

A species distribution modelling analysis of *Rafflesia pricei* (Rafflesiaceae), a parasitic flowering plant endemic to Borneo

Volodymyr Tytar¹, Iryna Kozynenko¹, Michael Navakatikyan²

¹I. Schmalhausen Institute of Zoology, NAS of Ukraine (Kyiv, Ukraine);

²University of New South Wales (Sydney, New South Wales, Australia)

article info

key words

Rafflesia pricei, Sabah, species distribution modelling, conservation priorities

correspondence to

Volodymyr Tytar; I. Schmalhausen Institute of Zoology, NAS of Ukraine, 15 Bohdan Khmelnytsky Street, Kyiv, 01054 Ukraine; Email: vtytar@gmail.com; orcid: 0000-0002-0864-2548

article history

Submitted: 09.04.2025. Revised: 29.05.2025. Accepted: 30.06.2025

cite as

Tytar, V., I. Kozynenko, M. Navakatikyan. 2025. A species distribution modelling analysis of *Rafflesia pricei* (Rafflesiaceae), a parasitic flowering plant endemic to Borneo. *GEO&BIO*, 27: 215–233. [In English, with Ukrainian summary]

abstract

Rafflesia pricei (Rafflesiaceae), an endangered holoparasitic plant with large, elusive flowers, faces conservation challenges in Sabah, Malaysia, due to its cryptic life cycle and habitat vulnerability. This study used species distribution models (SDMs) to analyse environmental factors shaping its distribution, aiming to identify areas suitable for new populations and inform conservation efforts. Creating SDMs for a poorly known species as *R. pricei* can be a difficult task because occurrence data is limited, therefore the method ‘ensemble of small models’ was used, implementing three standard algorithms: generalised linear models, generalised boosted regression models, and support vector machines. Analysing elevation, bioclimatic variables, edaphic characteristics, and cloud cover data revealed a complex interplay of drivers. Initial assessments highlighted elevation, consistent with known sub-montane occurrences. However, a refined model identified the mean daily mean air temperature of the coldest quarter as the most significant predictor, suggesting a critical role for subtle temperature variations in flowering, host physiology, and pollination. Moisture-related bioclimatic factors also had considerable influence, while edaphic characteristics were less prominent. The study highlights the fragmented and undersized nature of habitats suitable for *R. pricei* in Sabah. This fragmentation, combined with sensitivity to environmental changes and threats from deforestation and agriculture, necessitates targeted conservation. Identifying key environmental drivers provides a scientific basis for selecting and managing potential reserves to safeguard *R. pricei* and its host. The study advocates for consolidating larger, interconnected habitat patches, particularly within the Heart of Borneo initiative, to maximise long-term survival and ecosystem preservation. By revealing *R. pricei*'s environmental dependencies, this research is anticipated to enhance the monitoring and conservation management of the species. Further research into eco-physiological responses to key factors, including El Niño impacts, is recommended.

© 2025 The Author(s); Published by the National Museum of Natural History, NAS of Ukraine on behalf of GEO&BIO. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY-SA 4.0), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Особливості та умови поширення виду *Rafflesia pricei* (Rafflesiaceae), паразитичної квіткової рослини, ендеміка Борнео: аналіз на основі моделювання екологічної ніші

Володимир Титар, Ірина Козиненко, Михайло Навакатикян

Резюме. *Rafflesia pricei* є вразливою голопаразитичною рослиною з великими квітками. Цей вид зростає в Сабаху, Малайзія, та стикається з проблемами збереження через свій прихований життєвий цикл та руйнування середовища існування. У цьому дослідженні використано моделі поширення видів (SDM) для аналізу екологічних факторів, що визначають його поширення, а також виявлення нових популяцій та інформування про природоохоронні зусилля. Створення SDM для маловідомого виду, такого як *R. pricei*, може бути складним завданням через обмеженість даних, тому було використано метод «ансамблю малих моделей», що включає три стандартні алгоритми: узагальнені лінійні моделі, узагальнені бустингові регресійні моделі та опорних векторних машин. Аналіз висоти над рівнем моря, біокліматичних змінних, едафічних характеристик та даних про хмарність виявив складну взаємодію факторів. Початкові оцінки підкреслили важливість висоти, що узгоджується з відомими місцями зростання у субмонтанних лісах. Однак, уточнена модель визначила середню добову температуру повітря найхолоднішого кварталу як найзначніший предиктор, що свідчить про критичну роль навіть незначних температурних коливань у цвітінні, фізіології рослини-господаря та запиленні. Біокліматичні фактори, пов'язані з вологістю, також мали помітний вплив, тоді як едафічні характеристики виявилися менш важливими. Дослідження підкреслює фрагментованість та недостатній розмір придатних для *R. pricei* середовищ існування в Сабаху. Ця фрагментація, у поєднанні з чутливістю виду до змін навколишнього середовища, зумовлює необхідність цілеспрямованих природоохоронних заходів. Є нагальна потреба у консолідації більших, взаємопов'язаних ділянок проживання, особливо в рамках ініціативи «Серце Борнео», для максимізації довгострокового виживання виду та збереження екосистеми. Виявляючи складні екологічні залежності *R. pricei*, це дослідження, як очікується, покращить моніторинг та управління збереженням цього виду. Рекомендуються подальші дослідження еколого-фізіологічних реакцій на ключові фактори, включаючи вплив таких кліматичних явищ, як Ель-Ніньйо.

Ключові слова: *Rafflesia pricei*, Сабах, моделювання поширення видів, природоохоронні пріоритети

Адреса для зв'язку: Титар Володимир; Інститут зоології ім. І. Шмальгаузена НАН України; вул. Богдана Хмельницького 15, Київ, 01054 Україна; Email: vtytar@gmail.com; orcid: 0000-0002-0864-2548

Introduction

Throughout the biosphere, parasitic relationships exist across various life forms, including vascular plants [Nickrent 2020]. The peculiar world of parasitic plants has always fascinated scientists. Their bizarre structures, complex life cycles, and intricate relationships with their hosts continue to spark curiosity and inspire research [Gonzalez *et al.* 2024]. These parasitic plants often utilise specialised organs called haustoria to penetrate into and extract water and nutrients from their host plant, typically attaching to shoots or roots [Westwood *et al.* 2010]. Parasitic plants can be further categorised based on their photosynthetic capabilities: hemiparasites retain the ability to photosynthesise, while holoparasites are entirely reliant on their host for sustenance [Kuijt 1969]. The family Rafflesiaceae exemplifies a fascinating realm of holoparasitic plants. Notably, members of this family lack both chlorophyll and most of their functional vegetative system, including roots and shoots [Bendiksby *et al.* 2010]. Their existence hinges entirely on a specific host plant from the genus *Tetrastigma* (Miq.) Planch., a group of vines of the grape family, Vitaceae, found within the tropical rainforests of Southeast Asia. Molecular data analysis reveals that the Rafflesiaceae family encompasses three distinct genera: *Rafflesia* R. Br. ex Gray, *Rhizanthus* Dumort., and *Sapria* Griff. [Bendiksby *et al.* 2010].

Rafflesia has long been known and celebrated in Malaysia where thirteen species have been recorded, of which eight occur in Peninsular Malaysia and a further five in the states of Sabah and Sarawak (Borneo).



Fig. 1. *Rafflesia pricei* in full bloom in the Rafflesia Forest Reserve, 19 December 2019; photo taken by M. Navakatikyan.

Рис. 1. *Rafflesia pricei* в повному розквіті в заповіднику Rafflesia Forest Reserve, 19 грудня 2019 р.; фото зроблено М. Навакатикьяном.

Among these, of our special interest is a species named *R. pricei* Meijer, 1984 found largely in Sabah (Fig. 1). This remarkable plant thrives in humid, tropical climates, typically established in the understory of rainforests. It prefers well-drained, nutrient-rich soil, which supports the growth of its host vines. The presence of shade is crucial as sunlight can be detrimental to its growth. Given these conditions, *R. pricei* is particularly sensitive to environmental changes, making it vulnerable to deforestation and habitat loss due to agricultural expansion and human encroachment. Conservation efforts emphasise the importance of protecting not only this unique flowering plant but also its host plants and the entire ecosystem in which it lives. In Sabah, all *Rafflesia* species, together with their host *Tetrastigma* species, are listed under Total Protected Plants under the Wildlife Conservation Enactment 1997 [Malabrigo *et al.* 2023].

Despite having one of the biggest single flowers produced by a plant (the mature flower of *R. pricei* can reach a diameter of up to 30 centimetres), it is almost impossible to see when it is not blooming. For most of its life the parasite lives inside its host in a very reduced form as a thread just a cell wide [Nikolov *et al.* 2014; Nikolov & Davis 2017]. Eventually, a flower bud emerges, taking about nine months to bloom, after which it lasts for only a few days. Precisely where the vines are, and the parasites inside them are, is largely unknown as Borneo's forests are difficult to survey. Time and financial constraints are preventing a comprehensive population survey in Sabah. Therefore, it is challenging to establish protected areas to securely preserve Sabah's national flowers, in particular the species *R. pricei*.

Fortunately, recent decades have brought forth some practical tools for assessing population status and risks over large areas, including habitat suitability model predictions [Renjana *et al.* 2022]. Amongst such are species distribution models (SDMs), which have become one of the most widely used tools in ecology [Guisan & Zimmermann 2000; Guisan & Thuiller 2005; Rathore & Sharma 2023], and their use has been increasing over the last two decades [Hao *et al.* 2019]. The principle of SDMs is to relate species occurrences to environmental data in order to gain ecological insights [Elith & Leathwick 2009]. This family of methods has gained a broad range of applications, including the identification of areas for conservation and guiding the efforts to locate new populations of known species of conservation concern [Fois *et al.* 2018]. The resulting habitat maps show areas where the considered species is predicted to reside. This information is crucial for conservation efforts and selecting suitable reserve locations.

Therefore, considering the cryptic character and invisibility of *R. pricei* for much of its life, this study is aimed to:

- Find the effective environmental factors that play roles in its distribution by using SDMs;
- Conclude where it is most promising to search for new populations, and find which current areas could be turned into reserves to foster conservation of the species.

The results of this study can be used to enhance monitoring efforts and improve conservation-based management and can help to inform and update the distribution of *R. pricei* in the future, provide a deeper insight into the ecology of the species.

Materials and Methods

Study area

The Malaysian state of Sabah (73 904 km²) is located between latitudes 4°8' and 7°22' north of the equator on the north-eastern tip of the island of Borneo. The western and eastern regions are divided by the Crocker Range, which runs almost parallel to the west coast, extending from the southern end of Marudu Bay in the north and southwards along the western part of the state to the border with Sarawak. The narrow western lowland plains contain areas of low, flat ground that is densely populated, while eastern Sabah is characterised by low dissected hills, gentle slopes, and poorly drained flatland and low-lying swampy zones that have been extensively logged and converted to permanent agriculture where soil and terrain is suitable [Payne 1988].

According to Nais [2001], there are several populations of *R. pricei* in Sabah: Crocker Range, Mount Kinabalu, Mamut Copper Mine area in Ranau, Langanan Waterfall area, Poring Hot Springs, Bukit Lugas, Bukit Tunturugung near Bundu Tuhan, Ranau. This species was also found in Sarawak and Brunei Darussalam. Till date, Sabah has been recorded to have the largest distribution of *R. pricei* compared to other locations in Borneo [Lestari *et al.* 2020].

Occurrence records

The occurrence records of *R. pricei* were obtained from published inventory and expedition reports, and information summarised in the Global Biodiversity Information Facility database [GBIF 2024]. These were supplemented by a field survey conducted in the Rafflesia Forest Reserve in December 2019, where several sites with wild specimens of the parasitic flowering plant are reachable via hiking trails from the reserve's information centre (05°46.64'N, 116°20.60'E). The record localities were compiled into a CSV database (.csv).

Environmental datasets

Habitat data for species distribution modelling were obtained from previous research and downloaded from available websites using the 'geodata' R package [Hijmans *et al.* 2023] allowing to access geographic data including climate, elevation, land use, soil, crop, species occurrence, administrative boundaries, and other geospatial data.

We included to our research the Shuttle Radar Topography Mission digital elevation model (SRTM30) [Farr *et al.* 2007] to capture fine-scale variations in terrain known to influence species distributions. Particularly in rough terrain ecosystems, topographic features including aspect, slope angle, and elevation above sea can affect site parameters like sunlight distribution, humidity, and nutrient availability, each of which can be primary determinants of local species distributions, as well as plant growth and mortality patterns [Hadley 1994]. Among the SRTM30 elevation features the following were employed: eastness, elevation, northness, roughness, slope, the topographic position index (tpi), and the terrain ruggedness index (tri).

Valuable insights into climate patterns were obtained from the Chelsa Bioclim+ climate data set, a comprehensive and high-resolution dataset that offers a wide range of climate variables, including standard bioclimatic variables featuring temperature and precipitation metrics like annual mean

temperature, temperature seasonality, precipitation of wettest quarter, precipitation of driest quarter and a diverse cluster of others [Brun *et al.* 2022]. We employed the following: mean annual air temperature (bio1), mean diurnal air temperature range (bio2), isothermality (bio3), temperature seasonality (bio4), mean daily maximum air temperature of the warmest month (bio5), mean daily minimum air temperature of the coldest month (bio6), annual range of air temperature (bio7), mean daily mean air temperatures of the wettest quarter (bio8), mean daily mean air temperatures of the driest quarter (bio9), mean daily mean air temperatures of the warmest quarter (bio10), mean daily mean air temperatures of the coldest quarter (bio11), annual precipitation amount (bio12), precipitation amount of the wettest month (bio13), precipitation amount of the driest month (bio14), precipitation seasonality (bio15), mean monthly precipitation amount of the wettest quarter (bio16), mean monthly precipitation amount of the driest quarter (bio17), mean monthly precipitation amount of the warmest quarter (bio18), mean monthly precipitation amount of the coldest quarter (bio19), mean monthly climate moisture index (cmi_mean), accumulated precipitation amount on growing season days (TREELIM) (gsp), mean temperature of the growing season (TREELIM) (gst), mean monthly near-surface relative humidity (hurs_mean), net primary productivity (npp), mean monthly surface downwelling shortwave flux in air (rsds_mean), mean monthly vapor pressure deficit (vpd_mean).

The information on edaphic characteristics also plays a crucial role in determining the suitable locations for *Rafflesia* species to thrive and reproduce, as well as their susceptibility to environmental changes and disturbances [Siahaan *et al.* 2024]. Therefore, soil data was added to the research by downloading data from the SoilGrids 2.0 source [Poggio *et al.* 2021]. SoilGrids 2.0 covers a wide range of soil properties, of which we used (accounting for the depth interval of 5 to 15 cm): bulk density of the fine earth fraction (bdod), cation exchange capacity of the soil (cec), volumetric fraction of coarse fragments (> 2 mm) (cfvo), proportion of clay particles (< 0.002 mm) in the fine earth fraction (clay), total nitrogen (N) (nitrogen), soil pH (phh2o), proportion of sand particles (> 0.05/0.063 mm) in the fine earth fraction (sand), proportion of silt particles (≥ 0.002 mm and $\leq 0.05/0.063$ mm) in the fine earth fraction (silt), soil organic carbon content in the fine earth fraction (soc), organic carbon density (ocd).

In our work we specifically evaluated the effect of cloud cover frequency on predicting the potential distribution of *R. pricei* in Sabah. This is because cloud cover frequency can influence numerous important ecological processes, including reproduction, growth, survival, and behaviour, yet the assessment of its importance has remained remarkably limited [Wilson & Jetz 2016; Pacayán *et al.* 2019; Gerstner *et al.* 2024]. We included the MODIS-derived mean annual cloud cover product [Wilson & Jetz 2016], which integrates 15 years of twice-daily remote-sensing-derived cloud observations. For our purpose we used: mean monthly cloud frequencies, their mean annual, inter- and intra-annual variability, and the seasonal cloud concentration index.

All variables were employed at a spatial resolution of 30 arcseconds.

Disturbingly, correlations between environmental features can pose challenges in SDM development, as highly correlated variables can lead to multicollinearity, which possibly can affect model performance and interpretation [Brun *et al.* 2020]. To avoid problems with multicollinearity and reduce the dimensionality of predictor layers, we conducted a principal component analysis on the layers and retained the predictor axes that cumulatively explained at least 70% of data variation [Stevens 1992], which were used as new predictors. The use of PCA-derived variables is advised because, instead of dropping variables, which may contain useful or interpretable information about the regional environment, it allows summarising the broad axis of feature variation across the study area, favouring both efficient prediction and interpretability [De Marco & Nóbrega 2018; Velazco *et al.* 2019].

Species distribution modelling

The 'flexsdm' R (v. 3.3.3) modelling package [Velazco *et al.* 2022] was used for determining the potential geographic distribution of *R. pricei* across Sabah. The choice fell on 'flexsdm' because of its

high manipulation and parametrisation capacity based on different functions and arguments. These attributes enable users to define a complete or partial modelling workflow specific for a modelling situation.

Creating SDMs for such a poorly known species as *R. pricei* can be a difficult task because occurrence data is limited, which can lead to model overfitting. To address the issues associated with modelling the spatial distributions of such species, Lomba *et al.* [2010] and Breiner *et al.* [2015] proposed the method ‘ensemble of small models’ or ESM. ESM is a good choice for distribution modelling since it minimises the model overfitting [Radomski *et al.* 2022], which is acceptable given the limited sample size of presence points available in this investigation. In ESM, many bivariate models are created with pairwise combinations of each predictor variable, then an ensemble is performed. In ‘flexsdm’, the ensemble is created using the average of suitability across all of the ‘small models’, weighted by Somers’ D [Somers 1962]. In our case we used three standard algorithms to model the distribution of *R. pricei*: generalised linear models (GLM), generalised boosted regression models (GBM), and support vector machines (SVM). We defined our model’s calibration area as a 5 km buffered area of a minimum convex polygon around the presence points and randomly selected pseudo-absence locations. It has been suggested that random creation of pseudo-absences could be an appropriate strategy to overcome the limitations of presence-only data sets [Chefaoui & Lobo 2008].

The ensemble predictions were evaluated with the continuous Boyce index [Boyce *et al.* 2002], one of the most reliable presence-only evaluation metrics, not prone to overestimate the predictive power of models built on a limited amount of data. It is continuous and varies between -1 and $+1$. Positive values indicate a model that presents predictions that are consistent with the distribution of presences in the evaluation dataset, values close to zero mean that the model is not different from a random model, and negative values indicate counter predictions [Hirzel *et al.* 2006]. Values exceeding 0.9 can be considered as evidence of excellent predictive performance. Final ensemble predictions were produced by averaging the models and weighting them according to the Boyce index.

Further we developed a predictive model for the whole of Sabah to obtain a distribution map of the habitat suitability of *R. pricei* with management purposes. In this case, we use the extent of the State of Sabah as our prediction area.

Model interpretability

A critical challenge in the adoption of machine-learning models is their inherent lack of interpretability, often referred to as the ‘black box’ problem. Shapley Additive exPlanations (SHAP) [Lundberg & Lee 2017; Lin & Gao 2022] represents a significant advancement in addressing this concern by providing a framework for understanding how models arrive at their predictions. SHAP offers various advantages over other variable importance methods, including its model-agnostic nature. Leveraging game theory concepts, SHAP provides a robust framework for variable importance attribution in machine-learning models, despite the number of used variables. The SHAP value quantifies the magnitude and direction of the variable’s influence on the prediction [Lundberg & Lee 2017; Lundberg *et al.* 2018]. The R package ‘shap-values’ (<https://github.com/pablo14/>) in a modified version was used to calculate SHAP values for selected models. A summary plot offers a comprehensive view of the most influential variables in a model and ranks them based on their effect on the model’s predictions. Also, the package can produce SHAP dependence plots, a model agnostic visualisation tool that helps understand the relationship between a variable and the model’s prediction [Niemann *et al.* 2020]. In other words, SHAP values can be integrated into global explanations such as variable importance, but using a completely different from a range of widely applied so far variable selection methods [Bourhis *et al.* 2023; Song & Estes 2023; Wadoux *et al.* 2023; Frazier & Song 2025].

In this work we apply a two-stage approach, where we first train our models using all the predictors within each of the mentioned above categories: elevation features (SRTM30), bioclimatic features (Chelsa Bioclim+), edaphic characteristics (SoilGrids 2.0), and MODIS-derived cloud cover, and

calculate SHAP values for features within these groups to get a group-level importance score. Features with the highest SHAP values are considered the most important. Further we choose to keep only the top five most important within each group and discard the rest. In the following step we re-train our model using only the selected predictors in an expectation that this can lead to a simpler, more interpretable model with potentially improved performance by removing noise from less important features. Once again, we rank the employed features based on their effect on the model's predictions.

Maps of habitat suitability in the GeoTIFF format were processed and visualised in SAGA GIS [Conrad *et al.* 2015]. Statistical data was analysed using the PAST software package [Hammer *et al.* 2001] and the R environment (<https://www.r-project.org>).

Results and Discussion

Important environmental factors that play roles in the distribution of *R. pricei*

Elevation features

In general, the genus *Rafflesia* has been discovered to grow in lowland and hill primary, overlogged, and old secondary forests at altitudes ranging from 300 to 1600 m a.s.l. [Nais 2001]. *Rafflesia pricei* is reportedly found at elevations between 400 and 1300 metres in sub-mountainous hilly forests. For instance, in 2016 a new locality of *R. pricei* was found in a montane forest, on a hillside 60° slope, at 1277 m a.s.l. [Lestari *et al.* 2020]. Willem Meijer [Meijer 1984] suggested that possibly *R. pricei* prefers higher altitudes than the other species of *Rafflesia*. Using the 'Add Grid Values to Points' tool in SAGA GIS, information easily can be extracted from the elevational grid data at given specific occurrence locations. This shows that specimens of *R. pricei* in Sabah have been found at elevations between 20.7 and 1600 m a.s.l., with an average of 1141 ± 46 SE; however, around 62% of known records have been made at locations lying between 1200 and 1400 m a.s.l. These findings align with physico-geographical descriptions of low montane forests near Mount Kinabalu, the Crocker Range and its surroundings in Borneo, dominated by a mix of dipterocarp and other broadleaf trees, and where *Tetrastigma leucostaphylum* (Dennst.) Alston, a host for *R. pricei*, found for instance in the Rafflesia Forest Reserve [Chung *et al.* 2012]. Typically, the vine is considered to grow between elevations of 600 and 1400 m a.s.l. [Rafiqpoor & Nieder 2006; Wahab *et al.* 2021].

After the original values from the input data have been replaced by its SHAP values, passing the matrix of SHAP values to the bar plot function of the package 'shap-values' creates a global feature importance plot, where feature importance is calculated by averaging the absolute SHAP values for each feature across all predictions. This provides a measure of how important each feature is on average for all predictions (Fig. 2). The further a value is from zero, the more influence it will have on the prediction, whereas values near zero will have less of an impact on predictions. In this case top five elevation features have been ranked in the following order: elevation [0.1906], roughness [0.0053], terrain ruggedness index (tri) [0.0050], slope [0.0037], topographic position index (tpi) [0.0020]. Figures in square brackets are averaged absolute SHAP values. These results indicate that elevation is indeed the most important factor in this analysis, with an averaged absolute SHAP value of 0.1906. This value is significantly higher than the other elevation features considered, suggesting that elevation plays a dominant role in shaping the *R. pricei* habitat and influencing the processes favouring or suppressing the species.

Analysis of the corresponding SHAP dependence plot points out that the impact of elevation on the model's predictions follows a non-linear pattern. A locally weighted scatterplot smoothing (using the 'smoothing spline' module in PAST) indicates a global trend, peaking at an elevation of 1275 metres. This suggests that the influence of elevation on the model's output is most pronounced around this specific altitude.

However, it is important to note that the other factors, while less influential based on the SHAP values, still contribute to the overall picture. Roughness, terrain ruggedness index, slope, and topographic position index all have non-negligible impacts, albeit to a lesser degree than elevation, but

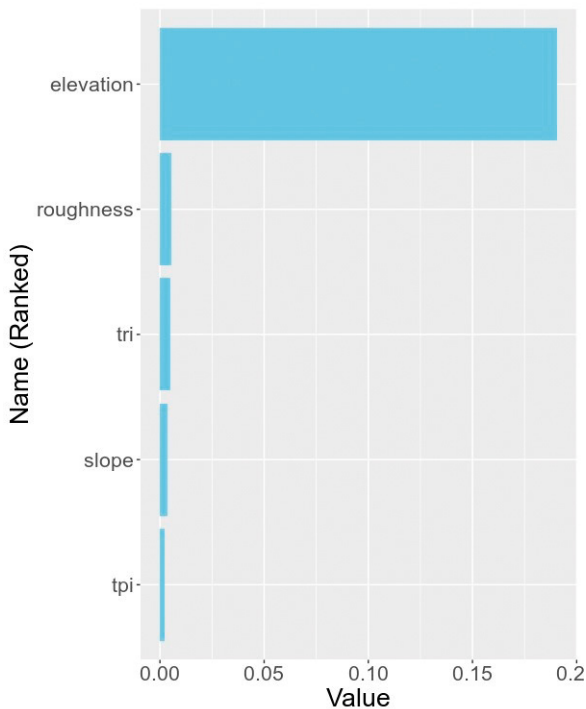


Fig. 2. Bar plot of global elevation feature importance based on the mean absolute magnitude of the SHAP values: elevation, roughness, tri (terrain ruggedness index), slope, tpi (topographic position index); value = mean absolute SHAP values.

Рис. 2. Стовпчаста діаграма глобальної важливості ознак висоти на основі середньої абсолютної величини значень Шеплі: висота над рівнем моря, нерівність місцевості, tri (індекс пересіченості рельєфу), нахил, tpi (індекс топографічного положення); value = середні абсолютні значення Шеплі.

each provide valuable information about the complexity and variability of the landscape, which can be crucial for various applications such as hydrological modelling, ecological studies and geomorphological analysis of the *R. pricei* habitat. A deeper understanding of the underlying relationships between these factors and the outcome of interest would require further investigation.

Bioclimatic features

Climate is a core aspect of a species' niche, and climate-based SDMs should therefore be useful tools for assessing the suitability of different locations for species [Abdulwahab *et al.* 2022].

Rafflesia pricei exhibits a strong preference for Southeast Asian locations with well-established tropical rainforest ecosystems. These ecosystems are characterised by consistently high temperatures, high humidity, and regular precipitation, the conditions that are pivotal for the successful growth and development of the host vines upon which the parasite is obligately dependent.

As mentioned, we chose the bioclimatic variables provided by CHELSA. One key reason for such a selection is that the underlying algorithm for composing this data set incorporates orographic predictors, which perform well in predicting important factors modified by elevation [Karger *et al.* 2020]. This is significant because elevation proved to be a leading variable amongst the considered elevation features.

Once again, using the SHAP framework for elucidating variable importance attribution in machine-learning models, top five bioclimatic features have been ranked in the following order (Fig. 3): mean monthly climate moisture index (cmi_mean) [0.0494], mean daily mean air temperatures of the wettest quarter (bio8) [0.0158], annual precipitation amount (bio12) [0.0147], precipitation amount of the wettest month (bio13) [0.0070], mean monthly precipitation amount of the wettest quarter (bio16) [0.0067]. As previously, averaged absolute SHAP values are in square brackets. In this row, the SHAP value furthest from zero, with a considerable magnitude, corresponds to the mean monthly climate moisture index, highlighting its significant and overwhelming influence on the prediction in terms of the considered bioclimatic variables.

The climate moisture index is a key variable that provides valuable information about water availability and is calculated as the difference between precipitation and potential evapotranspiration

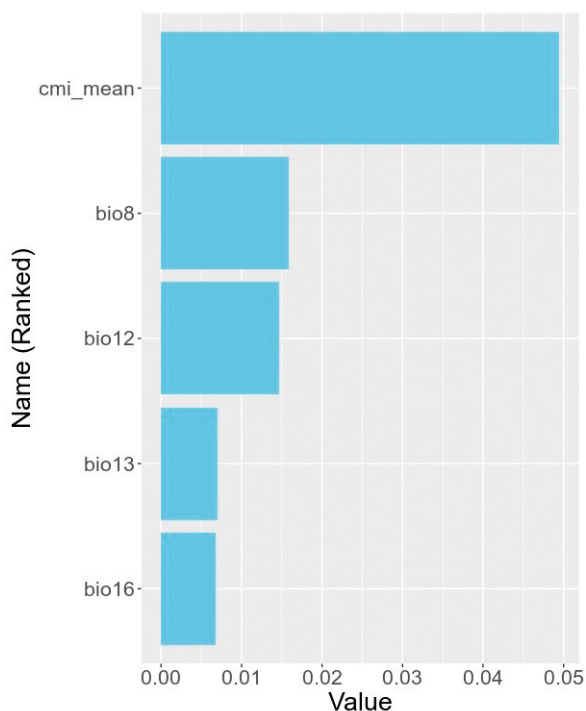


Fig. 3. Bar plot of global Chelsea Bioclim+ feature importance based on the mean absolute magnitude of the SHAP values: cmi_mean (mean monthly climate moisture index), bio8 (mean daily mean air temperatures of the wettest quarter), bio12 (annual precipitation amount), bio13 (precipitation amount of the wettest month), bio16 (mean monthly precipitation amount of the wettest quarter); value = mean absolute SHAP values.

Рис. 3. Стовпчаста діаграма важливості глобальних ознак Chelsea Bioclim+ на основі середнього абсолютного значення Шеплі: cmi_mean (середній місячний індекс кліматичної вологості), bio8 (середня добова середня температура повітря найвологішого кварталу), bio12 (річна кількість опадів), bio13 (кількість опадів найвологішого місяця), bio16 (середня місячна кількість опадів найвологішого кварталу); value = середні абсолютні значення Шеплі.

[Hogg 1997]. Positive values indicate that precipitation exceeds the amount of water that could potentially evaporate, suggesting sufficient moisture availability. The corresponding SHAP dependence plot reveals a non-linear relationship between the index and the SHAP values. Specifically, high initial index values are associated with a substantial increase in SHAP values. However, this trend reverses around 160 kg/m²/month, after which SHAP values begin to decrease.

In general, there appears to be an emphasis on moisture factors. While temperature (bio8) is included, the prominence of variables directly or indirectly related to moisture suggests its highly significant role in the model's predictions. Indeed, in the tropics, rainfall has a more significant effect because it varies greatly in amount and in seasonal distribution, while temperature has largely insignificant seasonal variation [Kedri *et al.* 2018].

Moisture plays a crucial role in the growth and survival of *Tetrastigma*, the host vine of *R. pricei*. Moisture is a critical factor for the health and survival of *Tetrastigma* vines, therefore adequate moisture levels ensure proper growth, nutrient uptake, and overall vigour, which are essential for supporting the parasite. Fortunately for *Rafflesia*, vines are one of the best sources of water. Vines move massive quantities of water, which may be one of the physiological reasons that *Rafflesia* colonise them, suggests in an interview Charles Davis, professor of evolutionary biology at Harvard University [Shaw 2017]. The flowers mostly consist of water themselves and the exponential growth of the blooms in the final stages of development is made possible 'primarily by pumping massive quantities of water into the flower,' says the professor.

However, excessive moisture is likely to create a range of problems, including root rot, nutrient leaching, and increased disease susceptibility. As indicated by the curvilinear behaviour of the SHAP dependence plot, maintaining a balance of moisture is essential for the health of the *Tetrastigma* vine and, consequently, for the survival of *R. pricei*.

Edaphic characteristics

As said, *Rafflesia* species are parasitic plants that rely entirely on their host plants for survival. The genus *Tetrastigma* is a common host for *Rafflesia*, and the edaphic characteristics of the soil where these host plants grow are destined to play a crucial role in the success of the intricate parasitic

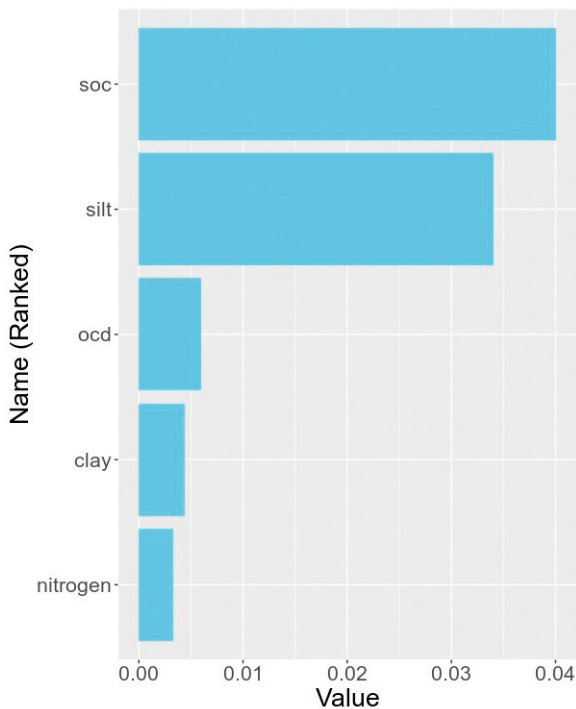


Fig. 4. Bar plot of global edaphic feature importance based on the mean absolute magnitude of the SHAP values: soil organic carbon content in the fine earth fraction (soc), proportion of silt particles in the fine earth fraction (silt), organic carbon density (ocd), proportion of clay particles in the fine earth fraction (clay), total nitrogen (nitrogen); value = mean absolute SHAP values.

Рис. 4. Стовпчикова діаграма глобальної важливості едафічних характеристик на основі середньої абсолютної величини значень Шеплі: вміст органічного вуглецю у дрібноземній фракції (soc), частка мулистих частинок у дрібноземній фракції (silt), щільність органічного вуглецю (ocd), частка глинистих частинок у дрібноземній фракції (clay), загальний азот (nitrogen); value = середні абсолютні значення Шеплі.

relationship [Siahaan *et al.* 2024]. Tetrastigma species are commonly found growing in the hillsides and valleys of shady and moist primary rainforests [Lianah *et al.* 2015]. These vines typically prefer well-drained soils with a loamy texture [Wahab *et al.* 2021a], which provide a good balance of water retention and aeration, essential for the healthy growth of the host plant.

Employing the SHAP framework for feature importance analysis in machine learning models, the top five influential edaphic variables were ranked as follows (Fig. 4): soil organic carbon content in the fine earth fraction (soc) [0.0401], proportion of silt particles in the fine earth fraction (silt) [0.0341], organic carbon density (ocd) [0.0060], proportion of clay particles in the fine earth fraction (clay) [0.0044], and total nitrogen (nitrogen) [0.0033]. The averaged absolute magnitude of the SHAP values for each feature is presented in square brackets.

This information provides important insights into the key edaphic factors driving ecological processes in the investigated environment. Saying specifically, soil organic carbon (soc) is considered a significant contributor to soil water retention [Ramírez *et al.* 2023] and heavily influences soil fertility [Bagwan *et al.* 2023; Singh *et al.* 2016]. An analysis of the corresponding SHAP dependence plot indicates that the impact of soil organic carbon on the model's predictions follows a bell-shaped pattern, reaching a maximum of around 66 g/kg. This can be considered a high concentration as far as the global average of soil organic carbon content is significantly lower, typically ranging from 10 to 30 g/kg. It is pertinent to note here that tropical forests, in particular, store up to 32% their carbon in the soil [Pan *et al.* 2011].

In another methodologically related paper soil organic carbon was found to be the most important modelled biophysical variable affecting the occurrence of the host plants of *R. arnoldii* on Sumatra [Renjana *et al.* 2022]. In this ecosystem, likely due to the relatively dense canopy cover, the vegetation creates a warm and moist microclimate, accelerating litter decomposition [Wang *et al.* 2010], which may enhance the accumulation of soil organic carbon.

The corresponding feature importance analysis also strongly points towards the soil texture, particularly the presence of silt, which is believed to play a crucial role in determining water and nutrient availability [Lal 2020], which prevents waterlogging, root damage, and decaying.

MODIS-derived cloud cover

As is the case with all tropical regions, our study area is affected by persistent cloud cover, which during a high cloud-cover season can be considerably blanketed by clouds, compared to a low cloud-cover season [Perbet *et al.* 2019].

Using the SHAP framework for exploring variable importance attribution in machine-learning models, top five features in this category of predictors representing mean monthly cloud frequencies have been ranked in the following order (Fig. 5): September (mean09) [0.0415], July (mean07) [0.0161], May (mean05) [0.0108], February (mean02) [0.0056], and August (mean08) [0.0046]; averaged absolute SHAP values are in square brackets.

Among the MODIS-derived cloud cover variables, the mean monthly cloud coverage for September exhibited the greatest influence on the prediction, as indicated by its SHAP value with a profound absolute magnitude in this row. The nature of this influence is not clear, but perhaps could be linked with its transitional character in Sabah, Malaysia, falling between the southwest monsoon (May to September) and the northeast monsoon (November to March). September marks the end of the southwest monsoon, which is generally associated with increased rainfall and cloud cover. The northeast monsoon is the primary source of rainfall during this period and is essential for maintaining the high levels of humidity and soil moisture that support the rich forest ecosystems housing the *Tetrastigma* and *Rafflesia* species. Possibly, conditions in September play a crucial role in setting the stage for the arrival of the corresponding monsoon, however this is only an assumption.

In the second phase of our approach, we re-train our model using only previously selected variables, five from each pool of environmental predictors, i.e. elevation and bioclimatic features, edaphic characteristics, cloud cover, and ranked the employed features based on their effect on the model's predictions in terms of the SHAP approach. The top five appeared in the following order (Fig. 6): mean daily mean air temperatures of the coldest quarter (bio11) [0.0318], mean monthly cloud frequency for September (mean09) [0.0229], elevation [0.0185], mean monthly cloud frequency for August (mean08) [0.0136], and precipitation amount of the driest month (bio14) [0.0125]. SHAP values are in square brackets. Surprisingly, soil properties were entirely absent from the five most significant

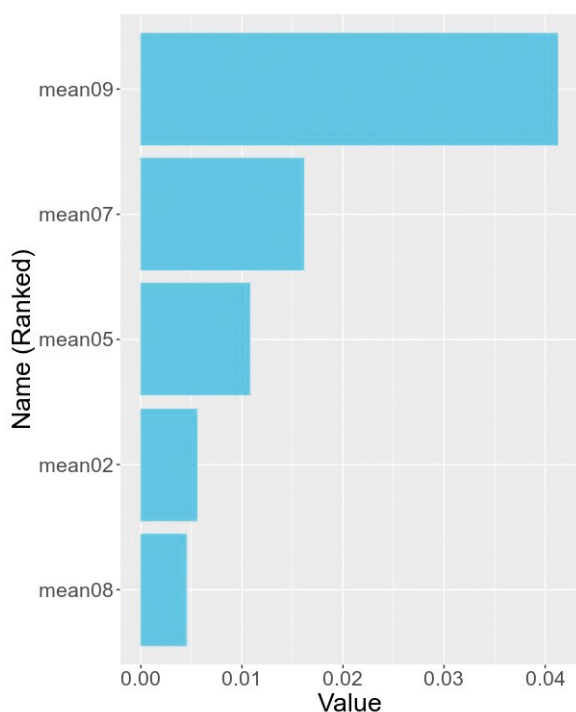


Fig. 5. Bar plot of global MODIS-derived cloud cover feature importance based on the mean absolute magnitude of the SHAP values: mean monthly cloud frequencies for September (mean09), July (mean07), May (mean05), February (mean02), August (mean08); value = mean absolute SHAP values.

Рис. 5. Стовпчаста діаграма важливості глобальних ознак хмарного покриття, отриманих за допомогою MODIS, на основі середньої абсолютної величини значень Шеплі: середні місячні частоти хмарності за вересень (mean09), липень (mean07), травень (mean05), лютий (mean02), серпень (mean08); value = середні абсолютні значення Шеплі.

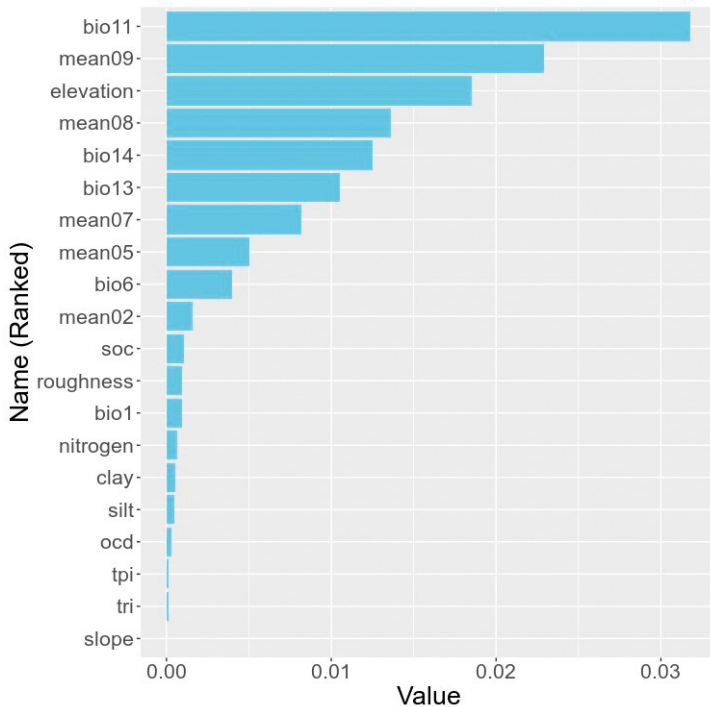


Fig. 6. Bar plot of global feature importance in the re-trained model using only previously selected variables: mean daily air temperatures of the coldest quarter (bio11), mean monthly cloud frequency for September (mean09), elevation, precipitation amount of the driest month (bio14), precipitation amount of the wettest month (bio13), mean monthly cloud frequency for July (mean07), mean monthly cloud frequency for May (mean05), mean daily minimum air temperature of the coldest month (bio6), mean monthly cloud frequency for February (mean02), soil organic carbon content (soc), roughness, mean annual air temperature (bio1), total nitrogen (N) (nitrogen), proportion of clay particles (clay), proportion of silt particles (silt), organic carbon density (ocd), topographic position index (tpi), terrain ruggedness index (tri), slope; value = mean absolute SHAP values.

Рис. 6. Стовпчаста діаграма глобальної важливості ознак у моделі з використанням лише попередньо відібраних змінних: середня добова температура повітря найхолоднішого кварталу (bio11), середня місячна частота хмарності за вересень (mean09), висота над рівнем моря, кількість опадів найсухішого місяця (bio14), кількість опадів найвологішого місяця (bio13), середня місячна частота хмарності за липень (mean07), середня місячна частота хмарності за травень (mean05), середня добова мінімальна температура повітря найхолоднішого місяця (bio6), середня місячна частота хмарності за лютий (mean02), вміст органічного вуглецю в ґрунті (soc), нерівність місцевості, середня річна температура повітря (bio1), загальний вміст азоту (N) (азот), частка глинистих частинок (clay), частка мулистих частинок (silt), щільність органічного вуглецю (ocd), індекс топографічного положення (tpi), індекс пересіченості рельєфу (tri), нахил; value = середні абсолютні значення Шеплі.

predictive factors. For instance, soil organic carbon content, ranking eleventh in the row of twenty predictors, accounted for just 3.4% of the influence seen from the coldest quarter's mean daily air temperatures (see: Fig. 6).

Because Sabah experiences a tropical climate, meaning relatively consistent warm temperatures throughout the year, this circumstance inherently reduces the extreme variations one might see in temperate climates. Accordingly, the overall difference between the 'coldest' and 'warmest' quarters is generally not as drastic as in regions with distinct seasons, although subtle variations do occur. Therefore, even more unexpected is the profound significance of the mean daily mean air temperatures of the coldest quarter. In this respect elevation must play a crucial role in temperature variations. Higher elevations, such as those in the mountainous regions of the Crocker Range and areas surrounding Mount Kinabalu, experience significantly cooler temperatures compared to coastal areas, especially at night. This means that while coastal areas may have relatively small differences between the warmest and coldest quarters, highland areas will have larger differences in temperature ranges. Information found in published sources indicate that within the Crocker Range, such as in Gunung Alab, daily temperature ranges of 13–25°C [Nainar *et al.* 2022]. This shows that there is a large swing in daily temperatures, which would also lend to larger swings in quarterly temperatures. Analysis of the corresponding SHAP dependence plot shows that the impact of coldest quarter's temperature on the model's predictions can be described by a horizontal trend followed by a steep descent after reaching the mean quarterly day temperature of about 22 °C, above which SHAP values quickly drop to zero and negative figures.

The coldest quarter's temperature likely acts as an indicator of broader seasonal changes, even if those changes are subtle. It might correlate with shifts in rainfall patterns, humidity, or sunlight, which collectively influence vegetation. The model might be picking up on this broader environmental signal through the coldest quarter's temperature. Ultimately, the importance of this variable is determined by its contribution to the model's predictive power. The fact that it ranked as the most important predictor suggests that it provides valuable information for understanding and predicting vegetation patterns in this mid-altitude zone of Sabah.

In more specific terms analysing the daily mean air temperature of the coldest quarter could possibly give a better understanding of the ecological requirements of *R. pricei* and its *Tetrastigma* host. As already mentioned, *Rafflesia* has a very unusual and complex life cycle, entirely dependent on its *Tetrastigma* host. It exists as a parasitic form within the host, only producing its massive flowers at certain times. The environmental cues that trigger flowering are still not fully understood, but temperature fluctuations, even subtle ones, could play a role. The coldest quarter's temperature might be one of the factors that influence the timing of *Rafflesia*'s flowering, affecting its reproductive success. *Tetrastigma* vine's health and vigour are crucial for *Rafflesia*'s survival. If the host vine is stressed due to temperature variations during the coldest quarter, it could negatively impact *Rafflesia*'s development and flowering. The coldest quarter's temperature might affect the vine's physiological processes, such as photosynthesis and nutrient uptake, indirectly influencing *Rafflesia*. Both the parasite and host might have specific temperature requirements or tolerances for optimal growth and reproduction. The mean daily mean temperature of the coldest quarter could be a critical factor determining whether these requirements are met. Even slight deviations from their preferred temperature range could affect their survival and distribution. The impact of temperature on *Rafflesia* and *Tetrastigma* could be intertwined with other environmental factors, such as rainfall, humidity, and sunlight. The coldest quarter's temperature might interact with these factors to create a specific microclimate that influences *Rafflesia*'s development and flowering. Along with that a non-seasonal flowering pattern seems to be characteristic of *Rafflesia* species in general and the blooming period depends on the individual bud's development cycle, which can vary significantly, but here, not everything is clear yet as far as in some species flowering appears to be seasonal [Hidayati *et al.* 2016].

Rafflesia flowers are pollinated by flies (Diptera: Brachycera), attracted by their carrion-like scent. For instance, species of the genera *Lucilia* and *Chrysomia* were reported to be pollinators of *R. pricei* [Beaman *et al.* 1988]. The activity of these pollinators is likely influenced by temperature. Cooler temperatures during the coldest quarter might affect the flies' behaviour, potentially impacting *Rafflesia*'s pollination success. In this respect, as exemplified by the species *R. tuan-mudae*, the ability of *Rafflesia* flowers to generate internal heat [Patiño *et al.* 2002] could be seen as an adaptation not only to enhance the dispersal of the flower's pungent odour but also create a favourable microenvironment within the central floral chamber that allows visiting pollinators to operate at a lower metabolic cost [Shaw 2017].

Considering these factors, it is possible that the mean daily mean air temperature of the coldest quarter plays a significant role in the ecology of *R. pricei* and its *Tetrastigma* host. It could influence the timing of flowering, the health and vigour of the host vine, and even the activity of pollinators. Further research specifically focusing on the temperature requirements and tolerances of these species would be needed to confirm these hypotheses.

In addition, we consider it might be important to give a brief mention of the El Niño phenomenon, which significantly influences the climate in Sabah, particularly during the December to February period, and which aligns with the northeast monsoon season and the coldest quarter of the year. El Niño has had a significant impact on Malaysia, particularly in Sabah and Sarawak, causing hot and dry weather [Thirumalai *et al.* 2017; Kemarau & Eboy 2021], leading to significantly higher daytime temperatures in Sabah, often exceeding 35 °C. These effects of temperature, though currently fairly irregular, could perhaps negatively affect the well-being of *Rafflesia* species and its vine host.

Model performance evaluation

From various sources we had at our disposal 57 occurrences scattered across northern Borneo, after excluding duplicate records and those outside of Sabah, 39 georeferenced occurrences were left for the modelling.

The results of the performance evaluation of the ESMs constructed based on different modelling techniques are shown in Table 1. The SVM algorithm produced the highest values (mean 0.988 ± 0.003 SE), while GLM had the lowest (mean 0.959 ± 0.001 SE), with a statistically significant difference ($p < 0.05$). Because of their excellent predictive performance, all produced models were subsequently employed in the construction of a weighted ESM. This was done despite the published fact that the use of different modelling techniques within an ESM framework was shown not to improve performance significantly [Breiner *et al.* 2015]. Nonetheless, we consider that other basic ideas in this case should also be taken into account, such as that different algorithms may capture different aspects of the species–environment relationship due to their varying assumptions and methodologies [Peterson & Soberón 2012].

Table 1. Ensembles of small models (ESM); accuracies of performance are given by the continuous Boyce index

Таблиця 1. Ансамблі малих моделей (АММ); оцінки точності моделей зазначені безперервним індексом Бойса

Ensembles of small models (ESM)	Elevation features (SRTM30)	Bioclimatic features (Chelsa Bioclim+)	Edaphic characteristics (SoilGr ids 2.0)	MODIS-derived cloud cover
GLM*	0.959	0.955	0.962	0.958
GBM	0.976	0.989	0.966	0.978
SVM	0.993	0.991	0.979	0.987

* GLM: generalised linear models; GBM: generalised boosted regression models; SVM: support vector machines.

Mapping habitat suitability and identification of conservation areas

Model outputs were used to map habitat suitability throughout Sabah (see: Fig. 6). Regions that have the value more than 0.5 are indicative of those regions which accommodate more suitable habitats for the considered species, while the value less than 0.5 are indicative of the opposite [Renjana *et al.* 2022; Tsegmed *et al.* 2023].

The predicted highly suitable habitats (i.e. > 50%) of *R. pricei* cover, in total, an area of around 2160 sq. km scattered across 83 patches within the administrative unit, comprising a mere 3% of Sabah's territory. In truth, it must be said that more than 70% of these patches of projected high suitability are each in size of less than one tenth of a square kilometre, suggesting that the highly specialised ecological requirements of the parasite can hardly be met within such a small-scale area. So, for example, because *R. pricei* relies entirely on its host vine, *Tetrastigma* spp., for survival, 'patch' viability is more about the health and extent of these host vines. In smaller, isolated patches, the distribution and health of *Tetrastigma* vines are disrupted, reducing the availability of suitable hosts. A viable *Rafflesia* population likely needs a continuous network of *Tetrastigma* vines, perhaps of an area of several square kilometres. In essence, the futility of undersized patches is underscored by the ongoing processes of shrinking and fragmentation of rainforests, which are directly jeopardising the survival of *Rafflesia* by destroying their core habitat and disrupting their delicate ecological balance [Malabrigo *et al.* 2023]. In this respect perhaps an account also should be taken of the creation of 'edge effects', where the conditions at the forest edge differ significantly from the interior [Magnago *et al.* 2015; Qie *et al.* 2017]. These edge effects can be detrimental to any species that thrive in the stable, interior conditions of the rainforest, which, however, sooner or later, find themselves, for varying reasons, to occur in small forest patches that are subjected to instability and may not be sufficient to support their livelihood.

Given that the predicted highly suitable habitat for *R. pricei* is fragmented across numerous patches representing a small fraction of Sabah's land, prioritising the search for new populations should focus

on exploring areas associated with existing highly suitable patches and within the same or nearby elevational and environmental ranges. Furthermore, the identification of potential reserve areas should concentrate on consolidating clusters of these highly suitable patches, particularly those exhibiting the largest contiguous areas and connectivity, to maximise conservation effectiveness and long-term viability of the species. In these terms it is most promising to search for new populations and find which current areas could be turned into reserves to foster conservation of the species by focusing on the largest projected contiguous area of habitat suitability located between longitudes of 115.89° and 116.69°E, and latitudes of 5.25° and 6.04°N (Fig. 7).

This area covers of around 1400 sq. km and comprises approximately 65% of predicted by the SDM highly suitable habitats of *R. pricei* in Sabah. While this region represents the largest and most promising concentration of predicted highly suitable habitat, it is important to acknowledge that not every part of this extensive area will be uniformly suitable, but it undoubtedly holds the greatest potential for discovering new *R. pricei* populations and identifying key areas for targeted conservation efforts.

This significant contiguous area of high habitat suitability for *R. pricei* falls within the broader context of the ‘Heart of Borneo’ (HoB) initiative, a transboundary conservation agreement aimed at protecting the biodiversity and rainforests of Borneo [Sloan *et al.* 2019] (Fig. 8). Recognising this overlap presents a compelling opportunity to integrate *R. pricei* conservation efforts within the larger

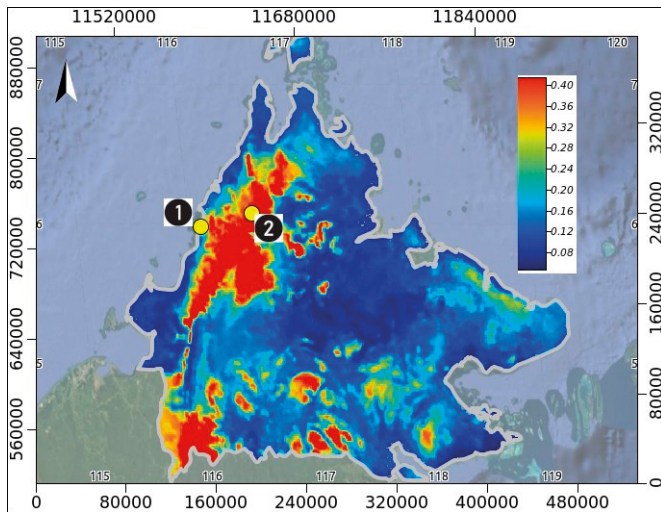


Fig. 7. The ensemble model output used to map habitat suitability for *Rafflesia pricei* throughout Sabah; colours show potential habitat suitability ranging from high (red) to low (blue); (1) the capital city of Kota Kinabalu; (2) Mount Kinabalu.

Рис. 7. Ансамблева модель поширення придатних територій для *Rafflesia pricei* по всьому Сабаху; кольори показують потенційну придатність середовища існування, що варіюється від високої (червоний) до низької (синій); (1) столиця Кота-Кінабалу; (2) гора Кінабалу.

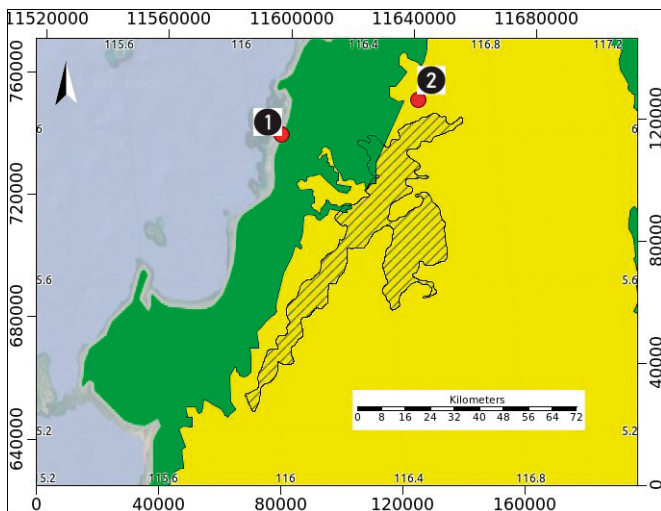


Fig. 8. Polygon representing the largest projected contiguous area of habitat suitability exceeding 50%, for *Rafflesia pricei* in Sabah, filled with a backward diagonal pattern; yellow colour displays the area designated as the ‘Heart of Borneo’; (1) the capital city of Kota Kinabalu; (2) Mount Kinabalu.

Рис. 8. Полігон, що представляє найбільшу прогнозовану суцільну ділянку придатності середовища існування, що перевищує 50%, для *Rafflesia pricei* в Сабаху, заповнений візерунком із діагональних ліній, спрямованих назад; жовтим кольором відображено територію, визначену як «Серце Борнео»; (1) столиця Кота-Кінабалу; (2) гора Кінабалу.

HoB framework, potentially leveraging existing conservation strategies, funding mechanisms, and collaborative partnerships already established within this ecologically critical region. The HoB has been particularly successful in Sabah, where it covers 54% of the state's territory and is acknowledged within the Sabah Biodiversity Strategy [Sabah... 2012]. Furthermore, focusing conservation efforts within this core area aligns strategically with the HoB's goals of maintaining large, interconnected forest landscapes essential for the survival of numerous endemic and endangered species, including *R. pricei*.

Conclusions

In conclusion, while various environmental factors contribute to the habitat suitability of *R. pricei*, this analysis highlights the significant and sometimes surprising influence of specific variables. Elevation and the mean daily mean air temperature of the coldest quarter emerged as particularly important predictors, underscoring the complex interplay of topographical and climatic conditions in shaping the distribution of this unique parasitic plant. Further research into the specific eco-physiological responses of *R. pricei* and its host to these key environmental factors, presumably also in the context of phenomena like El Niño, is crucial for effective conservation strategies.

In summary, species distribution modelling proves to be an invaluable tool in identifying and quantifying the fragmented yet critical habitats of *R. pricei*. By revealing that highly suitable areas are scattered and often undersized, the modelling underscores the pressing need for focused conservation efforts. Specifically, it directs attention towards exploring regions associated with existing and projected suitable patches and consolidating larger, interconnected habitat clusters, notably within the strategic framework of the 'Heart of Borneo' initiative. This targeted approach, informed by SDM insights, offers the most efficient pathway for discovering new *R. pricei* populations and establishing robust, long-term reserves essential for the survival of this unique species and the preservation of its vulnerable rainforest ecosystem.

Declarations

Funding. This study was not supported by any specific funding.

Conflict of interests. The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

- Abdulwahab, U. A., E. Hammill, C. P. Hawkins. 2022. Choice of climate data affects the performance and interpretation of species distribution models. *Ecological Modelling*, **471**: art. 110042. <https://doi.org/10.1016/j.ecolmodel.2022.110042>
- Bagwan, W. A., R. S. Gavali, A. Maity. 2023. Quantifying soil organic carbon (SOC) density and stock in the Urmodi River watershed of Maharashtra, India: implications for sustainable land management. *Journal of Umm Al-Qura University for Applied Sciences*, **9** (4): 548–564. <https://doi.org/10.1007/s43994-023-00064-3>
- Beaman, R. S., P. J. Decker, J. H. Beaman. 1988. Pollination of *Rafflesia* (Rafflesiaceae). *American Journal of Botany*, **75** (8): 1148–1162. <https://doi.org/10.1002/j.1537-2197.1988.tb08828.x>
- Bendiksby M., T. Schumacher, G. Gussarova, J. Nais, K. Mat-Salleh, [et al.]. 2010. Elucidating the evolutionary history of the Southeast Asian, holoparasitic, giant-flowered Rafflesiaceae: pliocene vicariance, morphological convergence and character displacement. *Molecular Phylogenetics and Evolution*, **57** (2): 620–33. <https://doi.org/10.1016/j.ympev.2010.08.005>
- Bourhis, Y., J. R. Bell, C. R. Shortall, W. E. Kunin, A. E. Milne. 2023. Explainable neural networks for trait-based multispecies distribution modelling—A case study with butterflies and moths. *Methods in Ecology and Evolution*, **14** (2): 1531–1542. <https://doi.org/10.1111/2041-210X.14097>
- Boyce, M. S., P. R., Vernier, S. E. Nielsen, F. K. A. Schmiegelow. 2002. Evaluating resource selection functions. *Ecological Modelling*, **157** (2–3): 281–300. [https://doi.org/10.1016/S0304-3800\(02\)00200-4](https://doi.org/10.1016/S0304-3800(02)00200-4)
- Breiner, F. T., A. Guisan, A. Bergamini, M.P. Nobs. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, **6**(10): 1210–1218. <https://doi.org/10.1111/2041-210X.12403>
- Brun, P., W. Y. Thuiller, L. Chauvier, L. Pellissier, R.O. Wüest, [et al]. 2020. Model complexity affects species distribution projections under climate change. *Journal of Biogeography*, **47** (1): 130–142. <https://doi.org/10.1111/jbi.13734>

- Brun, P., N. E. Zimmermann, C. Hari, L. Pellissier, D. N. Karger. 2022. CHELSA-BIOCLIM+ A novel set of global climate-related predictors at kilometre-resolution. *Earth System Science Data*, **14** (12): 5573–5603. <https://doi.org/10.5194/essd-14-5573-2022>
- Chefaoui, R. M., J. M. Lobo. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, **210** (4): 478–486. <https://doi.org/10.1016/j.ecolmodel.2007.08.010>
- Chung, A. Y. C., S. K. F. Chew, R. Majapun, R. Nilus, F. Kugan. 2012. Rafflesia Forest Reserve: Conserving not only *Rafflesia* but also endemic, rare & interesting insects. Paper presented at the Malaysia International Biological Symposium, Selangor, 1–3. <https://www.researchgate.net/publication/329152466>
- Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, **8** (7): 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- De Marco, P. Jun., C. C. Nóbrega. 2018. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PLoS ONE*, **13** (9): e0202403. <https://doi.org/10.1371/journal.pone.0202403>
- Elith, J., Leathwick, J. R. 2009. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40** (1): 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Farr, T. G., P.A. Rosen, E. Caro, R. Crippen, R. Duren, [et al]. 2007. The shuttle radar topography mission. *Reviews of Geophysics*, **45** (2): RG2004. <https://doi.org/10.1029/2005RG000183>
- Fois, M., A. Cuena-Lombraña, G. Fenu, G. Bacchetta. 2018. Using species distribution models at local scale to guide the search of poorly known species: Review, methodological issues and future directions. *Ecological Modelling*, **385**: 124–132. <https://doi.org/10.1016/j.ecolmodel.2018.07.018>
- Frazier, A. E., L. Song. 2025. Artificial intelligence in landscape ecology: Recent advances, perspectives, and opportunities. *Current Landscape Ecology Reports*, **10** (1): 1–13. <https://doi.org/10.1007/s40823-024-00103-7>
- GBIF. 2024. *GBIF Occurrence*. Download (14 November 2024). <https://doi.org/10.15468/dl.z8m4y4>
- Gerstner B. E., M. E. Blair, P. Bills, C.A. Cruz-Rodriguez, P. L. Zarnetske. 2024. The influence of scale-dependent geodiversity on species distribution models in a biodiversity hotspot. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, **382** (2269): 20230057. <https://doi.org/10.1098/rsta.2023.0057>
- Gonzalez, A. M., M. F. Romero, H. A. Sato. 2024. Exploring hidden connections: endophytic system and flower morphostem development of *Pilostyles berteroi* (Apodanthaceae) and interaction with its host *Adesmia trijuga* (Fabaceae). *Plants*, **13** (21): 3010. <https://doi.org/10.3390/plants13213010>
- Guisan, A., N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135** (2–3): 147–186. [http://dx.doi.org/10.1016/S0304-3800\(00\)00354-9](http://dx.doi.org/10.1016/S0304-3800(00)00354-9)
- Guisan, A., W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8** (9): 993–1009. <http://dx.doi.org/10.1111/j.1461-0248.2007.01044.x>
- Hadley, K. S. 1994. The role of disturbance, topography, and forest structure in the development of a montane forest landscape. *Bulletin of the Torrey Botanical Club*, **121** (1): 47–61. <https://doi.org/10.2307/2996883>
- Hammer, Ø., D. A. T. Harper, P. D. Ryan. 2001. Past: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, **4** (1): art. 4, 1–9. http://palaeo-electronica.org/2001_1/past/issue1_01.htm
- Hao, T., J. Elith, G. Guillera-Arroita, J. J. Lahoz-Monfort. 2019. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity & Distributions*, **25** (5): 839–852. <https://doi.org/10.1111/ddi.12892>
- Hidayati, S. N., J. L. Walck. 2016. A review of the biology of *Rafflesia*: what do we know and what's next? *Botanic Gardens Bulletin*, **19** (2): 67–78. <https://doi.org/10.14203/bkr.v19i2.166>
- Hijmans, R. J., M. Barbosa, A. Ghosh, A. Mandel. 2023. *Geodata: download geographic data*. <https://doi.org/10.32614/CRAN.package.geodata>
- Hirzel, A. H., G. Le Lay, V. Helfer, C. Randin, A. Guisan. 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, **199** (2): 142–152. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>
- Hogg, E. H. 1997. Temporal scaling of moisture and the forest-grassland boundary in western Canada. *Agricultural and Forest Meteorology*, **84** (1-2): 115–122. [https://doi.org/10.1016/S0168-1923\(96\)02380-5](https://doi.org/10.1016/S0168-1923(96)02380-5)
- Karger, D. N., D. R. Schmatz, G. Dettling, N. E. Zimmermann. 2020. High-resolution monthly precipitation and temperature time series from 2006 to 2100. *Scientific Data*, **7** (1): 248. <https://doi.org/10.1038/s41597-020-00587-y>
- Kedri, F. K., Z. Hamzah, N. S. Sukri, S. H. Yaacob, N. A. Majid, [et al]. 2018. Distribution and ecology of *Rafflesia* in Royal Belum state park, Perak, Malaysia. *International Journal of Engineering & Technology*, **7** (2.29): 292–296. <https://doi.org/10.14419/ijet.v7i2.29.13335>
- Kemarau, R. A., O. V. Ebov. 2021. The impact of El Niño–Southern Oscillation (ENSO) on temperature: a case study in Kuching, Sarawak. *Malaysian Journal of Social Sciences and Humanities*, **6** (1): 289–297. <https://doi.org/10.47405/mjssh.v6i1.602>
- Kuijt, J. 1969. *The Biology of Parasitic Flowering Plants*. University of California Press, Berkeley, 1–246. <https://doi.org/10.1525/9780520417793>

- Lal, R. 2020. Soil organic matter and water retention. *Agronomy Journal*, **112** (5): 3265–3277. <https://doi.org/10.1002/agj2.20282>
- Lestari, D., R. Mahyuni, R. Iryadi. 2020. *Rafflesia pricei* Meijer (Rafflesiaceae): A new locality in Borneo. *Berita Biologi*, **19** (2): 177–184. <https://doi.org/10.14203/beritabiologi.v19i2.3856>
- Lianah, L., K. Krisantini, P. Sopade. 2015. Status of traditional herb *Tetrastigma glabratum* (Blume) Planch in Mt Prau, Central Java, Indonesia. *Journal of Pharmacognosy and Phytochemistry*, **4** (4): 179–184. <https://www.phytojournal.com/archives/2015/vol4issue4/PartC/4-3-38.pdf>
- Lin K, Y. Gao. 2022. Model interpretability of financial fraud detection by group SHAP. *Expert Systems with Applications*, **210**: 118354. <https://doi.org/10.1016/j.eswa.2022.118354>
- Lomba, A., L. Pellissier, C. Randin, J. Vicente, F. Moreira, [et al.]. 2010. Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, **143** (11): 2647–2657. <https://doi.org/10.1016/j.biocon.2010.07.007>
- Lundberg, S. M., S. I. Lee. 2017. A unified approach to interpreting model predictions. *Neural Information Processing Systems*, **30**: 4768–4777. <https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>
- Lundberg, S. M., B. Nair, M.S. Vavilala, M. Horibe, M.J. Eisses, [et al.]. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, **2** (10): 749e760. <https://doi.org/10.1038/s41551-018-0304-0>
- Magnago, L. F. S., M. F. Rocha, L. Meyer, S.V. Martins, J. A. A. Meira-Neto. 2015. Microclimatic conditions at forest edges have significant impacts on vegetation structure in large Atlantic forest fragments. *Biodiversity & Conservation*, **24**: 2305–2318. <https://doi.org/10.1007/s10531-015-0961-1>
- Malabrigo Jr, P., A. B. Tobias, J. Witono, S. Mursidawati, A. Susatya, [et al.]. 2023. Most of the world's largest flowers (genus *Rafflesia*) are now on the brink of extinction. *Plants, People, Planet*, **7** (2): 331–346. <https://doi.org/10.1002/ppp3.10431>
- Meijer, W. 1984. New species of *Rafflesia* (Rafflesiaceae). *Blumea*, **30** (1): 209–215. <https://repository.naturalis.nl/pub/524883>
- Nainar, A., M. Mahali, K. U. Kamlun, N. A. Besar, L. Majuakim, [et al.]. 2022. Higher water yield but no evidence of higher flashiness in tropical montane cloud forest (TMCF) headwater streams. *Hydrology*, **9** (10): 162. <https://doi.org/10.3390/hydrology9100162>
- Nais, J., 2001. *Rafflesia of The World. Sabah Parks in association with Natural History Publications*. Kota Kinabalu. 1–243.
- Nickrent, D. L. 2020. Parasitic angiosperms: How often and how many? *Taxon*, **69** (1): 5–27. <https://doi.org/10.1002/tax.121953>
- Niemann, U., B. Boecking, P. Brueggemann, W. Mebus, B. Mazurek, [et al.]. 2020. Tinnitus-related distress after multimodal treatment can be characterized using a key subset of baseline variables. *PLoS ONE*, **15** (1): e0228037. <https://doi.org/10.1371/journal.pone.0228037>
- Nikolov, L. A., P. B. Tomlinson, M. Sugumaran, P. K. Endress, E. M. Kramer, [et al.]. 2014. Holoparasitic Rafflesiaceae possess the most reduced endophytes and yet give rise to the world's largest flowers. *Annals of Botany*, **114** (2): 233–242. <https://doi.org/10.1093/aob/mcu114>
- Nikolov, L. A., C. C. Davis. 2017. The big, the bad, and the beautiful: Biology of the world's largest flowers. *Journal of Systematics and Evolution*, **55** (6): 516–524. <https://doi.org/10.1111/jse.12260>
- Pan, Y., R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, [et al.]. 2011. A large and persistent carbon sink in the World's forests. *Science*, **333** (6045): 988–994. <http://dx.doi.org/10.1126/science.1201609>
- Pacayán, S., F. D. Alfaro, W. Pérez-Martínez, I. Briceño-de-Urbaneja. 2019. Potential distribution model of *Leontochir ovallei* using remote sensing data. *Revista de Teledetección*, **54**: 59–69. <https://doi.org/10.4995/raet.2019.12792>
- Patiño, S., T. Aalto, A.A. Edwards, J. Grace. 2002. Is *Rafflesia* an endothermic flower? *New Phytologist*, **154** (2): 429–437. <https://doi.org/10.1046/j.1469-8137.2002.00396.x>
- Payne, J. 1988. Orang-utan Conservation in Sabah, WWF Malaysia. Project No. 96/86 and WWF International Project No. 3759. WWF Malaysia, Kuala Lumpur, 137 pp. <http://myagric.upm.edu.my/id/eprint/8716>
- Perbet, P., M. Fortin, A. Ville, M. Béland. 2019. Near real-time deforestation detection in Malaysia and Indonesia using change vector analysis with three sensors. *International Journal of Remote Sensing*, **40** (19): 7439–7458. <https://doi.org/10.1080/01431161.2019.1579390>
- Peterson, A. T., J. Soberón. 2012. Species distribution modeling and ecological niche modeling: Getting the concepts right. *Natureza & Conservação*, **10** (2): 102–107. <https://doi.org/10.4322/natcon.2012.019>
- Poggio, L., L. M. de Sousa, N. H. Batjes, G. B. M. Heuvelink, B. Kempen, [et al.]. 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL*, **7** (1): 217–240. <https://doi.org/10.5194/soil-7-217-2021>
- Qie, L., S. L. Lewis, M. J. P. Sullivan, G. Lopez-Gonzalez, G. C. Pickavance, [et al.]. 2017. Long-term carbon sink in Borneo's forests halted by drought and vulnerable to edge effects. *Nature Communications*, **8**: 1966. <https://doi.org/10.1038/s41467-017-01997-0>
- Radomski, T., D. Beamer, A. Babineau, C. Wilson, J. Pechmann, [et al.]. 2022. Finding what you don't know: testing SDM methods for poorly known species. *Diversity & Distributions*, **28** (9): 1769–1779. <https://doi.org/10.1111/ddi.13536>

- Rafiqpoor, M. D., J. Nieder. 2006. Altitudinal zonation of climate and vegetation in a global megadiversity centre, Mount Kinabalu (North Borneo). *ERDKUNDE*, **60** (4): 362–374. <https://doi.org/10.3112/erdkunde.2006.04.06>
- Ramírez, P. B., S. Machado, S. Singh, R. Plunkett, F. J. Calderón. 2023. Addressing the effects of soil organic carbon on water retention in US Pacific Northwest wheat–soil systems. *Frontiers in Soil Science*, **3**: 1233886. <https://doi.org/10.3389/fsoil.2023.1233886>
- Rathore, M. K., L. K. Sharma. 2023. Efficacy of species distribution models (SDMs) for ecological realms to ascertain biological conservation and practices. *Biodiversity & Conservation*, **32** (10): 1–35. <http://dx.doi.org/10.1007/s10531-023-02648-1>
- Renjana, E., I. P. Astuti, E. Munawaroh, S. Mursidawati, J. R. Witono, [et al.]. 2022. Assessing potential habitat suitability of parasitic plant: A case study of *Rafflesia arnoldii* and its host plants. *Global Ecology & Conservation*, **34**: p.e02063. <http://doi.org/10.1016/j.gecco.2022.e02063>
- Sabah... 2012. Sabah State Government, 2012. Sabah Biodiversity Strategy 2012–2022. Available at: <http://ww2.sabah.gov.my/phb/wp-content/uploads/Sabah-Biodiversity-Conservation-Strategy-ilovepdf-compressed1.pdf>
- Shaw, J. 2017. Colossal Blossom. Pursuing the peculiar genetics of a parasitic plant. *Harvard Magazine*, March–April 2017; (online publication <https://www.harvardmagazine.com/2017/02/colossal-blossom>).
- Siahaan, F.A., R. Iryadi, D. Lestari. 2024. Edaphic characteristics of *Rafflesia* habitats in Indonesia: Implications for conservation and propagation. *Journal of Tropical Biodiversity & Biotechnology*, **9** (3): 1–11. <https://doi.org/10.1007/s10531-023-02648-1>
- Singh, A., P. Santra, M. Kumar, N. Panwar, P. R. Meghwal. 2016. Spatial assessment of soil organic carbon and physicochemical properties in a horticultural orchard at arid zone of India using geostatistical approaches. *Environmental Monitoring & Assessment*, **188** (9): 529. <https://doi.org/10.1007/s10661-016-5522-x>
- Sloan, S., M. J. Campbell, M. Alamgir, A. M. Lechner, J. Engert, [et al.]. 2019. Trans-national conservation and infrastructure development in the Heart of Borneo. *PLoS One*, **14** (9): e0221947. <https://doi.org/10.1371/journal.pone.0221947>
- Somers, R. H. 1962. A new asymmetric measure of association for ordinal variables. *American Sociological Review*, **27** (6): 799–811. <https://doi.org/10.2307/2090408>
- Song, L., L. Estes. 2023. itsdm: Isolation forest-based presence-only species distribution modelling and explanation in R. *Methods in Ecology & Evolution*, **14** (3): 831–840. <https://doi.org/10.1111/2041-210X.14067>
- Stevens, J. 1992. *Applied Multivariate Statistics for the Social Sciences (4th ed.)*. Lawrence Erlbaum Associates, Inc., 1–720. <https://doi.org/10.4324/9781410604491>
- Thirumalai, K., P. N. DiNezio, Y. Okumura, C. Deser. 2017. Extreme temperatures in Southeast Asia caused by El Niño and worsened by global warming. *Nature Communications*, **8**: 15531. <https://doi.org/10.1038/ncomms15531>
- Tsegmed, Z., S. Baasanmunkh, K. Oyundelger, B. Oyuntsetseg, U. Bayarsaikhan, [et al.]. 2023. Predicting the current and future suitable habitats, species distribution, and conservation assessment of *Fritillaria dagana* (Liliaceae). *Journal of Asia-Pacific Biodiversity*, **16** (3): 384–390. <https://doi.org/10.1016/j.japb.2023.01.004>
- Velazco, S. J. E., F. Villalobos, F. Galvão, P. A. De Marco (Jr.). 2019. A dark scenario for Cerrado plant species: Effects of future climate, land use and protected areas ineffectiveness. *Diversity & Distributions*, **25** (4): 660–673. <https://doi.org/10.1111/ddi.12886>
- Velazco, S. J. E., M. B. Rose, A. F. A. Andrade, I. Minoli, J. Franklin. 2022. flexsdm: An R package for supporting a comprehensive and flexible species distribution modelling workflow. *Methods in Ecology & Evolution*, **13** (8): 1661–1669. <https://doi.org/10.1111/2041-210X.13874>
- Wadoux, A., P. A. Saby, M. P. Martin. 2023. Shapley values reveal the drivers of soil organic carbon stock prediction. *SOIL*, **9** (1): 21–38. <https://doi.org/10.5194/soil-9-21-2023>
- Wahab, R., G. Gunsalam, F. Karolus, D. Dasini, M. Fadil, [et al.]. 2021. Floristic community composition in *Rafflesia*'s habitat at Kinabalu Park, Sabah, *Borneo Science. The Journal of Science and Technology*, **42** (2): 1–19. <https://doi.org/10.51200/bsj.v42i2.4468>
- Wahab, R., A. A. Mus, K. Saibeh, S. Khamis, H. Mujih, [et al.]. 2021a. Soil physico-chemistry in the habitat of *Rafflesia* in Kinabalu Park, Sabah, Malaysia. *Journal of Tropical Biology & Conservation (JTBC)*, **18**: 149–165. <https://doi.org/10.51200/jtbc.v18i.3450>
- Wang, S., H. Ruan, Y. Han. 2010. Effects of microclimate, litter type, and mesh size on leaf litter decomposition along an elevation gradient in the Wuyi Mountains, China. *Ecological Research*, **25** (6): 1113–1120. <https://doi.org/10.1007/s11284-010-0736-9>
- Westwood, J. H., J. I. Yoder, M. P. Timko, C. W. DePamphilis. 2010. The evolution of parasitism in plants. *Trends in Plant Science*, **15**(4): 227–235. <https://doi.org/10.1016/j.tplants.2010.01.004>
- Wilson, A. M., W. Jetz. 2016. Remotely sensed high resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. *PLoS Biology*, **14**(3): e1002415. <https://doi.org/10.1371/journal.pbio.1002415>