the above methods. To select most realistic threshold, maps of no further than 5%, 10%, 15%, 20% and 25% values from the highest were made. For the species with two, three or four occurrences, the resulting binary rasters based on each point were summed to make the final binary raster.

To check the effect of the threshold values, species from lowland and highland areas were independently tested.

2.3.3 Test the accuracy of SPE

In order to determine the precision of the method of SPE, a narrow endemic species but with reasonably high number of occurrence data (*B. paoana)* was modelled using the both methods, Maxent and Single Point Extrapolation. Then, a comparative analysis was done with the resulted maps of two methods.

2.4. Synthetic analysis

2.4.1 Producing the Stacked Species distribution map

Binary maps obtained from Maxent were stacked together in ArcMap version 10.5.1 to get the final prediction for the species with more than five occurrences. Similarly, binary maps of the species with low occurrences resulted from the SPE in R, 3.6.1 were also stacked together to make the "stacked model" for the species with less than five geo-referenced records. Then the two maps were sum up in R to make the final *Begonia* richness map of Borneo.

2.4.2 *Begonia* hotspot analysis

Begonia hotspots with highest *Begonia* richness on Borneo were identified and the species growing on those hotspots were analysed.

Further, the places where little or no collections have made to date were analysed for future collection purposes.

The environmental factors which preferred by those species were also determined. In addition, the possible ecological and historical factors such as; forest types, limestone and the distribution of protected areas across Borneo were compared in order to determine past and future potential factors would/will affect the *Begonia* richness of the island.

CHAPTER 3: RESULTS AND ANALYSIS

3.1 Databasing, Morphogrouping and Geo-referencing

The final dataset for Borneo is made up of 180 identified *Begonia* species and 7 morphogroups representing unknown species with sample sizes ranging from one to 45 collections, altogether 695 specimen records. Among those 135 species had less than five geo-referenced records, 49 species had more than five while only 24 species had more than ten geo-referenced records. The data set comprises species both growing in lower altitudes as well as higher altitudes (Figure 8). All the species with less than five geo-referenced data (135) were used for modelling using the SPE method explained in Chapter 2.

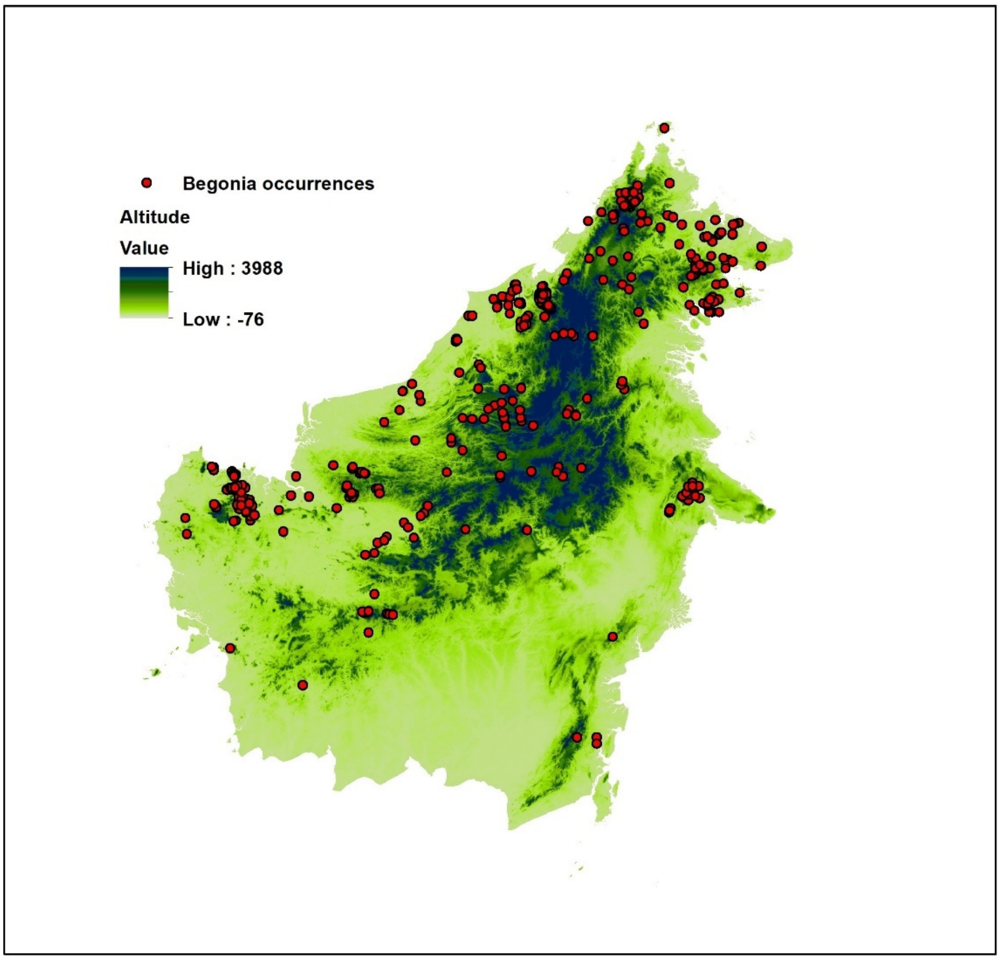

Figure 8: Distribution of Geo-referenced specimen records from Borneo; Sabah, Sarawak, Kalimantan and Brunei. Map showing the distribution of data across different elevations ranging from low (light green) to high (blue).

The number of specimen records (Table 1) were highest from Brunei (142), Sarawak (302) and Sabah (123) while many species in Kalimantan (128) were poorly represented in data set compared to the area of extent of each region (Figure 9). However, every effort was made to collate all available specimen information for these species.

Table 1: Summary of data used for species distribution analysis.

	No. of *Begonia* species	**No. of geo-referenced specimens**	**Area of extent (km^2)**	**Species/1000km^2**
Brunei	23	142	5,765	3.99
Sabah	82	123	73,904	1.11
Sarawak	140	302	124,451	1.12
Kalimantan	24	128	743,330	0.03

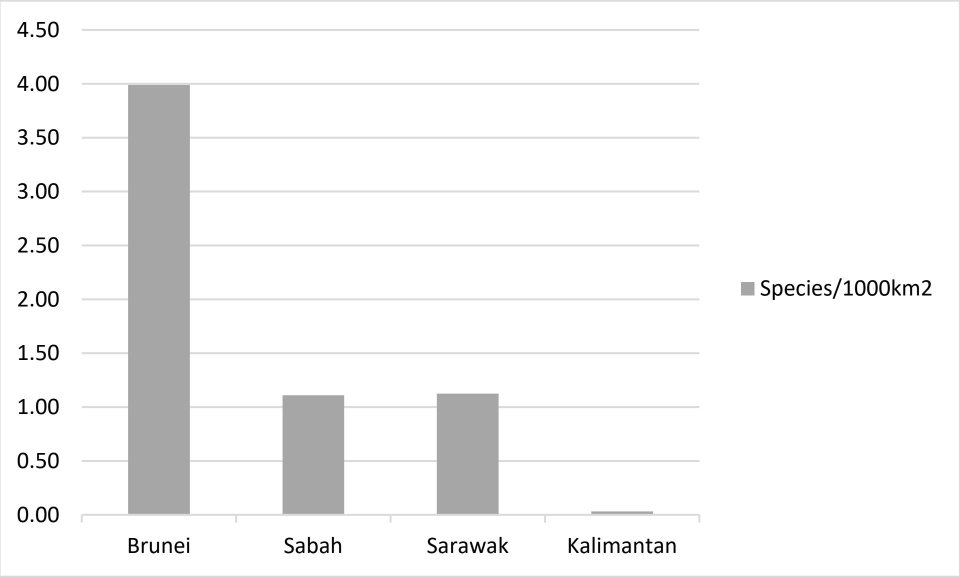

Figure 9: Variation of Begonia density recorded from Brunei, Sabah, Sarawak and Kalimantan.

3.2 Maxent distribution Modelling

3.2.1 Optimized “regulization multiplier” for narrow endemics

The distribution of *B. calcarea* in Borneo was modelled by changing the value of the parameter “regulization multiplier (RM)” in order to determine the most biologically realistic prediction for narrowly endemic species. The results shown that the distribution of the species is over fitted with default value “1”, 1.5, 2 and 3 while under fitted with 0.01, 0.1, 0.25 and 0.5 as well. The most biologically realistic prediction was given with 0.75 (Figure 10e) for the species.

Maxent predictions for default RM (1) and for 0.75 were compared in order to check the accuracy of selected RM, using stacked binary presence-absence rasters (Figure 11). Both maps predicted high *Begonia* richness in Kubah national park, Brunei and Tawau hills. However,

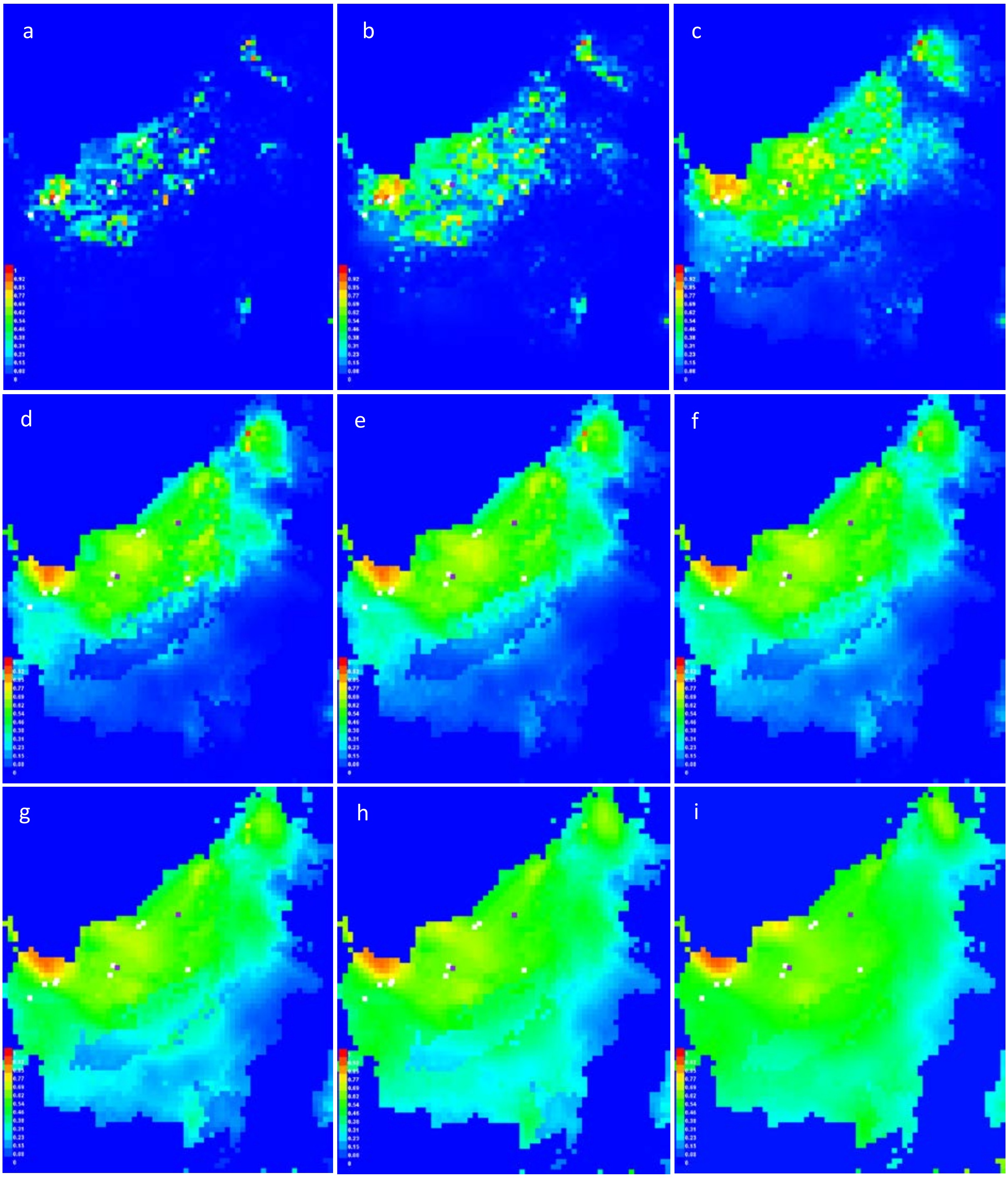

Figure 10: Predicted suitable areas for the distribution of B. calcarea given with different Regulization multiplier values in MaxEnt. Maps showing the effect of Regulation multiplier (a) 0.01, (b) 0.1, (c) 0.25, (d) 5, (e) 0.75, (f) 1, (g) 1.5, (h) 2 and (i) 3. The predictions show the suitability from low (blue) to high (red) and squares correspond to calibration (white) and evaluation (purple).

the map produced using RM 0.75 (Figure 11b) gives much higher score in Tawau hills and Northern and eastern Kalimantan areas.

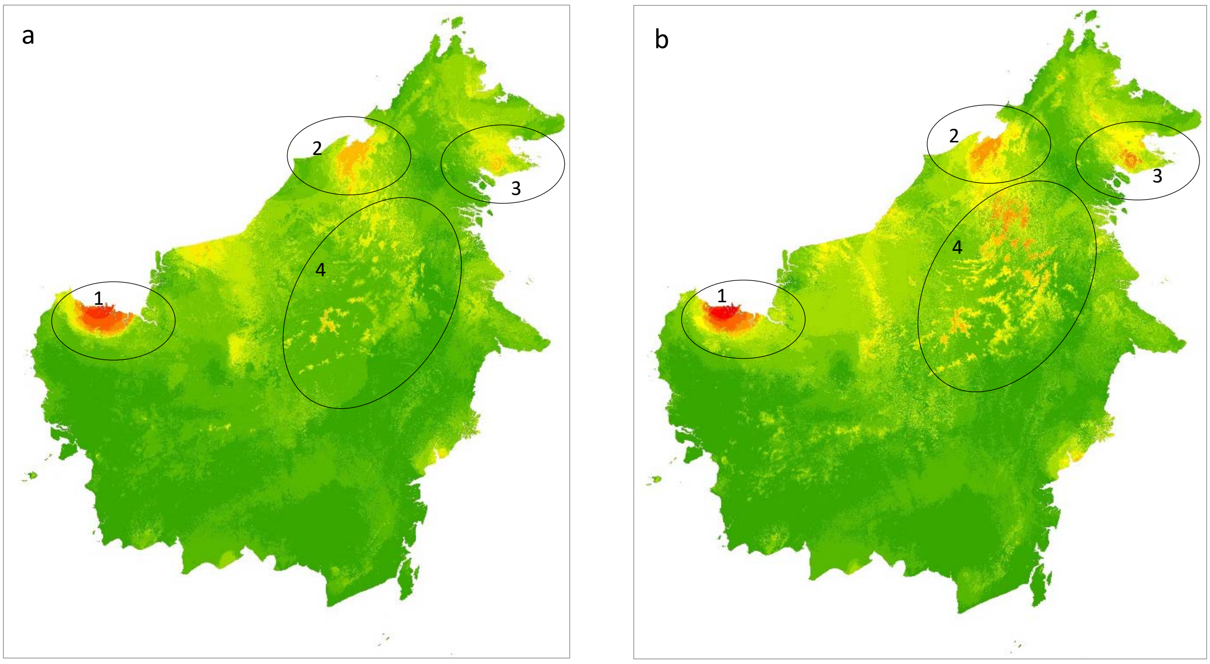

Figure 11: Maps showing the difference between predicting Begonia hotspots in Maxent species distribution modelling when using the default settings and the optimized value where (a) MaxEnt prediction when Regulization multiplier is default (1) and (b) MaxEnt predictions when the Regulization multiplier is 0.75. Predictions showing the hotspots (1) Kubah national park, (2) Brunei, (3) Tawau hills and (4) Northern and eastern Kalimantan areas while indicating the probability of occurrence from low (green) to high (red).

3.2.2 Confirmation of optimum threshold for presence-absence prediction

B. fuscicetosa and *B. paoana*, fairly widely distributed and well-known species was used to investigate the effect of different threshold values for turning the Maxent models into binary presence-absence maps. The effect of the threshold value on the distribution of species was checked to optimize the biological reliability of the predictions based on expert opinions. For *B. fuscicetosa*, most biologically realistic distributions were predicted with “Maximum test sensitivity plus specificity” as well as with “Equal test sensitivity and specificity” thresholds (Figure 12, Table 2). The prediction given using the “Balance training omission threshold” was highly over predictive (Figure 12a) while the resulted maps were slightly under predictive with both “maximum training sensitivity plus specificity” and “equal training sensitivity and specificity” thresholds.

Table 2: Maxent output of Omission Threshold Table given for B. fuscicetosa.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.014	Fixed cumulative value 1	0.312	0.000	0.000	2.96E-3
5.000	0.075	Fixed cumulative value 5	0.194	0.000	0.000	2.76E-4
10.000	0.130	Fixed cumulative value 10	0.145	0.059	0.000	6.364E-5
8.229	0.114	Minimum training presence	0.159	0.000	0.000	1.022E-4
29.284	0.276	10 percentile training presence	0.059	0.059	0.600	3.043E-2
29.213	0.276	Equal training sensitivity and specificity	0.059	0.059	0.600	3.064E-2
29.284	0.276	Maximum training sensitivity plus specificity	0.059	0.059	0.600	3.043E-2
13.210	0.160	Equal test sensitivity and specificity	0.124	0.059	0.200	1.05E-3
13.174	0.160	Maximum test sensitivity plus specificity	0.124	0.059	0.000	2.902E-5
2.842	0.043	Balance training omission, predicted area and threshold value	0.234	0.000	0.000	6.947E-4
15.273	0.179	Equate entropy of thresholded and original distributions	0.112	0.059	0.400	1.182E-2

Similar results were obtained for *B. paoana* (Figure 13), a species with a narrower distribution than *B. fuscisetosa.* However, for this species, the training omission rate is significantly high (0.667) (Table 3) with both “maximum test sensitivity and specificity” and “equal test sensitivity and specificity” which means that 66.7% percent of test data located in an area where model predicts unsuitable for the species. Therefore, when the training omission is significantly high, the thresholds of "maximum training sensitivity and specificity" and "equal training sensitivity and specificity" were taken as the cut-off.

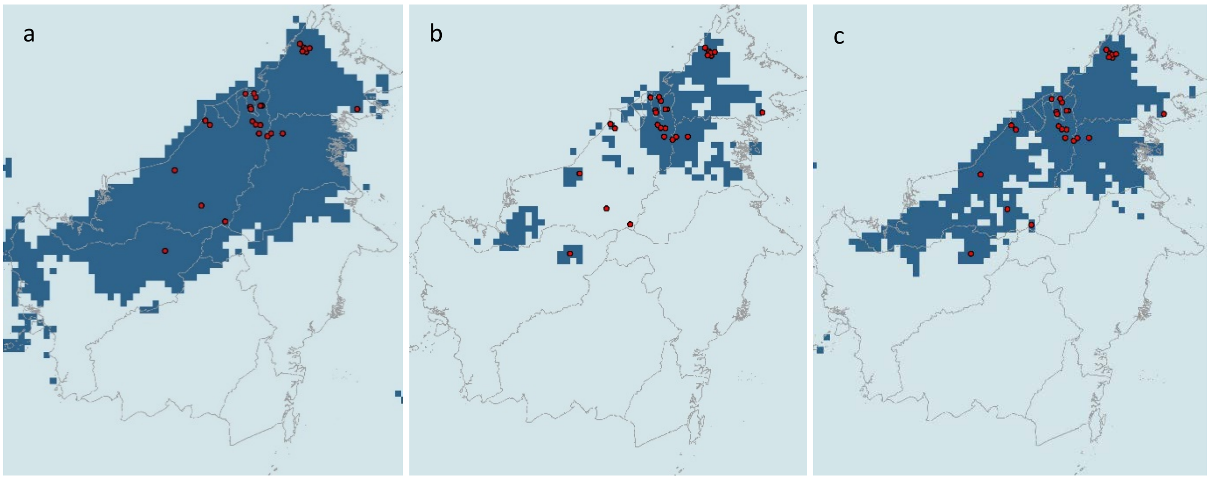

Figure 12: Maxent binary predictions for suitable areas of Begonia fuscicetosa showing the effect of thresholds. Figures a, b, and c showing the effect of threshold for B. fuscicetosa (a)with Balance training omission threshold; (b) with Maximum training sensitivity plus specificity threshold and Equal training sensitivity and specificity threshold; (c) with Maximum test sensitivity plus specificity threshold and Equal test sensitivity and specificity threshold.

Table 3: Maxent output of Omission Threshold Table given for B. paoana.

Cumulative threshold	**Logistic threshold**	**Description**	**Fractional predicted area**	**Training omission rate**	**Test omission rate**	**P-value**
1.000	0.025	Fixed cumulative value 1	0.398	0.000	0.000	1.582E-1
5.000	0.081	Fixed cumulative value 5	0.256	0.000	0.000	6.577E-2
10.000	0.147	Fixed cumulative value 10	0.192	0.000	0.000	3.698E-2
25.357	0.293	Minimum training presence	0.101	0.000	0.000	1.021E-2
25.357	0.293	10 percentile training presence	0.101	0.000	0.000	1.021E-2
25.357	0.293	Equal training sensitivity and specificity	0.101	0.111	0.000	1.021E-2

Table 3, continued.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
25.357	0.293	Maximum training sensitivity plus specificity	0.101	0.000	0.000	1.021E-2
72.740	0.728	Equal test sensitivity and specificity	0.008	0.667	0.000	6.602E-5
72.740	0.728	Maximum test sensitivity plus specificity	0.008	0.667	0.000	6.602E-5
3.892	0.062	Balance training omission, predicted area and threshold value	0.281	0.000	0.000	7.875E-2
13.395	0.188	Equate entropy of thresholded and original distributions	0.165	0.000	0.000	2.709E-2

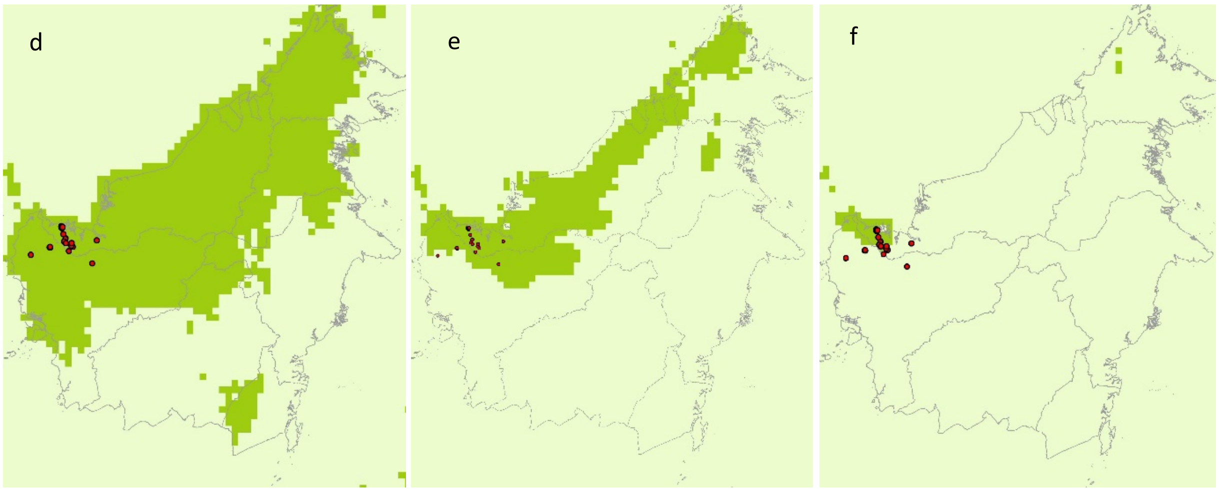

Figure 13: Maxent binary predictions for suitable areas of Begonia paoana showing the effect of thresholds. Figure d, e, and f showing the effect of threshold for B. paoana (d) with Balance training omission threshold; (e) with Maximum training sensitivity plus specificitythreshold and Equal training sensitivity and specificity threshold; (f) with Maximum test sensitivity plus specificity threshold and Equal test sensitivity and specificity threshold.

3.2.3 *Begonia* hotspots predicted by MaxEnt distribution model

According to the Maxent predicted distribution (Figure 14), Kubah forest in southwest Borneo shows the highest species richness of *Begonia* across the island. Secondly, high *Begonia* richness also predicted in northern Borneo, Sabah where in Tawau hills and Mt. Kinabalu where only showing a very narrow prediction. Similar pattern is also showing along the central mountain range of Borneo in northern and east Kalimantan area while in Balikpapan district in eastern Kalimantan also showing much suitability for *Begonia* occurrence. A wider distribution of *Begonia* is showing across the country of Brunei Darussalam including Labu forest, Labi forest, Batu Apoi forest and Ulu Temburong national parks. The lowest scores of *Begonia* richness were obtained from western and central Kalimantan areas.

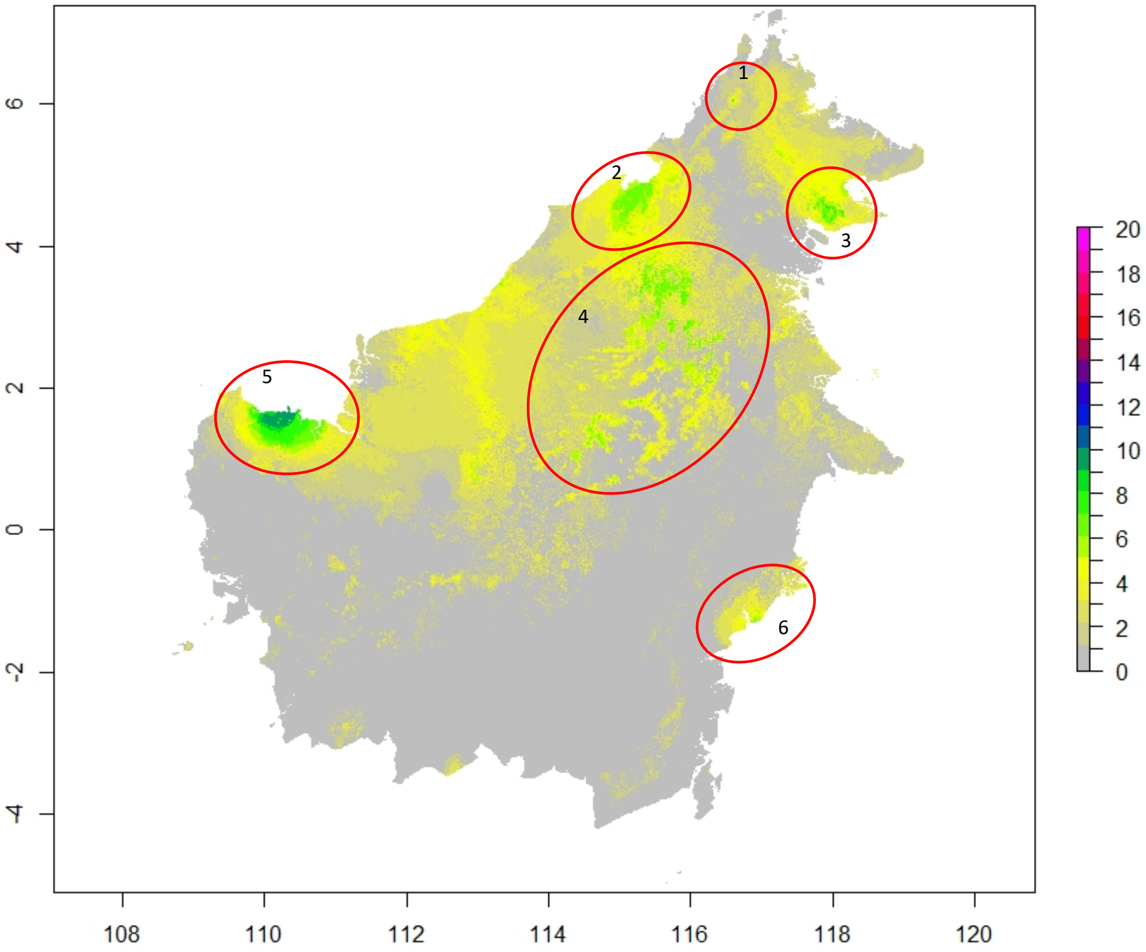

Figure 14: Predicted hotspots for Begonia distribution in Borneo using MaxEnt distribution modelling. Suitability of occurrence of species is showing from low (grey) to high (magenta) while red circles showing the Begonia hotspots identified (1) Kinabalu, (2) Brunei Darussalam, (3) Tawau hills, (4) Northern and eastern Kalimantan, (5) Kubah forest and surrounding area and (6) Balikpapan

3.3 Single Point Extrapolation

3.3.1 Selection of bioclimatic layers most influential to *Begonia* distribution

Bioclimatic variables with highest permutation importance along with highest percent contribution for model building in maxent were checked for good quality models given for *Begonia* species. Six bioclimatic variables; *Mean Diurnal Range (Mean of monthly (max temp – min temp)), Temperature Annual Range, Mean Temperature of Wettest Quarter, Precipitation of Wettest Month, Precipitation of Wettest Quarter* and *Precipitation of Driest quarter* were selected for model building in SPE method.

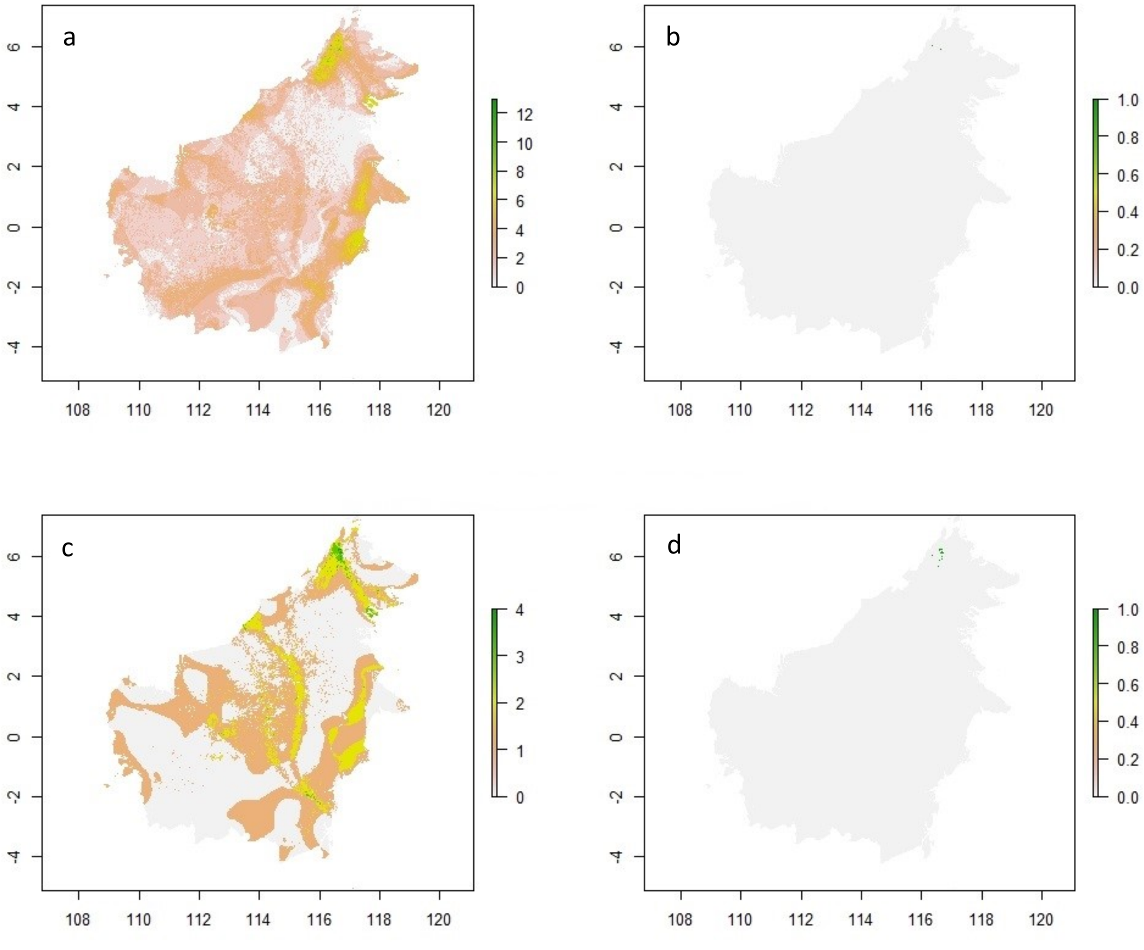

Figure 15: SPE prediction for distribution of B. adliniana not further 20% from the occurrence point. Maps a & b showing the results with all 19 bioclimatic variables while c & d showing when only with selected bioclimatic variables within 20% similarity from the highest probability of occurrence. Figure a & c are showing the heat maps while b & d showing the binary predictions in each situation.

Binary maps for the predicted distribution of some of the narrow endemic species was given an under prediction while others showed an over prediction when using all 19 bioclimatic layers. Predicted distribution for *B. adliniana* (Figure 15 a & b) only restricted to very narrow area in Mt. Kinabalu which seems to be not biologically realistic. However, the prediction given with only selected bioclimatic variables was showing a much realistic amount of distribution for narrow endemic begonias (Figure 15 c & d). Further, a highly over predicted distribution was showing for *B. addrinii* (Figure 16 a & b) in central Kalimantan and Malaysian regions with all bioclimatic variables. However, *B. addrinii* also given a realistic prediction with only selected bioclimatic variables (Figure 16 c & d).

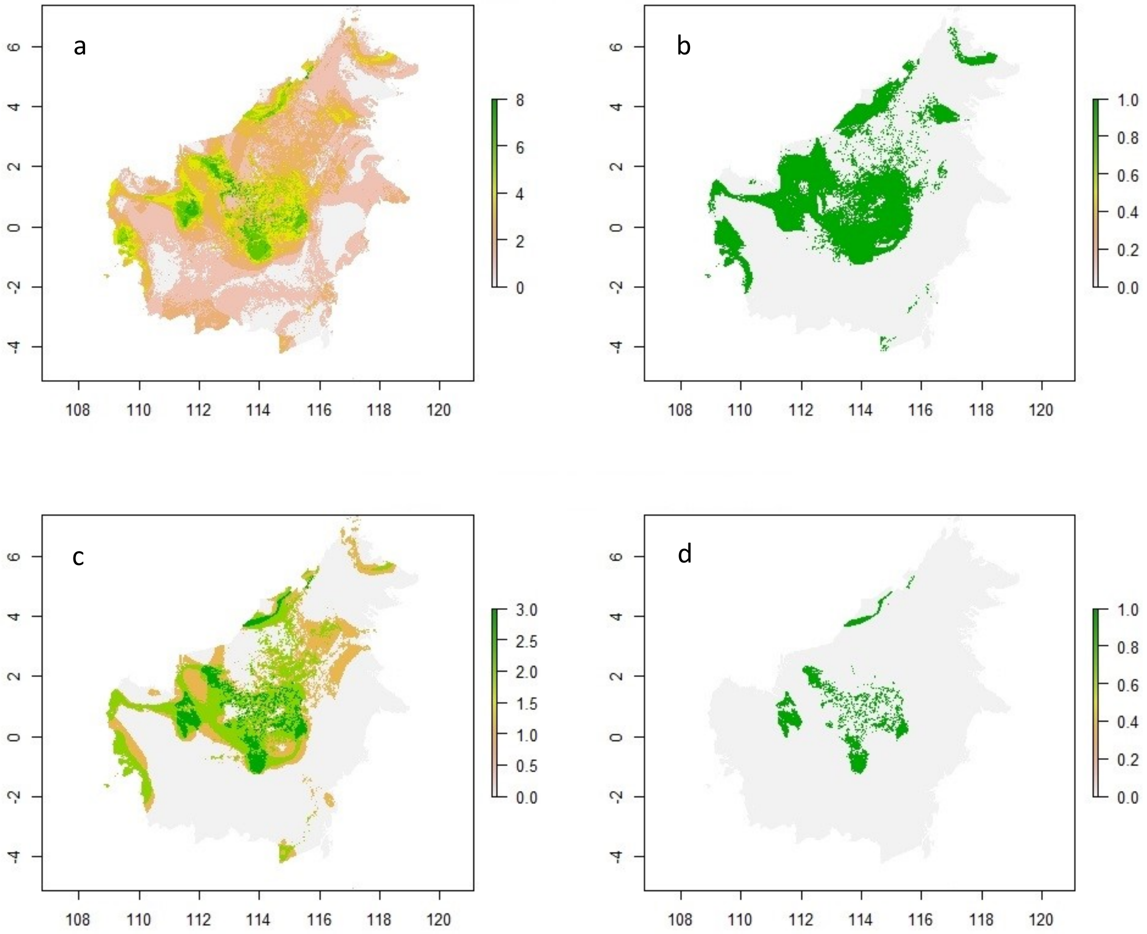

Figure 16: Predicted distribution of B. addrinii not further 20% from the occurrence point. Maps a & b showing the results with all 19 bioclimatic variables while c & d showing when only with selected bioclimatic variables within 20% similarity from the highest probability of occurrence. Figure a & c are showing the heat maps while b & d showing the binary predictions in each situation.

3.3.2 Optimum standardize range for model building using single point extrapolation

The results showing that the most biologically realistic prediction for narrowly distributed species was within the 5% of similarity from the optimum point (Figure 17a). The areas predicted within 10% and 15% similarities showing much over prediction for narrow endemics (Figure 17 b & c). Therefore, 5% similarity was taken as the suitability area for the species occurrence.

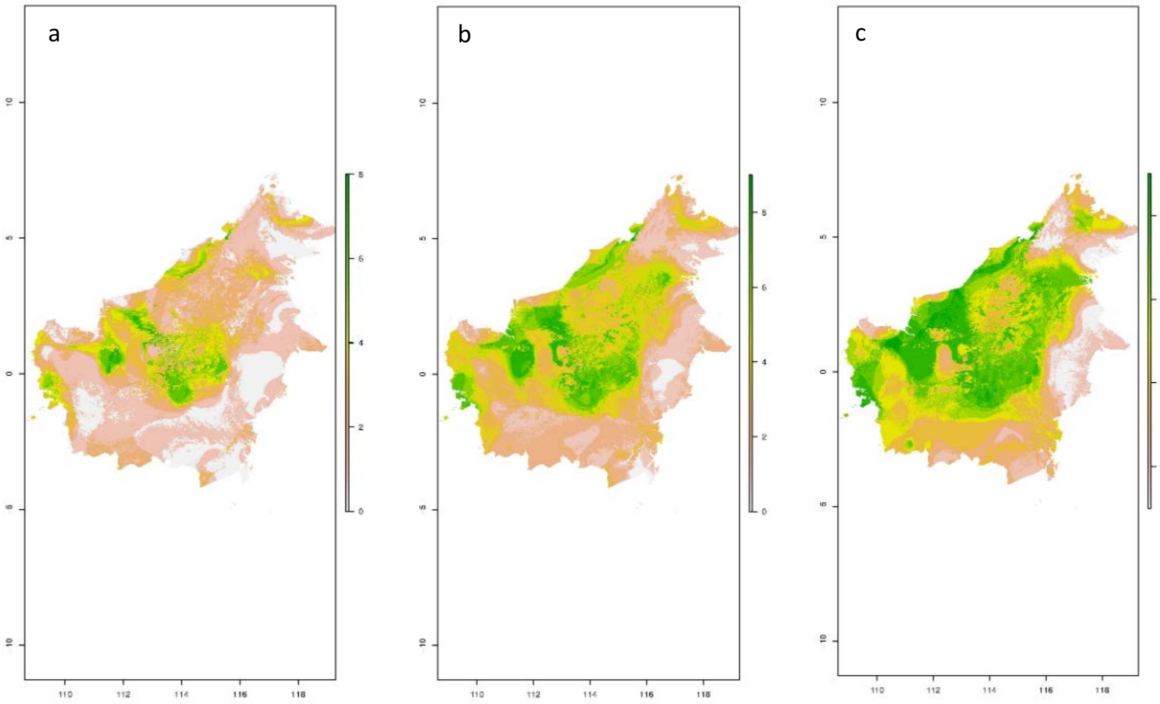

Figure 17: Predicted suitable areas for B. addrinii by SPE method, showing the effect of (a) 5%, (b) 10% and (c) 15% of similarities from the point of occurrence. Predictions showing the suitability from low (white) to high (green)

3.3.3 Selection of Threshold for presence-absence prediction

Most realistic threshold were tested from the values no further than 5%, 10%, 20% and 25% of the highest probability of occurrence to produce binary rasters. *Begonia keithii* (Figure 18), a lowland limestone micro-endemic species gave a very low areas of predictions for 5% and 10% similarities of the highest probability of occurrence. Much realistic predictions were given with the thresholds not further than 20% and 25% of the highest.

The similar pattern was observed in *B. vaccinioides* (Figure 19), a species grows in higher elevations showing under predictions up to 10% of the highest probability of occurrence while the thresholds not further than 20% and 25% of the highest are giving the most realistic predictions.

Since the most realistic distribution patterns for both lowland and highland species was the 25% similarity to the highest probability of occurrence, 25% was taken as the threshold on producing the binary rasters under SPE.

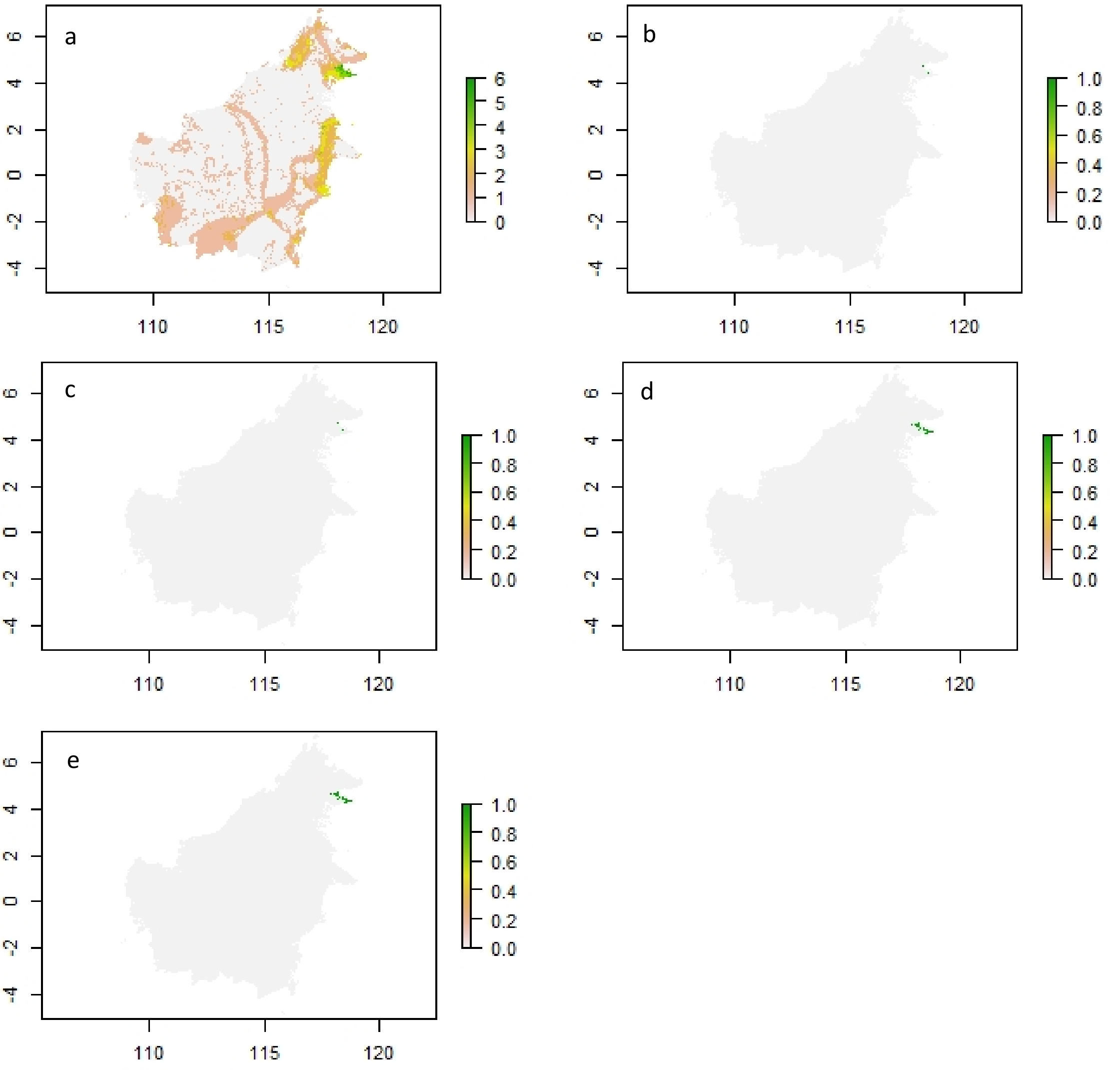

Figure 18:Predicted distribution of B. keithii with no further than 5, 10, 20 and 25 percent of similarity from the highest probability of occurrence using SPE method. Figure (a) shows the heat map while (b) 5%, (c) 10%, (d) 20% and (e) 25% showing the binary predictions of distribution of the species. Predictions showing the suitability from low (white) to high (green).

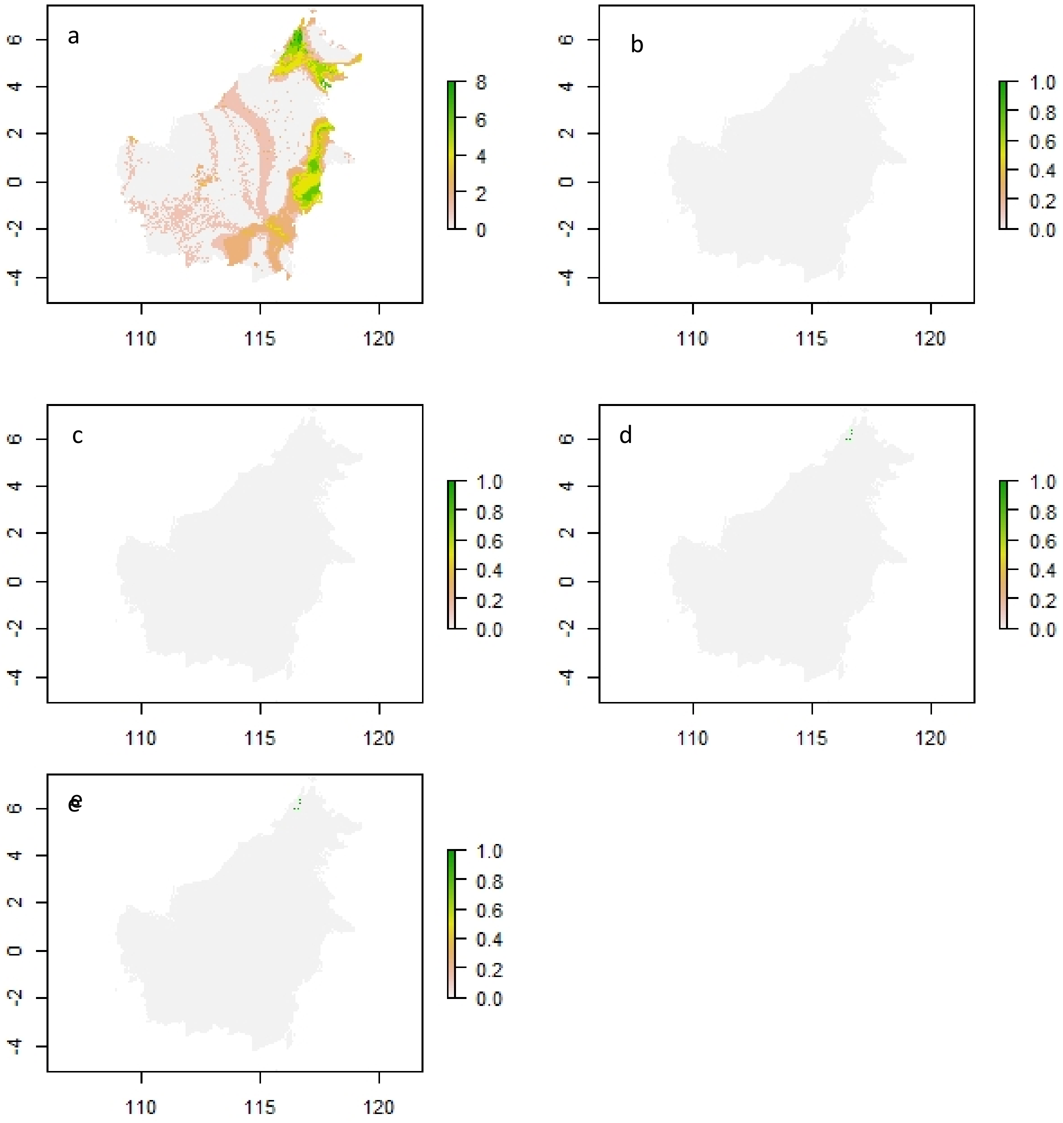

Figure 19: Predicted distributions of B. vaccinioides with no further than 5, 10, 20 and 25 percent of similarity from the highest probability of occurrence using SPE method. Figure (a) shows the heat map while (b) 5%, (c) 10%, (d) 20% and (e) 25% showing the binary predictions of distribution of the species. Predictions showing the suitability from low (white) to high (green).

3.3.4 *Begonia* hotspots predicted by Single Point Extrapolation

Unlike in Maxent prediction, the map from SPE (Figure 20) of micro endemic begonias is clearly emphasizing the hotspots in Sabah with highest richness compare to the maxent prediction. Overall score is highest particularly in Mt. Kinabalu and Tawau hills. Second highest scores observe in the Batang Ai national park and upper- Sintang district in Sarawak. However, the areas with high scores in Maxent prediction like Brunei, Kubah national park and northern and central Kalimantan show slightly lower scores in SPE method. Species used for this method are not predicting western and central Kalimantan while only very few areas are preferred to grow in northern and eastern Kalimantan.

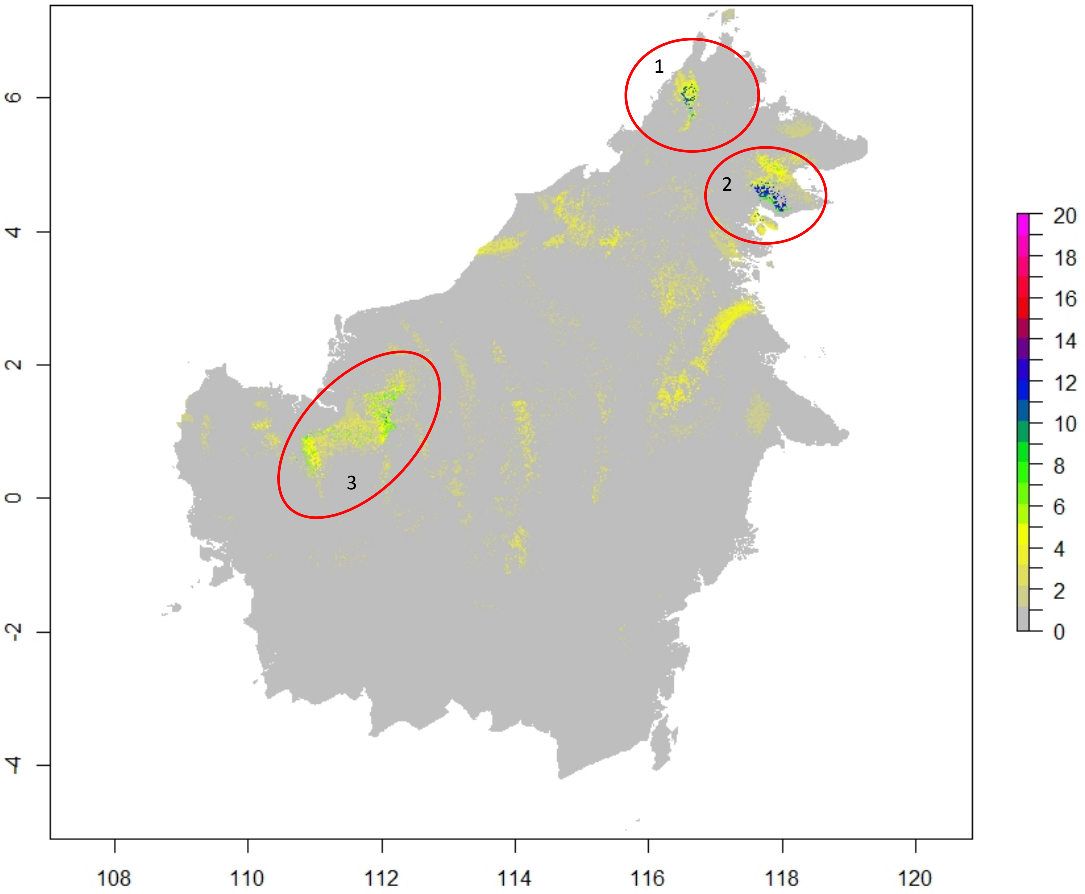

Figure 20: Predicted Begonia hotspots using SPE method for all Begonia species with less than five occurrences. Map showing the hotspots (1) Kinabalu, (2) Tawau hills and (3) Batang Ai and northwest of Sintang district by indicating the suitability of occurrence from low (grey) to high (magenta).

3.4 Synthetic analysis

3.4.1 Stacked – *Begonia* hotspot map in Borneo

Combined richness map of *Begonia* distribution (Figure 21) has clearly spotted *Begonia* hotspots in Borneo. Overall highest score is given to Tawau hills in Sabah and the second highest is the Mt. Kinabalu. Following that, Danun valley forest, Kubah national park, Batang Ai national park, northern-Sintang district and Pulong Tau forest in Sarawak, Labu forest, Labi forest, Batu Apoi and Ulu Temburong national parks in Brunei and Kayan Mentarang forests and east of Bulongan district in northern Kalimantan are also showing a higher *Begonia* richness. However, map also predicts Balikpapan district in eastern Kalimantan with significantly high score. Western, central and most of the parts of eastern Kalimantan areas giving the lowest probability of *Begonia* occurrence.

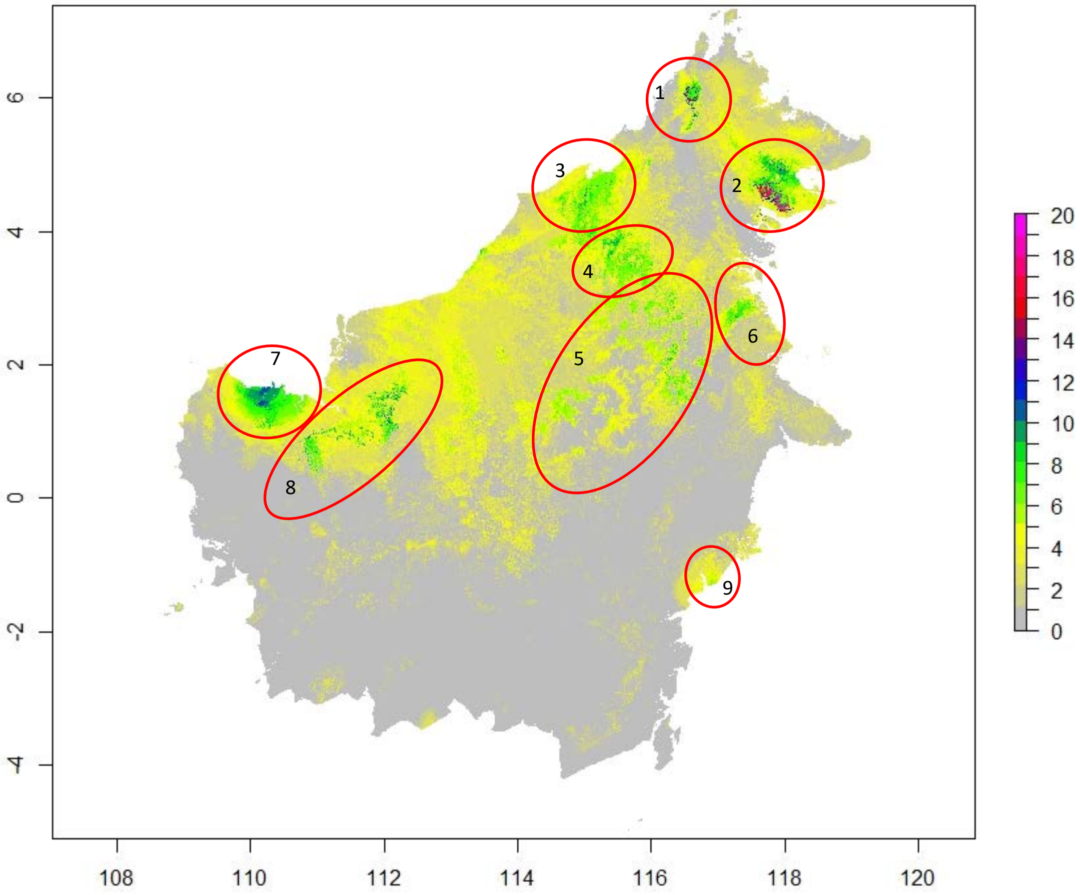

Figure 21: Stacked-Begonia distribution map for all Bornean Begonia species obtained summing up both the MaxEnt predicted distribution and the SPE predicted distribution. Map showing the combination of hotspots predicted from two different approaches (1) Kinabalu, (2) Tawau hills-Danum valley, (3) Brunei, (4) Pulong Tau and north-Sintang district, (5) Kayan Mentarang range, (6) East of Bulongan district, (7) Kubah NP, (8)Batang Ai- northwest of Sintang district and (9) near Balikpapan bay where indicating the suitability from low (grey) to high (magenta).

3.4.2 MaxEnt prediction vs Single Point Extrapolation

Both Maxent and SPE models (Figure 22) showing mostly similar prediction for the distribution of *B. calcarea* across Borneo. Maxent gives some wider prediction specially in

Sibu, Selangau, Tatau districts in Srawak. SPE prediction showing much restricted spread of the species which seems to be more specific when compare to the Maxent prediction. However, both maps predicting the hotspots like Batang Ai forest reserve which really important for *Begonia* ecology.

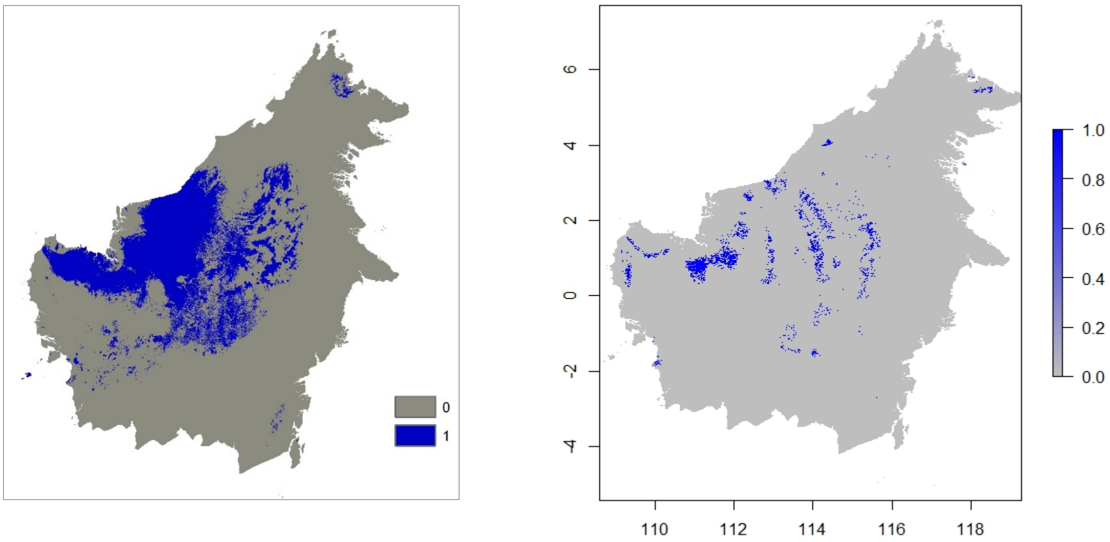

Figure 22: Maxent vs SPE binary prediction for the distribution of B. calcarea in Borneo. Map (a) showing the predicted distribution by MaxEnt while map (b) showing the distribution obtained frm SPE method(a) Maxent, (b) SPE prediction, representing presence (blue) and absence (grey) of Begonia.

CHAPTER 4: DISCUSSION

This study attempted to identify the areas richest in *Begonia* diversity in Borneo island. This was done by modelling the *Begonia* distribution in Borneo using species occurrence data recorded in Hughes et al. (2015). Since a majority of the species have less than five geo-referenced collection records, modelling was done using two separate methods as discussed in Chapter two and stacking 154 individual species distribution models in order to produce the final *Begonia* distribution map of Borneo.

4.1 What are the *Begonia* hotspots in Borneo?

According to the stacked *Begonia* distribution map (Figure 21), many *Begonia* rich localities could be identified. Some of them are very rich in predicted *Begonia* occurrence, called as "*Begonia* hotspots" in this study. In general, these hotspots are located among all the main political regions; Brunei, Sabah, Sarawak and Kalimantan; northern Borneo, Sabah is home to the key *Begonia* hotspots, Kinabalu and Tawau hills.

Mount Kinabalu, the highest peak of Sundaland, is a well-known biodiversity hotspot with ca. 5000 vascular plants in 1200 km^2 (Beaman, 2005) and is one of the most well-studied regions in Borneo. Present results show that Kinabalu is a key hotspot for *Begonia* diversity too (Figure 23a), finding many micro endemic *Begonia* species like *B. minutiflora*, *B. ramlanii*, *B. adliniana*, *B. gibbsiae* and *B. humericola* which have been found only in either one or two localities, and also species like *B. fuscisetosa* with fairly a wide range growing on the site. Most species which grow on Kinabalu enjoy lower montane habitat ranging from 1000 m to 2000 m elevation gradient with preferably 2000-2300 mm precipitation and 10 ^{o}C to 24 ^{o}C mean annual temperature while only one collection has been made (*B. fuscisetosa*) from out of those boundaries (at 183 m elevation with 26 ^{o}C temperature). Besides, many other narrow endemics including *B. chongii*, *B. diwolii*, *B. inostegia*, *B. imbricata*, *B. isabellina* and *B. minutiflora* predicted to grow in Mt. Kinabalu and surrounding area who prefers the same ecological conditions where *B. isabelline* is known only from Kalimantan up to now. Our results confirm the findings of many *Begonia* studies who also referred the greatest *Begonia* richness in Mt. Kinabalu (Beaman, Anderson and Beaman, 2001; Kiew *et al.*, 2015; Repin, Karim and Mujih, 2015) while Repin, Karim and Mujih (2015) further found that some begonias (*B. adliniana*, *B. ramlanii* and *B. vanderentii*) are endemic on ultramafic formations around Kinabalu, which are renowned as hotspots for plant diversity and endemism. Furthermore, Merckx *et al.* (2015)

state that Kinabalu's biota is younger than the mountain origin (six million years). Kinabalu endemics would have not evolved by niche shifts from local ancestors, but rather by immigration of biota from somewhere else. This is supposed because of its rapid emergence and unique alpine summit conditions while climate changes may have restricted the geographical range over past.

Tawau hills and the surrounding area (Figure 23b) is the largest *Begonia* hotspot predicted by our results where home to narrowly endemic *B. paracauliflora*, *B. tawaensis*, *B. gelasensis* and also for *B. fuscisetosa*. Unlike Kinabalu begonias, these species preferred lower lying dipterocarp forests located between 250-850 m elevational gradient, much warmer climate (Annual mean temperature 22 ^{o}C – 25 ^{o}C) and 2000-2200 mm annual rainfall. Tawau soils are comprised of alluvial volcanic soils which are rich in nutrient content while there are also some patches of *kerangas* where begonias do not grow (Repin, Sampang and Limbawang, 2015). Narrow endemism in the Tawau hills has been reported by Repin, Sampang and Limbawang (2015), stating that many of the begonias in the area have restricted distributions and most of them known only from single population while a large number of begonias are found in a small part of Tawau hills. Seino, T., Kanehiro and Lakim (2007) mentioned the Tawau hill soils are rich in soil nutrients due to accumulation of volcanic ash and therefore, there are significant differences in floristic composition to the rest of Sabah which might also be the reason for high *Begonia* richness.

Apart from the two main hotspots, four other *Begonia* rich areas can be identified in Sabah; Danum valley and Madai forest, Trus Madi forest, Tangkulap-Pinangab-Deramakot forests and southwest and northeastern parts of Crocker range where Danum valley- Madai forest and Trus Madi forests are predicting higher *Begonia* richness. Danum valley is a lower laying primary dipterocarp forest located in between 100-240 m. begonias which prefer much warmer (26-27 ^{o}C annual mean temperature), moist climates like *B. incompta*, *B. vespropinqua*, *B. baturongensis* and *B. madaiensis* have been recorded in this forest. Danum valleys' high *Begonia* richness has previously recorded by Yun *et al.* in (2015) who encountered 15 *Begonia* species within three days confirming the extremely high *Begonia* richness on the site. Most importantly Danum valley is rich in limestone which is a home for 17 *Begonia* species. Additionally, the absence of dry periods except occasionally E1 Nino years and receiving high level of rain are the reasons for this high *Begonia* richness. It is also suggested this could be due to the high soil nutrient content and water retention capacity discussed in Hazebroek, Adlin and Sinun (2012). A further reason for this diversity might be due to the species attracting different pollinators by their morphologically distinct flowers which discorages hybridization, allowing them to co-exist in the same habitat (Yun *et al.*, 2015). Tangkulap-Pinangab-Deramakot forests area is between Danum valley and Mt. Kinabalu, sharing rather similar climatic conditions but much higher rainfall (2500-2600 mm) than the Danum valley hotspot. Though its highly predicted for high *Begonia* diversity, there is only one *Begonia* collection record from this area where *B. bengohensis* has been collected. However, the modelled distribution predicts that the species like *B. murudensis*, *B. gibbsiae*, *B. gusilii*, *B. postari* prefer same ecological conditions and possibly grow in that area too. Southwest of Crocker Range is the next predicted *Begonia* hotspot located in Sabah where narrowly endemic *B. tamaniensis* and *B. kinahimiae* have been collected. These forests are rich in dipterocarps and lay between 100-240 m elevation where annual mean temperature 26 ^{o}C and 1900-2000 mm annual rainfall is received.

As a summary, Sabah has a total of 82 *Begonia* species while 42 of them have less than five known localities suggesting that northern Borneo has very high *Begonia* diversity while many species have very restricted range and narrow endemism. This is also supported by Raes *et al.* (2009), stating that the northern Borneo has the highest botanical richness while the northern Crocker range with mount Kinabalu has the highest endemism; hence my results confirm the suggestion that northern Borneo is the centre of vascular plant diversity in southeast Asian region (Mutke and Barthlott, 2005).

Holding the highest species number in Borneo, a quite large number of *Begonia* hotspots has also been predicted in Sarawak when compared to the other regions in Borneo. In particular, Kubah National Park and the surrounding area (Figure 23e) commonly known as Matang or Gunung Matang among biologists, covering the Lundu, Bau, and Kuching districts and northern-Serian and western-Simunjan districts shows the highest *Begonia* richness in the region. Although it is widely spread among several district, it shares rather similar climatic and altitudinal conditions where begonias found at 20-640 m elevation, 23-27 ^{o}C annual mean temperature with 3500-4300 mm precipitation. The hilly terrain and heavily forested slopes and ridges dominated with lowland dipterocarp forests making ideal conditions for *Begonia* growth. Species recorded consists of several widespread species whilst a large number of narrow endemic species like *B. sarawakensis*, *B. corrugata*, *B. baik*, *B. sadirensis*, and *B. cincinnifera* have been found in Kuching. In addition, *B. hulleti* has only found in Matang range-Kubah National Park. Bau is the place where *B. oblongifolia* has been found which is primarily a limestone habitat which could be the main driving factor for the *Begonia* richness on the site.

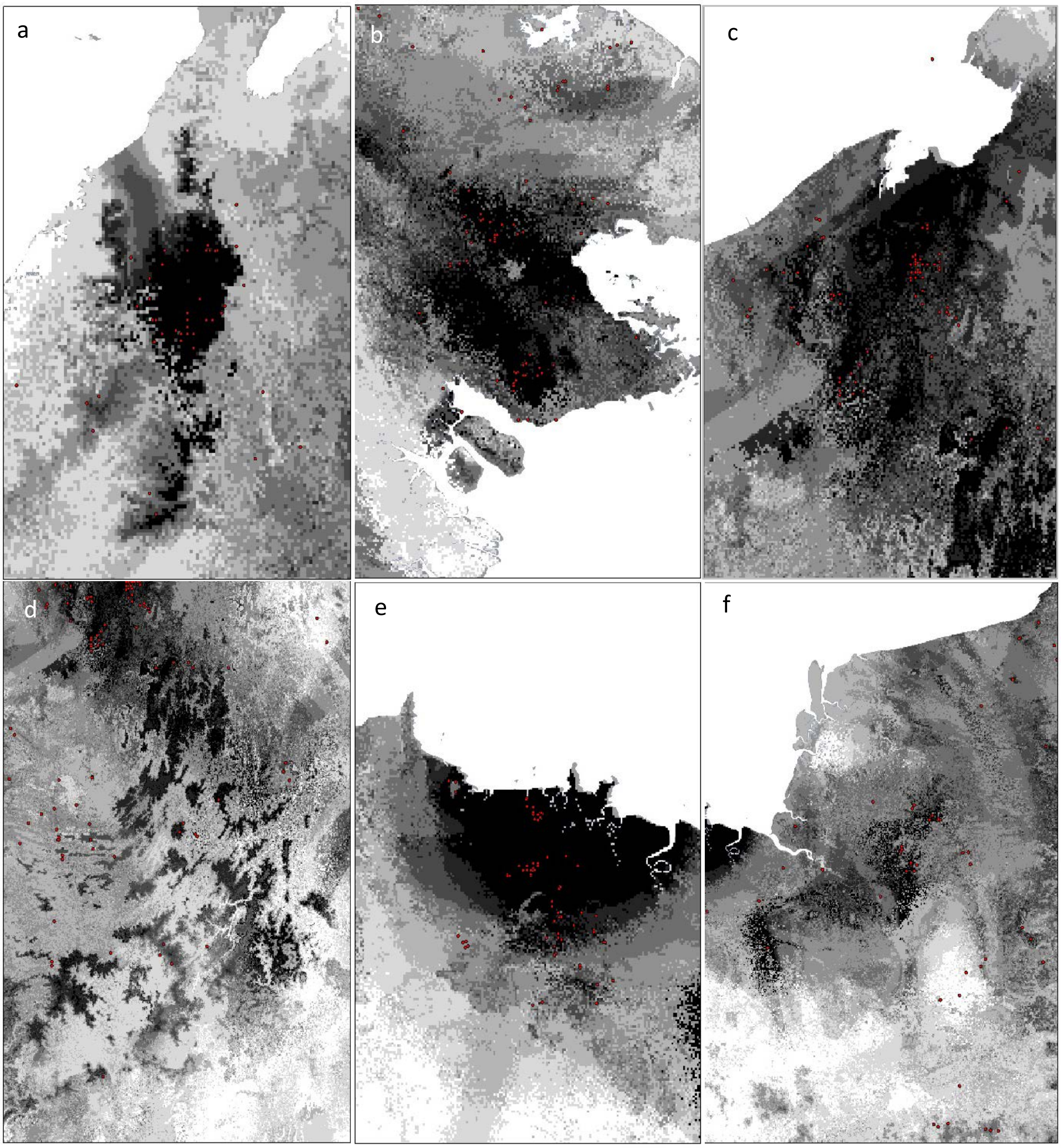

Figure 31: Major hotspots predicted by stacked- Begonia distribution model. (a) Mount Kinabalu and surrounding area, (b) Tawau Hills and Danum valley conservation area, (c) Brunei and Northern Sarawak, (d) Kayan-central mountain range in Kalimantan, (e) Kubah-Matang range and (f) Batang Ai and surrounding forests.

Batang Ai forest together with northwest-Sintang and northeast-Sanggau districts (Figure 23f) is the next important hotspot predicted in Sarawak which also harbours many range-restricted begonias which have been found as only one or two collections (*B. addrinii, B. celata, B. bayae, B. edgariana* and *B. tebiang*) while also home to some widespread species like *B. calcarea* and *B. acidulenta*. The forest is dominant with primary mixed dipterocarp forest and riparian forests with average 25 ^{o}C annual mean temperature and 3000 mm rainfall which provides better climatic conditions for *Begonia* growth. In addition to that, few other important *Begonia* hotspots have predicted by our distribution models. Those are the basin of Sungai Tutoh, Mt. Mulu and adjacent forests, Long Seridan-Sungai Magoh basin, Mt. Murud-Pulong Tau forest in northern Sarawak. Of these, there are no collection records from Sungai Tutoh basin and Long Seridan-Sungai Magoh basin, while the following species are predicted: *B. yii*, *B. linauensis, B. magentifolia* and *B. linauensis, B xhiphophylloides* respectively. Almost all the species grow in lowland forested habitats with nearly the same climate (24-27 ^{o}C annual mean temperature and 3000-3700 mm rainfall) whereas some species grow in Mt. Murud-Pulong Tau forest and enjoy some upper montane habitat (1500-2000 m) with cooler climates.

Sang *et al.* (2015) stated that Brunei is poor in *Begonia* diversity when compare to Sabah and Sarawak. One reason she pointed out was the absence of limestone rich habitats where usually the home to *Begonia* rich flora. She further mentioned that, habitats like *kerangas*, swamp forests and shale are not favourable for *Begonia* growth while those on shale are completely devoid by begonias. This confirms the predictions of botanical richness pattern of Raes *et al.* (2009) who also found some lower species richness in Brunei habitats. However, our predictions showing that Brunei is rich in *Begonia* diversity (Figure 23c) specially, Ulu Temburong-Paradayan, Labu forest, Labi forest along with Sungai Belait are is rich in Brunei endemics like *B. baramensis, B. temburogensis, B. retakensis* and *B. stenogyna* even though it shows lower prediction compared to the surrounding forests.

Current records showing that Indonesian Kalimantan, the largest region of Borneo holds the lowest *Begonia* density of the island (Hughes et al., 2015). It has suggested that this is because of the dominant vegetation types like heath forests, swamp forests where there is low in species diversity compared to lowland dipterocarp forests (Raes *et al.*, 2009). However, we found that northern and eastern Kalimantan (Figure 23d) are potentially rich in *Begonia* diversity mainly in the Kayan Mentarang national park, the central mountain range located at the borders of Malaysian-Indonesian regions and nearest scattered hills in northern Kalimantan. Considering the ecological factors, predicted areas are located between 1000-1500 m elevation with 18-23 °C annual mean temperature, receiving 2800-3500 mm annual rainfall. According to the available data, there are no collections made in the Kayan range whilst only *B. gueritziana* and a few unidentified species (Morphospecies, Chapter 2.1.2) recorded in adjacent scattered hills. However, *B. hirtitepala*, *B. julaihiana* and *B. melianuensis* are predicted to be there based on their SPE model. Similarly, the lowland of the Sungai Kayan basin in the east of northern Kalimantan and west-Berau and northwest-Kuala Timur districts in eastern Kalimantan also have predicted high *Begonia* richness, with *B. melianuensis*, *B. muluensis*, *B. murudensis*, *B. vulgare* and *B. xiphophylloides* modelled as present whereas there are no current species occurrence records.

Furthermore, our results also predicting a high suitability at the north of Balikpapan bay in the coast of eastern Kalimantan. The species like *B. conipila*, *B. cyanescense*, *B. stynogyna* and *B. fuscicetosa* have predicted for the area while many of them are Brunei endemics. The area predicted located near to the coast at less than 60 m elevation with much warmer climate (26-27 °C annual mean temperature) and 2500 mm average rainfall presenting mostly a similar ecology to Brunei habitats. Therefore, this might be the reason that Brunei endemics prefer to grow on the site even though currently it is fully urbanized and none of the begonias can expect. However, supporting to the our results, the similar area was predicted for high botanical richness by Raes *et al.* in 2009.

Additionally, it is identifiable that some of the areas have high number of single point endemics compare to other hotspots. For example, as mentioned earlier Batang Ai is home to the species like *B. addrinii*, *B. celata*, *B. tebiang*, *B. bayae*, *B. edgariana*, *B. acidulenta* and *B. hirtitepala* which has been found only single or two localities which might be due to lot of recent explorations of these areas compare to other places explored earlier (i.e., Kuching area).

4.2 Are limestones the only driving factor for *Begonia* distribution?

According to our observations most of the major predicted *Begonia* hotspots (Kinabalu, Tawau hills, Danum valley and Batang Ai) are not laying on limestone rich areas (Figure 24b). Only those at the north-Serian district and Kuching area in part near Kubah national park, the Danum valley forests, the Labi forest reserve in Brunei, a small area of the Lahad Datu district near Tawau hills and Bulongan district in eastern Kalimantan are located on limestone rich habitat.

None of the predicted areas in central Borneo is not limestone rich. Therefore, our results show that presence of limestone is not the only factor driving *Begonia* species richness. However, the map of distribution of limestone we used for our analysis is quite generalized and a more detailed map will be useful for further conclusions.

According to the factors discussed before, environmental variables like warm, humid environments, nutrient rich soils and variety of geological formations strongly influence the distribution and diversity of begonias in Borneo in addition to the presence of limestone. This might be further affected by geological history particularly the age of each region which might influence the evolution of *Begonia*. Further, Syfer*t et al,* (2014) states that species with small and narrow ranges are common at mid to high elevations while widespread species are more common at low elevations, indicating the altitude also an important factor for a successful establishment for begonias which will depends on the species interest. Our results also found many of the narrow endemic begonias prefer mid to high elevations while wide range species like *B. fuscicetosa* is generally growing on lower elevations. However, we also found some narrow range species like *B. paoana* growing between 40-70 m elevations northwest-Sitang and northeast-Sanggau areas near Batang Ai forests while most of the begonias growing in Borneo prefer steep slopes rather than growing on flat lands signifying topography is also more influential in *Begonia* diversity.

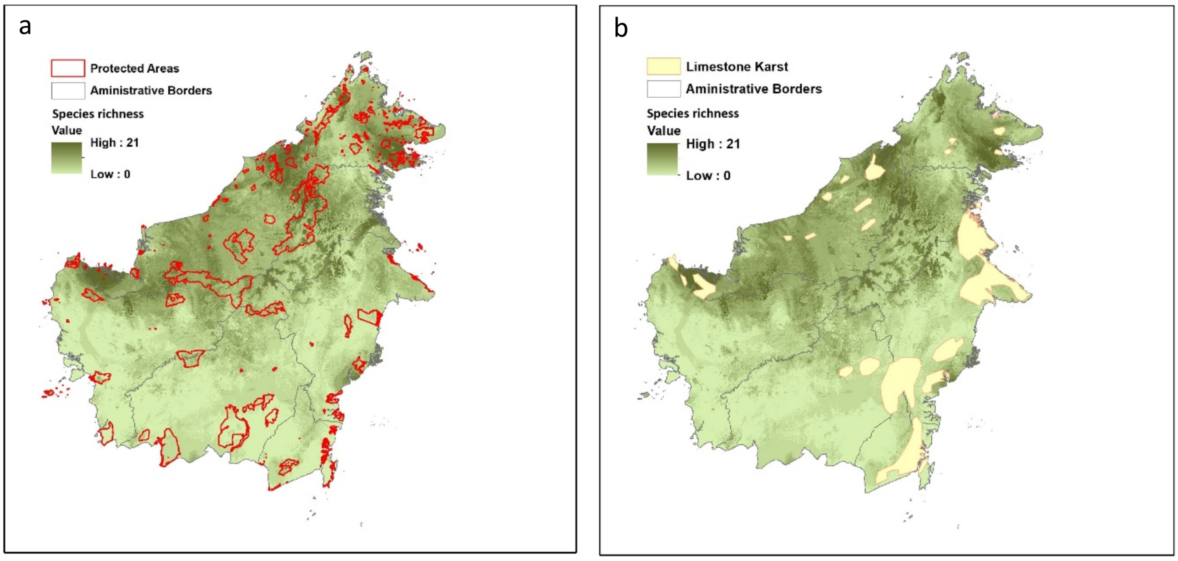

Figure 39: Distribution of (a) Protected areas and (b) Limestone across Borneo.

4.3 How protected they are?

In generally, narrow endemics and rare species are at high risk in extinction when they are not adequately protected by protected areas. According to the findings of this study, only some of the predicted *Begonia* hotspots are located inside forest reserves (Figure 24a). The places where highest *Begonia* richness in Sabah is located inside of Kinabalu and Tawau hills national parks while much of the area is still outside of the Tawau hills NP. However, the hotspot at Kubah National park is not fully protected and most of the area lying outside of the boundary. Similarly, much of the area of the hotspot near Batang Ai national park is located outside of the national park while the area predicted in upper-Sintang district and the lower part of Sungai Tutoh are fully lying outside of reserves. Further, the Brunei endemics are mostly protected by Labu, Labi, Batu Apoi and Ulu Temburong national parks while most of the high diversity areas in northern Kalimantan are mostly protected by the Kayan Mentarang forests reserve. However, the areas lying at the bottom of Malinau district, east of Bulongan district in northern Kalimantan and upper part of the Kutai Timur district in eastern Kalimantan are not protected. Furthermore, Balikpapan district in eastern Kalimantan with a significantly high diversity is highly urbanized and is not protected at all.

4.4 Where and why future collections should be done?

Sang, Kiew and Yiing (2015) states that begonias are highly restricted to primary forests and are highly vulnerable to disturbances. Further, these species can be easily destroyed by opening the canopy as well as flooding or silting the areas where they occur. The author mentioned that the begonias in the forest Bukit Setiam, in Tatau district, is only surviving on several vertical rock surfaces due to extreme use of land for oil palm cultivation and logging. The further, they emphasize the importance of demarcating wildlife corridors such as Bukit Mina wildlife corridor area in Tatau, where there are many begonias surviving.

As discussed above, many of the *Begonia* rich areas predicted by our results are underexplored yet (Figure 25), while some of the places are highly explored recently like the hotspots predicted in northern and southwestern Borneo where the collection densities are very high. In some of the places high in collection density, only a few species have been found, leading to a low *Begonia* richness. High collection density is inevitable in areas that collectors are more likely drive through and are easily accessible, however areas which are hard to access like the mountainous areas in north and east Kalimantan are predicted to have a high Begonia richness. Collection efforts are biased when it comes to determining whether the most of Bornean

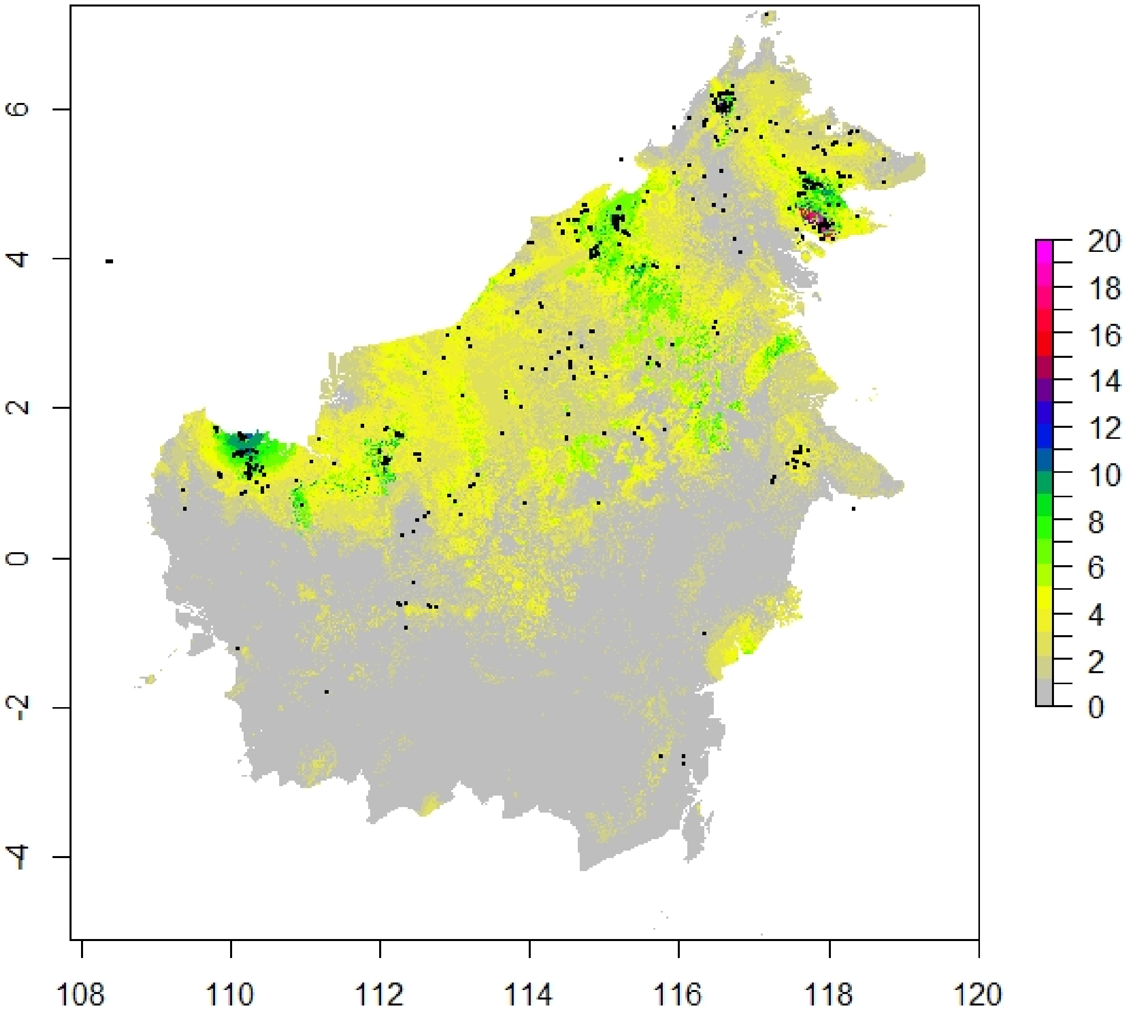

Figure 40: Map of variance between predicted Begonia hotspots and the actual occurence records in Borneo. Predicted Begonia distribution showing the suitability of occurence from low (grey) to high (magenta) while black dots indicate the geo-referenced Begonia collection records published in Hughes et al., (2015).

begonias are truly narrow endemics. However, species distribution modelling techniques allow us to predict geographic distributions of species even for areas where no collection has been made by overcoming incomplete and biased sampling ((Raes *et al.*, 2009).

4.6 MaxEnt vs SPE predicted hotspots

It was noticeable that the predicted distributions of narrowly distributed begonias are different when using all the bioclimatic layers than with only selected bioclimatic layers in the SPE method. The reason for this difference might be because not all the bioclimatic variables are

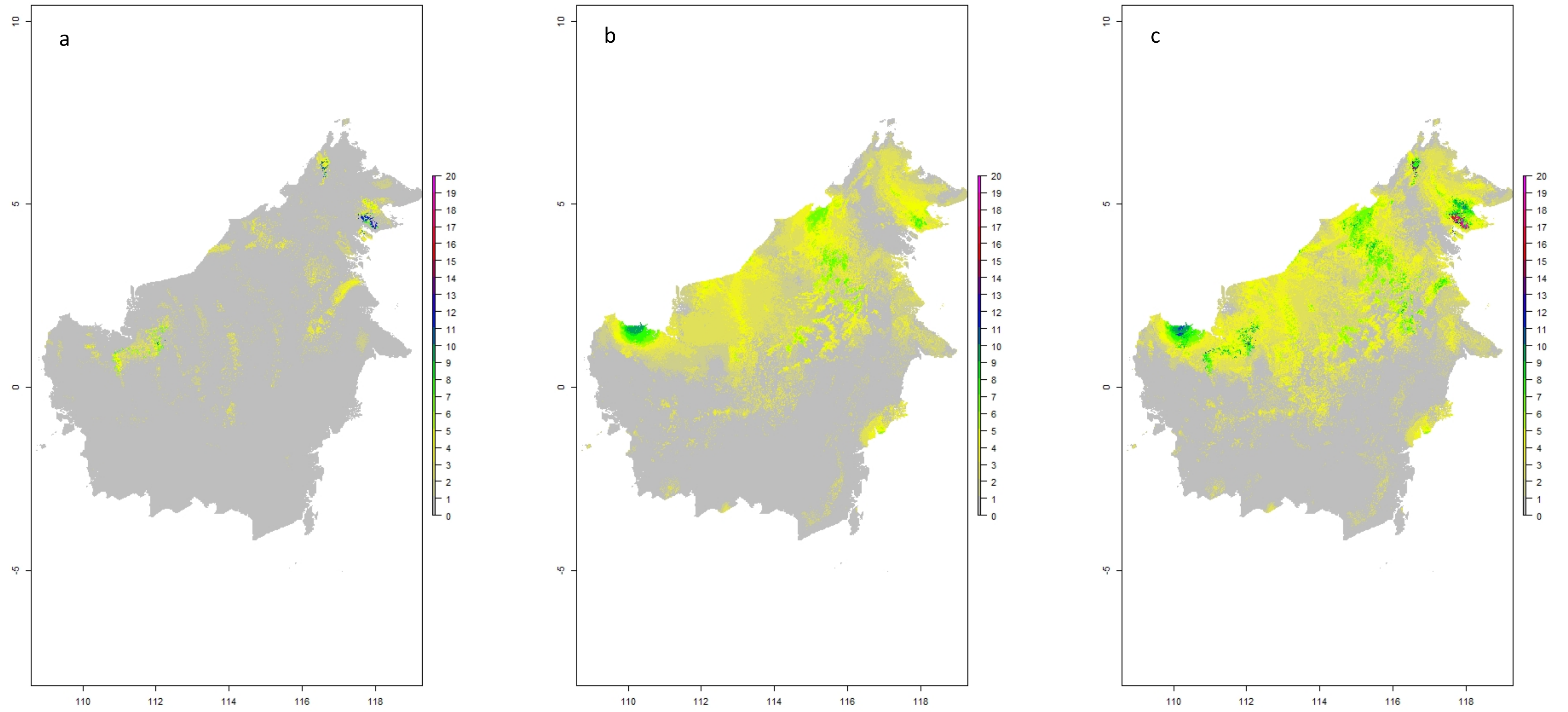

Figure 45: Comparison of predicted distribution of Begonia richness in Borneo using (a) SPE method, (b) MaxEnt and (c) Combined map obtained by summing the SPE and MaxEnt predictions. Maps showing the suitability of occurrence from low (grey) to high (magenta).

important for *Begonia* distribution, even though all of them are contributed some score for the model. Some of the variables are highly correlated with other factors and therefore, the predicted maps using all layers might not be biologically realistic with respect to *Begonia* distribution. When using the selected variables, only those which are highly important for *Begonia* distribution contributed to the model building; hence the predictions were much more realistic with narrow endemics. However, using all layers is not a problem with MaxEnt models because MaxEnt down weights correlated bioclimatic variables and hence, only the most suitable variables are contributing to the model.

According to our results, MaxEnt couldn't identify all *Begonia* hotspots in Borneo (Figure 26b), even the very rich biodiversity hotspot in Mt. Kinabalu. Although many studies have proven that MaxEnt can cope with small sample sizes (Wisz *et al.*, 2008; Fourcade *et al.*, 2014, van Proosdij *et al.*, 2016), our results showed that MaxEnt is not very effective in predicting hotspots with narrow endemics restricted to one or two localities. However, I have overcome this gap by developing the SPE approach which I used to predict the distribution of 131 narrow ranged species and incorporated them into the final map (Figure 26c). Importantly, the highest *Begonia* richness predicted in Mt. Kinabalu and the Tawau hills and adjacent forests were predicted by the SPE method (Figure 26a) and also contributed much to the predicted high *Begonia* richness in Batang Ai and nearest forests in the final map. The hotspot predicted in the Matang range and Kubah forests was predicted by MaxEnt through models of densely collected species. Therefore, in conclusion, using the combined approach of MaxEnt and SPE is given new insights of distribution diversity in this genus of micro endemics.

CONCLUSIONS

According to our predictions, *Begonia* hotspots are distributed in all regions in Borneo; Sabah, Sarawak, Kalimantan and Brunei while Sabah is home to the richest *Begonia* hotspots of Mt. Kinabalu and the Tawau hills.

In addition to the main *Begonia* hotspots,

i. Four other *Begonia* rich areas can be identified in Sabah; Danum valley and Madai forest, Trus Madi forest, Tangkulap-Pinangab-Deramakot forests and southwest and northeastern parts of Crocker range.

ii. Kubah National Park and the surrounding area, Batang Ai forest together with northwest-Sintang and northeast-Sanggau districts, Basin of the Sungai Tutoh, Mt. Mulu and adjacent forests, Long Seridan-Sungai Magoh basin and Mt. Murud-Pulong Tau forest are the predicted hotspots in Sarawak.

iii. In Brunei, Ulu Temburong-Paradayan, Labu forest, Labi forest along with Sungai Belait have predicted high *Begonia* richness.

iv. Kayan Mentarang national park, the Sungai Kayan basin, west-Berau and northwest-Kuala Timur districts and north of Balikpapan bay have predicted high begonia richness in Kalimantan.

Some of the predicted *Begonia* hotspots are located in limestone rich habitats. However, we found that limestone is not the only factor for begonia distribution, but also warm, humid environments, nutrient rich volcanic soils and a variety of geological formations also strongly influence the distribution and diversity of begonias in Borneo.

MaxEnt itself couldn’t identify all *Begonia* hotspots, however, Maxent-SPE combined approach overcome this gap and is a promising method to model species distributions for narrow endemics.

REFERENCES

Abolmaali, S. M. R., Tarkesh, M. and Bashari, H. (2018) 'MaxEnt modeling for predicting suitable habitats and identifying the effects of climate change on a threatened species, Daphne mucronata, in central Iran', *Ecological Informatics*. Elsevier, 43(October 2017), pp. 116–123. doi: 10.1016/j.ecoinf.2017.10.002.

Ahmad, J. A. *et al.* (2015) 'The Begoniaceae of Brunei Darussalam, Borneo, including Two New Species', *Sandakania*, 20, pp. 7–50.

Ashton, P. S. (1992) 'Plant conservation in the Malaysian region', in *Proceedings of International Conference of Tropical Biodiversity 'In harmony with nature'*, pp. 86–93.

Bathlott, W. and Rafiqpoor, D. (2005) 'Global Centers of Vascular Plant Diversity', *Nova Acta Leopoldina*, 92(342), pp. 61–83.

Beaman, J. H. (2005) 'Mount Kinabalu: hotspot of plant diversity in Borneo.', *Biologiske Skrifter*, 55, pp. 103–127.

Beaman, J. H., Anderson, C. and Beaman, R. S. . (2001) 'The plants of Mount Kinabalu. 4. Dicotyledon families Acanthaceae to Lythraceae', *Natural History Publications (Borneo) & Royal Botanic Gardens, Kew*, p. 570.

Beck, S. V. *et al.* (2017) 'Plio-Pleistocene phylogeography of the Southeast Asian Blue Panchax killifish, Aplocheilus panchax', *PLoS ONE*, 12(7), pp. 1–17. doi: 10.1371/journal.pone.0179557.

Brambach, F. *et al.* (2017) 'Diversity, endemism, and composition of tropical mountain forest communities in Sulawesi, Indonesia, in relation to elevation and soil properties', *Perspectives in Plant Ecology, Evolution and Systematics*, 27, pp. 68–79. doi: 10.1016/j.ppees.2017.06.003.

Bruyn, M. de *et al.* (2014) 'Borneo and Indochina are Major Evolutionary Hotspots for Southeast Asian Biodiversity', *Systematic Biology*, 63(6), pp. 879–901. doi: 10.1093/sysbio/syu047.

CAIMS (2005) *General Geology of Sabah*. Available at: http://ww2.sabah.gov.my/htan_caims/Other Info/geology.htm (Accessed: 17 August 2019).

Cannon, C. H. and Leighton, M. (2019) 'Tree species distributions across five habitats in a Bornean rain forest', *Journal of Vegetation Science*, 15(2), pp. 257–266.

Cannon, C. H., Morley, R. J. and Bush, A. B. G. (2009) 'The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance', in May, R. (ed.) *Proceedings of the National Academy of Sciences USA*, pp. 106 (27) 11188-11193. doi: https://doi org.ezproxy.is.ed.ac.uk/10.1073/pnas.0809865106.

Collenette, P. (1964) 'A short account of the geology and geological history of Mt Kinabalu', *Proceedings of the Royal Society of London. Series B. Biological Sciences*. The Royal Society London, 161(982), pp. 56–63. doi: 10.1098/rspb.1964.0076.

Crayn, D. M., Costion, C. and Harrington, M. G. (2015) ‘The Sahul-Sunda floristic exchange: Dated molecular phylogenies document Cenozoic intercontinental dispersal dynamics’, *Journal of Biogeography*, 42(1), pp. 11–24. doi: 10.1111/jbi.12405.

Dewitte, A. *et al.* (2011) ‘The Origin of Diversity in Begonia: Genome dynamism, population processes and phylogenetic patterns.’, in Grillo, O. and Venora, G. (eds) *The Dynamical Processes of Biodiversity - Case Studies of Evolution and Spatial Distribution*. InTech, pp. 27–52. doi: 10.5772/23789.

Elith, J. *et al.* (2006) ‘Novel methods improve predition of species’ distributions from occurrence data’, *Ecography*, 29, pp. 129–151.

Elith, J. *et al.* (2011) ‘A statistical explanation of MaxEnt for ecologists’, *Diversity and Distributions*, 17(1), pp. 43–57. doi: 10.1111/j.1472-4642.2010.00725.x.

Escalante, T. *et al.* (2013) *Identification of areas of endemism from species distribution models: Threshold selection and Nearctic mammals*, *Revista Especializada en Ciencias Químico-Biológicas*. Available at: http://www.gbif.org/.

ESRI (2017) ArcMap 10.1 for Desktop. Environmental Research Institute, Redlands, California, USA.

Ficetola, G. F., Thuiller, W. and Miaud, C. (2007) ‘Prediction and validation of the potential global distribution of a problematic alien invasive species - The American bullfrog’, *Diversity and Distributions*, 13(4), pp. 476–485. doi: 10.1111/j.1472-4642.2007.00377.x.

Fick, S. E. and Hijmans, R. J. (2017) *Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas*, *International Journal of Climatology*. Available at: http://worldclim.org/version2 (Accessed: 1 July 2019).

Fourcade, Y. *et al.* (2014) ‘Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias’, *PLoS ONE*, 9(5), pp. 1–13. doi: 10.1371/journal.pone.0097122.

Goodall-Copestake, W. P. *et al.* (2010) ‘The early evolution of the mega-diverse genus Begonia (Begoniaceae) inferred from organelle DNA phylogenies’, *Biological Journal of the Linnean Society*, 101(2), pp. 243–250. doi: 10.1111/j.1095-8312.2010.01489.x.

Grytnes, J. A. *et al.* (2006) ‘Elevational species richness patterns for vascular plants on Mount Kinabalu , Borneo’, *Journal of Biogeography*, 33, pp. 1838–1849. doi: 10.1111/j.1365-2699.2006.01554.x.

Hall, R. (1995) ‘Plate Tectonic Reconstructions of the Indonesian Region’, in *Proceedings Indonesian Petroleum Association*, pp. 71–82.

Hall, R. (2002) ‘Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: Computer-based reconstructions, model and animations’, *Journal of Asian Earth Sciences*, 20(4), pp. 353–431. doi: 10.1016/S1367-9120(01)00069-4.

Hall, R. (2009) ‘Southeast Asia’s changing palaeogeography’, *Blumea: Journal of Plant Taxonomy and Plant Geography*, 54(1–3), pp. 148–161. doi: 10.3767/000651909X475941.

Hall, R. (2011) ‘Australia–SE Asia collision: plate tectonics and crustal flow’, *Geological Society, London, Special Publications*, 355(1), pp. 75–109. doi: 10.1144/sp355.5.

Hall, R. (2012a) ‘Late Jurassic-Cenozoic reconstructions of the Indonesian region and the Indian Ocean’, *Tectonophysics*. Robert Hall, 570–571, pp. 1–41. doi: 10.1016/j.tecto.2012.04.021.

Hall, R. (2012b) ‘Sundaland and Wallacea: geology, plate tectonics and palaeogeography’, in David Gower, Kenneth Johnson, James Richardson, Brian Rosen, Lukas Rüber, S. W. (ed.) *Biotic Evolution and Environmental Change in Southeast Asia*. 1st edn. Cambridge: Cambridge University Press, pp. 32–78. doi: 10.1017/cbo9780511735882.

Hall, R. *et al.* (2013) *The Geology of Mountain Kinabalu*.

Hall, R. (2013) ‘The palaeogeography of Sundaland and Wallacea since the Late Jurassic co m er c ia l us e on on co m m er c ia l us e’, 72, pp. 1–17. doi: 10.4081/jlimnol.2013.s2.e1.

Hamilton, W. (1973) ‘Tectonics of the Indonesian Region’, *Geological Society of Malaysia, Bulletin*, 6(July), pp. 3–10. doi: 10.1016/0003-6870(73)90259-7.

Harley, M. M. and Morley, R. J. (1995) ‘Ultrastructural studies of some fossil and extant palm pollen , and the reconstruction of the biogeographical history of subtribes Iguanurinae and Calaminae’, *Review of Palaeobotany and Palynology*, 85, pp. 153–182.

Hazebroek, H. P., Adlin, T. Z. and Sinun, W. (2012) *Danum Valley: The Rain Forest*. Kota Kinabalu, Malaysia: Natural History Publications (Borneo).

Hernandez, P. A., Master, C. H. G. L. L. and Albert, D. L. (2006) ‘The effect of sample size and species characteristics on performance of different species distribution modeling methods’, *Ecography*, 29, pp. 773–785. doi: 10.1021/jp048527b.

Hughes, M. (2002) *Population structure and speciation in Begonia L . A thesis submitted to the University of Glasgow*. University of Glasgow.

Hughes, M. *et al.* (2015) ‘Begonia Resource Centre’. Online database available at: http://padme.rbge.org.uk/begonia.

Hughes, M., Girmansyah, D. and Ardi, W. H. (2015) ‘Further discoveries in the ever-expanding genus Begonia (Begoniaceae): fifteen new species from Sumatra’, *European Journal of Taxonomy*, 167(Hughes 2008), pp. 1–40.

Hughes, M. and Hollingsworth, P. M. (2008) ‘Population genetic divergence corresponds with species-level biodiversity patterns in the large genus Begonia’, *Molecular Ecology*. John Wiley & Sons, Ltd (10.1111), 17(11), pp. 2643–2651. doi: 10.1111/j.1365-294X.2008.03788.x.

IUCN 2019 (2019) The IUCN Red List of Threatened Species Version 2019-2.

Kessler, M. (2001) ‘Patterns of diversity and range size of selected plant groups along an elevational transect in the Bolivian Andes’, pp. 1897–1921.

Kiew, R. *et al.* (2015) *A Guide to Begonias of Borneo*. 1st edn. Kota Kinabalu: Natural History Publications (Borneo).

De Klerk, H. M. *et al.* (2002) ‘Patterns of species richness and narrow endemism of terrestrial bird species in the Afrotropical region’, *Journal of Zoology*, 256(3), pp. 327–342. doi: 10.1017/S0952836902000365.

Kreft, H. and Jetz, W. (2007) ‘Global patterns and determinants of vascular plant diversity’, in *Proceedings of the National Academy of Sciences*, pp. 5925–5930.

Lin, C., Chung, S. and Peng, C. (2014) ‘Three new species of Begonia (sect. Petermannia, Begoniaceae) from Sarawak, Borneo’, *Phytochemistry*, 191(1), pp. 129–140.

Linnaeus, C. (1753) ‘Nepenthes’, *Species Plantarum*, 2, p. 955.

Liu, C. *et al.* (2005) ‘Selecting thresholds of occurrence in the prediction of species distributions Canran’, *Ecography*, 3(December 2004), pp. 385–393.

Liu, C., Newell, G. and White, M. (2016) ‘On the selection of thresholds for predicting species occurrence with presence-only data’, *Ecology and Evolution*, 6(1), pp. 337–348. doi: 10.1002/ece3.1878.

Liu, C., White, M. and Newell, G. (2013) ‘Selecting thresholds for the prediction of species occurrence with presence-only data’, *Journal of Biogeography*, 40(4), pp. 778–789. doi: 10.1111/jbi.12058.

Lohman, D. J. *et al.* (2011) ‘Biogeography of the Indo-Australian Archipelago’, *The anual review of ecology, evolution and systematics*, 42(2011), pp. 205–226.

MacKinnon, K. *et al.* (1997) *The Ecology of Kalimantan*. United Kingdom: Oxford University Press.

Merckx, V. S. F. T. *et al.* (2015) ‘Evolution of endemism on a young tropical mountain’, *Nature*, 524(7565), pp. 347–350. doi: 10.1038/nature14949.

Metcalfe, I. (2013) ‘Palaeozoic − Mesozoic history of SE Asia’, pp. 7–35. doi: 10.1144/SP355.2.

Moonlight, P. W. (2017) *Linking Speciation and the Niche : Taxonomy , Phylogeny and Niche Evolution in Neotropical Begonia*. University of Glasgow.

Moonlight, P. W. *et al.* (2018) ‘Dividing and conquering the fastest-growing genus: Towards a natural sectional classification of the mega-diverse genus Begonia (Begoniaceae)’, *Taxon*. John Wiley & Sons, Ltd, 67(2), pp. 267–323. doi: 10.12705/672.3.

Morales, N. S., Fernández, I. C. and Baca-González, V. (2017) ‘MaxEnt’s parameter configuration and small samples: are we paying attention to recommendations? A systematic review’, *PeerJ*, 5, p. e3093. doi: 10.7717/peerj.3093.

Morley, R. J. (2000) *Origin and evolution of tropical rain forests*. Chichester : Wiley.

Mutke, J. and Barthlott, W. (2005) ‘Patterns of vascular plant diversity at continental to global scales’, *Biol. Skr.*, 55, pp. 521–531.

Norris, D. (2014) ‘Model thresholds are more important than presence location type: Understanding the distribution of lowland tapir (Tapirus terrestris) in a continuous Atlantic forest of southeast Brazil’, *Tropical Conservation Science*, 7(3), pp. 529–547. doi: 10.1177/194008291400700311.

Papeş, M. and Gaubert, P. (2007) ‘Modelling ecological niches from low numbers of occurrences: Assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents’, *Diversity and Distributions*, 13(6), pp. 890–902. doi: 10.1111/j.1472-4642.2007.00392.x.

Pearson, R. G. *et al.* (2007) ‘Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar’, *Journal of Biogeography*, 34(1), pp. 102–117. doi: 10.1111/j.1365-2699.2006.01594.x.

Philips, S. *et al.* (2017) ‘Opening the black box: an open-source release of Maxent’. doi: [10.1111/ecog.03049].

Phillips, S., Anderson, R. and Schapire, R. (2006) ‘Maximum entropy modeling of species geographic distributions.’, *Ecological Modelling*, 190, pp. 231–259.

Primack, R. B. and Hall, P. (1992) ‘Biodiversity and Forest Change in Malaysian Borneo and management of Asian rainforests’, *BioScience*, 42(11), pp. 829–837.

van Proosdij, A. S. J. *et al.* (2016) ‘Minimum required number of specimen records to develop accurate species distribution models’, *Ecography*, 39(6), pp. 542–552. doi: 10.1111/ecog.01509.

Qin, A. *et al.* (2017) ‘Maxent modeling for predicting impacts of climate change on the potential distribution of Thuja sutchuenensis Franch., an extremely endangered conifer from southwestern China’, *Global Ecology and Conservation*. Elsevier B. V., 10, pp. 139–146. doi: 10.1016/j.gecco.2017.02.004.

Radosavljevic, A. and Anderson, R. P. (2014) ‘Making better Maxent models of species distributions: Complexity, overfitting and evaluation’, *Journal of Biogeography*, 41(4), pp. 629–643. doi: 10.1111/jbi.12227.

Raes, N. *et al.* (2009) ‘Botanical richness and endemicity patterns of Borneo derived from species distribution models’, *Ecography*, 32(1), pp. 180–192. doi: 10.1111/j.1600-0587.2009.05800.x.

Rajbhandary, S., Hughes, M. and Shrestha, K. K. (2011a) ‘Distribution Patterns of *Begonia* species in the Nepal Himalaya’, *Botanica Orientalis: Journal of Plant Science*, 7, pp. 73–78. doi: 10.3126/botor.v7i0.4386.

Rajbhandary, S., Hughes, M. and Shrestha, K. K. (2011b) ‘Distribution Patterns of Begonia species in the Nepal Himalaya Distribution Patterns of Begonia species in the Nepal Himalaya’, (March). doi: 10.3126/botor.v7i0.4386.

Rebelo, H. and Jones, G. (2010) ‘Ground validation of presence-only modelling with rare species: A case study on barbastelles Barbastella barbastellus (Chiroptera: Vespertilionidae)’, *Journal of Applied Ecology*, 47(2), pp. 410–420. doi: 10.1111/j.1365-2664.2009.01765.x.

Repin, R., Karim, R. M. and Mujih, H. (2015) ‘Three New Begonia (Begoniaceae) Species from Ultramafic Outcrops in Kinabalu Park, Sabah, Malaysia’, *Sandakania*, 20, pp. 155–164.

Repin, R., Sampang, K. and Limbawang, S. (2015) ‘Begonia (Begoniaceae) from Tawau Hills Park, Sabah, Malaysia, including Ten New Species’, *Sandakania*, 20, pp. 165–204.

Richardson, J. E. *et al.* (2014) ‘The influence of tectonics , sea-level changes and dispersal on migration and diversification of Isonandreae (Sapotaceae)’, (1869), pp. 130–140.

Richardson, J. E., Costion, C. M. and Muellner, A. N. (2012) 'The Malesian Floristic interchange: plant migration patterns across Wallace's Line', in Gower, D. et al. (eds) *Biotic Evolution and Environmental Change in Southeast Asia*. Cambridge University Press, pp. 138–163. doi: https://doi.org/10.1017/CBO9780511735882.008.

Richter, B. *et al.* (1999) 'Paleomagnetism of Borneo', *Journal of Asian Earth Sciences*, 17(1–2), p. 3. doi: 10.1016/S0743-9547(98)00057-9.

Roos, M. C. *et al.* (2004) 'Species Diversity and Endemism of Five Major Malesian Islands : Diversity-Area Relationships Published by : Wiley Stable URL : http://www.jstor.org/stable/3554669 Accessed : 25-05-2016 08 : 10 UTC Your use of the JSTOR archive indicates your acceptance of', *Journal of Biogeog*, 31(12), pp. 1893–1908.

Sakurai, K. (1999) 'Soils and Agriculture in Borneo', *Tropics*, 9(1), pp. 27–40.

Sang, J. *et al.* (2015) 'Begonia (Begoniaceae) from Kubah National Park and Environs, Sarawak, Borneo, including a New Species', *Sandakania*, 20, pp. 87–106.

Sang, J. and Kiew, R. (2014) 'Diversity of Begonia (Begoniaceae) in Borneo "How many species are there?', *Reinwardtia*, 14(1), p. 233. doi: 10.14203/reinwardtia.v14i1.420.

Sang, J., Kiew, R. and Yiing, L. C. (2015) 'A Decade of Begonia (Begoniaceae) from Sarawak, Borneo', *Sandakania*, 20, pp. 129–154.

Sang, J. and Yiing, L. C. (2015) 'Eight New Species of Begonia (Begoniaceae) from Murum Dam, Sarawak, Borneo', *Sandakania*, 20, pp. 107–128.

Scrivenor, J. B. (1927) 'Notes on the Geology of Sarawak', *Journal of the Malayan Branch of the Royal Asiatic Society*, 5(2), pp. 288–294.

Seino, T., Kanehiro, K. and Lakim, M. (2007) 'Floristic composition and stand structure of mixed Dipterocarp Forest in Tawau Hills Park, Sabah , Malaysia. Sabah Parks', *Nature*, 8, pp. 63–82.

Slik, J. W. F. *et al.* (2003) 'A floristic analysis of the lowland dipterocarp forests of Borneo', *Journal of Biogeography*, 30(10), pp. 1517–1531.

Slik, J. W. F. *et al.* (2009) 'Environmental correlates for tropical tree diversity and distribution patterns in Borneo', *Diversity and Distributions*, 15, pp. 523–532. doi: 10.1111/j.1472-4642.2009.00557.x.

Stockwell, D. R. B. and Peterson, A. T. (2002) 'Effects of sample size on accuracy of species distribution models', *Ecological Modelling*, 148, pp. 1–13.

Syfert, M. M. *et al.* (2014) 'Using species distribution models to inform IUCN Red List assessments', *Biological Conservation*. Elsevier Ltd, 177, pp. 174–184. doi: 10.1016/j.biocon.2014.06.012.

Syfert, M. M. *et al.* (2018) 'Inferring diversity patterns along an elevation gradient from stacked SDMs: A case study on Mesoamerican ferns', *Global Ecology and Conservation*. Elsevier Ltd, 16, p. e00433. doi: 10.1016/j.gecco.2018.e00433.

Thomas, D. C. (2010) *Phylogenetics and historical biogeography of Southeast Asian Begonia L . (Begoniaceae)*. University of Glasgow.

Thomas, D. C. *et al.* (2012) 'West to east dispersal and subsequent rapid diversification of the mega-diverse genus Begonia (Begoniaceae) in the Malesian archipelago', *Journal of Biogeography*. John Wiley & Sons, Ltd (10.1111), 39(1), pp. 98–113. doi: 10.1111/j.1365-2699.2011.02596.x.

Thuiller, W. *et al.* (2009) 'BIOMOD - A platform for ensemble forecasting of species distributions', *Ecography*, 32(3), pp. 369–373. doi: 10.1111/j.1600-0587.2008.05742.x.

Tinoco, B. A. *et al.* (2009) 'Distribution, ecology and conservation of an endangered Andean hummingbird: The Violet-throated Metaltail (Mettallura baroni)', *Bird Conservation International*, 19(1), pp. 63–76. doi: 10.1017/S0959270908007703.

Tittensor, D. P. *et al.* (2009) 'Predicting global habitat suitability for stony corals on seamounts', *Journal of Biogeography*, 36(6), pp. 1111–1128. doi: 10.1111/j.1365-2699.2008.02062.x.

Tongkul, F. (1994) 'The geology of Northern Sabah, Malaysia: Its relationship to the opening of the South China Sea Basin', *Tectonophysics*. Elsevier, 235(1–2), pp. 131–147. doi: 10.1016/0040-1951(94)90021-3.

Vollering, J. *et al.* (2016) 'Phytogeography of New Guinean orchids: Patterns of species richness and turnover', *Journal of Biogeography*, 43(1), pp. 204–214. doi: 10.1111/jbi.12612.

Wilson, M. E. J. and Moss, S. J. (1999) 'Cenozoic palaeogeographic evolution of Sulawesi and Borneo', *Palaeogeography, Palaeoclimatology, Palaeoecology*, 145, pp. 303–337.

Wisz, M. S. *et al.* (2008) 'Effects of sample size on the performance of species distribution models', *Diversity and Distributions*, 14(5), pp. 763–773. doi: 10.1111/j.1472-4642.2008.00482.x.

Wong, K. M. (1998) 'Patterns of plant endemism and rarity in Borneo and the Malay peninsula', *Rare, Thretened and Endangered Fora of Asia*, pp. 139–169.

Yun, C. F. *et al.* (2015) 'Begonia (Begoniaceae) in the Danum Valley Conservation Area, Sabah, Borneo, including Eleven New Species', *Sandakania*, 20, pp. 51–85.

Yun, C. F., Guanih, V. S. and Repin, R. (2015) 'Begonia kibambanganensis and B. melanosticta (Begoniaceae), Two New Species from Sabah, Borneo', *Sandakania*, 20, pp. 1–6.

Zachos, F. E. and Habel, J. C. (2011) *Biodiversity Hotspots*. London New York: Springer Heidelberg Dordrecht. doi: 10.1007/978-3-642-20992-5.

Zhang, M. G. *et al.* (2012) 'Using species distribution modeling to improve conservation and land use planning of Yunnan, China', *Biological Conservation*. Elsevier Ltd, 153, pp. 257–264. doi: 10.1016/j.biocon.2012.04.023.

Zhang, M. G. *et al.* (2014) 'Major declines of woody plant species ranges under climate change in Yunnan, China', *Diversity and Distributions*, 20(4), pp. 405–415. doi: 10.1111/ddi.12165.

Zimmermann, N. E. *et al.* (2010) 'New trends in species distribution modelling', *Ecography*, 33(6), pp. 985–989. doi: 10.1111/j.1600-0587.2010.06953.x.