Trees along a highway: Species composition, risk rating, and tree performance

Shing Him, Louis LEE
Faculty of Design and Environment, Technological and Higher Education Institute of Hong Kong, Shing Tai Road, Chai Wan, Hong Kong, China
Hao ZHANG (allenzh@thei.edu.hk)
Faculty of Design and Environment, Technological and Higher Education Institute of Hong Kong, Shing Tai Road, Chai Wan, Hong Kong, China
Tze Kwun, Kathy NG
Landscape Division, Highways Department, Spectrum Tower, 53 Hung To Road, Kwun Tong, Kowloon, Hong Kong, China
Shun Cheong LO
Landscape Division, Highways Department, Spectrum Tower, 53 Hung To Road, Kwun Tong, Kowloon, Hong Kong, China
Siu Lun, Alan YU
Landscape Division, Highways Department, Spectrum Tower, 53 Hung To Road, Kwun Tong, Kowloon, Hong Kong, China

---

Research Article

Keywords: Highway greening strategy, Tree risk management, Forest structure, Species composition, Urban biodiversity, Urban tree monoculture

Posted Date: July 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1781793/v1

Abstract

Species composition is an important aspect of urban forestry. Some urban tree stands may be monospecific. It is important to identify factors explaining and predicting the occurrence of monospecific stands in previous less-explored space, such as transport land use. A large-scale tree survey was conducted, covering 53 slopes and 52 verges along San Tin Highway in Hong Kong. The aim of this study was to examine the occurrence of monospecific stands in relation to the biodiversity and the structure of urban forests adjacent to transport routes. 7,209 trees were found in the surveyed sites. However, the tree stock was dominated by *Corymbia citriodora* (72%) which was featured in monospecific tree stands, especially for the highway verges. Using chi-square tests, significant associations were found between monospecific stands, habitat type, and tree risk rating. A logistic regression model was constructed to predict the occurrence of monoculture. Every metre increase in maximum tree height, the odds of a stand being monospecific would be 1.22 times greater. Stands on verges had 5.26 times greater odds of being monospecific against slope. The associations and relationships were attributed to the dominance of *C. citriodora*. By boosting the logistic model, model reliability increased as kappa rose from 0.51 to 0.63, while balanced accuracy improved from 0.72 to 0.85. The occurrence of monospecific stands could be reliably predicted using maximum tree height and habitat type of tree stands. The findings could assist urban forest management in terms of species selection and biodiversity enhancement.

1. Introduction

Unlike their natural counterparts, urban forests represent a distinct type of vegetative community which is influenced by human management decisions. Trees serve different functions in different urban land uses. In transport land use, road users may expect a uniform tree line formed by structurally sound and nuisance-free trees. Expectations on urban forests evolve with time. Urban planners discovered possibilities in mainstreaming urban forestry in the name of biodiversity enhancement even in compact cities (Soga et al. 2014; O'Sullivan et al. 2017). Mixed-species stands provided a greater magnitude and a wider spectrum of ecosystem services (Felton et al. 2016; Huang et al. 2021). For instance, leaf colours and inflorescence can furnish visual aesthetics with seasonal shifts (Mu et al. 2022). Varied biomass structure can aid pollutant deposition and removal (Kaighn and Yu 1996). However, recent research has noted the limited tree biodiversity in transport land use (Sever-Mutlu et al. 2017; Liu and Slik 2022; Źróbek-Sokolnik et al. 2021). Roadside habitats such as verges can make up as much as one-third of public urban green space (Marshall et al. 2019). The large areal extent of well-managed urban forests can contribute to biodiversity conservation. Before large-scale implementation of any plans on vegetative biodiversity, a comprehensive survey can help contextualise the plans. More research is needed to draw empirical evidence about current urban tree biodiversity conditions in transport land use.

Researchers established some relationships between biodiversity parameters and the intrinsic properties of urban forests. A study utilising global urban forest inventories showed that abundance was a good predictor of species diversity (Kendal et al. 2014). Highway slopes can accommodate a lot of trees (Song et al. 2015). Thus, the effect of abundance can be tested. A Chinese study found that the size of habitat significantly influenced species compositional dissimilarity in remnant urban forest patches (Yang, et al. 2021b). A study on slope trees in Eastern China found that tree growth was attributed to slope angle (Yu et al. 2010). Environmental conditions may also affect vegetation composition (Orlóci and Stanek 1980). Much more findings may be found by digging into the relationships between biodiversity parameters and habitat properties. Such recommendable approach has yet been systematically implemented in the examination of green space in transport land use such as highways.

Transport infrastructures, such as highways and roads, are man-made entities, so are the tree plantings around them. In American, Asian and European studies, spatial patterns of species richness could be found in different urban locations (Divakara et al. 2021; Lugo-Pérez and Sabat-Guérnica 2011; Ma et al. 2020; Casanelles-Abella et al. 2021; Yang et al. 2021a). A wide range of artificial factors can affect biodiversity of tree stands in transport zone. For instance, road age was negatively related to plant diversity in China (Zeng et al. 2011). How landscape planners respond to habitat properties may affect the decision-making in the process of urban forest establishment. Some roadside urban forests are expected to tolerate environmental pollution (Modlingerová et al. 2012; Bandara and Dissanayake 2021). And sometimes traffic safety is highly important in planting design (Hale and Morzillo 2020). But if a study features multiple tree habitats created in a single transport facility, the confounding effects related to other anthropogenic drivers can be minimised. Therefore, research can be conducted in spatially proximal sites managed by the same authority.

Much of the existing urban tree research relies on modelling. Relevant subjects include allometry (Zheng et al. 2018), growing space requirement (Pretzsch et al. 2015), environmental justice (Violin et al. 2020), and spatial distribution pattern (Jim 1989). Very often, explanatory or predictive models are constructed using collected data. Additional insights may be gained by model training, tuning, and testing in machine learning. Such techniques can be applied on the study of tree biodiversity in transport land use. It is not uncommon to observe dominance by a few species in roadside urban green space. There could also be a possibility that a tree stand is monospecific. Therefore, it would be of research interest to explain and predict the occurrence of monoculture in tree stands along highways using a machine learning approach.

Data analysis procedures based on machine learning often requires an adequately large data set. In some urban forest sites, sampling may be difficult or even impossible. Due to site restriction and safety issues, conducting tree surveys on highway verges and slopes could be meticulous, resulting in small data set. But some research instances showed that small samples were acceptable given that appropriate analysis techniques and comprehensive model evaluation metrics are used (Shataee et al. 2012; Fassnacht et al. 2014; Zheng et al. 2018). Therefore, it is high time that urban foresters applied more advanced techniques in the investigation of issues such as managing and improving monoculture.

The focus of this research was urban trees along highways. Along a heavily used highway in Hong Kong, tree survey was conducted on some randomly selected slopes and verges. The aim of this study was to examine the occurrence of monospecific stands in relation to the biodiversity and the structure of urban forests adjacent to transport routes. Practical implications about urban tree management were distilled from the findings.

2. Methodology

2.1. Study area

This research was based in Hong Kong (22.32° N, 114.17° E). With more than 7.403 million residents on just 1,114 km^2 land area, Hong Kong is regarded a compact city (HKSARG Planning Department 2021; HKSARG Census and Statistic Department 2022). The transport of people and goods in Hong Kong relies on the Strategic Route and Exit Number System which is composed of inter-connected highways and expressways (HKSARG Transport Department 2020; HKSARG Highways Department 2022). This research focused on San Tin Highway, which is a major section of Route 9, which is the longest route in the territory (Fig. 1). San Tin Highway is a 6.3 km-long, three-lane, dual-carriageway in northwest Hong Kong. Yuen Long, a major residential and business area, lies in the southern exit of San Tin Highway, whereas the northern exit leads to the border between Hong Kong and Shenzhen, another prosperous city in mainland China.

Despite its heavy use, San Tin Highway is lined with rows of trees which grow on slopes and verges. Technical circulars have been published and continually updated to incorporate greenery in transport infrastructures in Hong Kong (e.g. HKSARG Development Bureau 2012; 2018). In the past, tree planting featured a limited palette of species, resulting in low-diversity stands and even monoculture. The technical circulars are applicable to the landscaping works of San Tin Highway.

52 verges and 53 slopes along San Tin Highway were randomly selected by the Highways Department in pursuit of a better understanding of the biodiversity and structure of the tree thereon. The characteristics and the management of verges and slopes were different. Thus, a distinction was drawn between them. In this study, all trees on a slope or a verge were collectively regarded as a stand. The Highways Department acknowledged that monospecific stands were common along San Tin Highway. This study was commissioned in order to investigate the factors influencing the monospecific stands.

2.2. Tree inventory and assessment

In this research, 53 slopes (Fig. 2a) and 52 verges (Fig. 2b) along San Tin Highway were randomly selected for sampling. Each slope or verge was visited once by a team of independent arborists and horticulturalists under the official coordination by the Highways Department. The data collection campaign lasted from 4th March, 2021 to 4th July, 2021. In this study, all trees on a slope or a verge was collectively regarded as a stand. The structure of the stands was quantified by abundance and maximum tree height. First, abundance was defined as the number of trees in a stand. Then, maximum tree height was defined as the height of the tallest tree in a stand, which was measured using a laser hyposometer (TruPulse 200, Laser Technology, Centennial).

The species composition of the stands was measured using different variables. Species richness was defined as the number of species found on each slope or verge. The identification of species was based on the official taxonomical publications in Hong Kong (Hong Kong Herbarium and South China Botanical Garden 2007; 2008; 2009; 2011; Hong Kong Herbarium 2012). Stands which consisted purely of one tree species were labelled as monospecific, but mixed-species if more than one species. Therefore, a stand could be dichotomously classified as monospecific or mixed-species. In order to quantify the biodiversity of the slopes and verges, three biodiversity indices, namely Shannon-Wiener diversity index (*H*), Simpson's index (*D*) and Pielou's evenness index (*J*), were derived as follows:

$$H = -\sum_{i=1}^{S} P_i \ln\left(P_i\right) \qquad (1)$$

where P_i was the proportion of tree species *i*, as computed by the number of tree species *i* over the total number of trees in each site. *S* was the total number of species thereon. A monoculture would be denoted by $H = 0$.

$$D = \sum_{i=1}^{S} \frac{n_i\left(n_i - 1\right)}{N(N-1)} \qquad (2)$$

where n_i was the number of tree species *i*. *N* was the total number of trees on each slope and verge. *D* was equal to the probability of two randomly and independently selected trees belonging to the identical species. In other words, a larger *D* value would imply a greater dominance by one or multiple certain species. *D* would be one in case of monoculture.

$$J = \frac{H}{\ln(S)} \qquad (3)$$

where *H* was Shannon-Wiener diversity as above. *S* was the number of species. The greater the *J*, the more even the species distribution was. At $J = 0$, monoculture would occur. *H*, *D*, and *J* have been applied in the study of Hong Kong's urban forests (Lee et al. 2019; 2021). The three biodiversity indices were forwarded to subsequent analysis.

In addition, tree risk assessment was conducted because tree risk was subsumed under road safety. In Hong Kong, the tree risk assessment protocol emanated from the best management practices published by the International Society of Arboriculture (American National Standard 2017; Dunster et al. 2017; Smiley et al. 2011). A limited visual tree risk assessment (Level 1 assessment) was carried out. Out of safety issues, time-consuming assessment exercises were ruled out in order to minimise the exposure of the personnel to busy road traffic. To uphold consistency in sampling, the assessment was performed by the same team of arborists for every stand. The outcome of each risk rating process was expressed by one of four ordinal categories, namely, low, moderate, high and extreme, at an ascending order of tree risk level. The risk rating was as a categorical variable.

2.3. Data analysis

Abundance, maximum tree height, species richness, *H*, *D*, *J* and risk rating, with respect to monoculture and habitat type, were sumamrised. Several statistical tests were run in order to distill specific management implications. Due to non-normality, abundance (Shapiro-Wilk $W = 0.712$, $p < 0.01$) and species richness (Shapiro-Wilk $W = 0.647$, $p < 0.01$) were log- and exponential-transformed respectively. Stand structure and species composition were compared between verges and slopes by independent *t* tests. Another round of *t* tests was conducted on abundance and maximum tree height between monospecific versus mixed-species sites. Then, *Chi*-square tests were administered to search for associations between monoculture status, habitat type and risk rating.

In order to understand the occurrence of monospecific stands, a binomial logistic regression analysis was conducted. The dichotomous response variable had two classes, namely monospecific and mixed-species. Abundance, maximum tree height, habitat type, and risk rating served as predictors. The former two and the latter two were quantitative and categorical variables respectively. Since the data set was small ($n = 105$), some precautions were taken when implementing a machine learning approach in evaluating model performance. With stratification, the data were split into a training set (75%) and a testing set (25%). Class imbalance in the training set was cured by Synthetic Minority Oversampling Technique (SMOTE), with 100% over-sampling and 200% under-sampling in R package *DMwR* (Torgo 2010).

A preliminary linear model with a logit link function was constructed using the training set, and evaluated using the testing set. In order to improve the model, boosting was applied on the logistic regression model. After a random hyperparameter search with 50 attempts, the optimal number of iterations in boosting was found as 30. For the limited sample size, leave-group-out cross-validation with 1,000 rounds of resampling was chosen due to the limited sample size. Parameters in the model diagnostics between the boosted and the preliminary models were compared. All analyses in this study were supported by RStudio (RStudio Team 2019), and the packages *caret* (Kuhn 2008; 2016), *caTools* (Tuszynski 2021), *ggpubr* (Kassambara 2020) and *tidyverse* (Wickham et al. 2019).

3. Results And Discussion

3.1. Urban forest structure of the surveyed slopes and verges

This study was conducted in San Tin Highway in Hong Kong. 53 slopes and 52 verges were surveyed in this research. 7,209 trees were found. When all stands were considered together, the mean and median abundance were 69 and 43 trees respectively (Fig. 3a). The divergence between mean and median indicated skewness, which could be caused by 12 outliers with high abundance. These outliers were entirely composed of stand on slopes. The influence of the outliers was magnified by considering the slopes only, whose mean tree abundance (93 trees) nearly doubled the median value (50 trees). But if only verges were considered, the skewness was suppressed as the mean and median values were closer, being 44 and 42 trees respectively. The two habitat types, namely slope and verge, were compared in terms of tree abundance. Neither mean nor median difference was significant, as revealed by *t*-test ($t = 1.54$, $p > 0.05$) and Mann-Whitney U test ($U = 1{,}596$, $p > 0.05$). In other words, the number of trees in the stands along San Tin highway was fairly uniform.

The mean and median abundance of monospecific stands was 60 and 47 trees respective. In contrast, mixed-species stands had a higher mean abundance (85 trees) but a lower median (33 trees). Nevertheless, the differences in mean and median abundance between monospecific and mixed-species stands were insignificant by *t*-test ($t = -0.55$, $p > 0.05$) and Mann-Whitney U test ($U = 1{,}259$, $p > 0.05$). Therefore, differences in tree abundance with respect to habitat type and monoculture status were insignificant among the surveyed stands.

Considering the maximum tree height of all stands, the mean and median values were 17.21 m and 18.00 m respectively (Fig. 3b). Unlike other variables, skewness was limited for maximum tree height due to convergence of mean and median values. Such convergence was observed on both slopes (mean = 13.09 m, median = 12.00 m) and verges (mean = 21.40 m, median = 20.00 m). However, the maximum tree height on the verges was noticeably greater than that on slopes. Such a large slope-versus-verge differences were highly significant, as concluded from the results of *t*-test ($t = -10.1$, $p < 0.001$) and Mann-Whitney U test ($U = 310$, $p < 0.001$). In other words, a greater value of maximum tree height could be found on verges.

The box of monospecific stands stood above that of mixed-species stands (Fig. 3b). It showed that a greater maximum tree height in the monospecific stands. In fact, the mean and median values of the tallest tree among the monospecific stands, which were mostly *Corymbia citriodora* monocultures, were 19.35 m and 20.00 m respectively. Such statistics reflected the height of mature specimens of *C. citriodora*. On the other hand, the mean and median height of the tallest trees in the mixed-species stands were 13.27 m and 12.00 m respectively, which were significantly lower than those of monospecific stands, as confirmed by *t*-test ($t = 5.58$, $p < 0.001$) and Mann-Whitney U test ($U = 1{,}952$, $p < 0.001$).

In this research, a large number of trees was recorded on just a 6.3 km-long highway section. A gigantic tree stock in the transport land use was thus implied in Hong Kong's 2,193 km-long extensive road network. Sharp contrasts were made with other land uses. A study on half of the public residential estates found 47,801 trees (Lee et al. 2019), while a tree survey in a town centre found just 1,111 trees (Lee 2022). Such comparison highlighted the contribution of transport land use to urban forestry in Hong Kong, echoing similar situations overseas (e.g. Marshall et al. 2019). In Hong Kong, highways are relatively long-distance transport corridors connecting various town centres. They cut through rural areas and sometimes serpent the countryside due to terrain constraints. During highway development, tree planting has been integral to compensatory planting and landscape improvement. Such greening efforts resulted in an enormous tree stock.

The tree stock might be made up by a consideration proportion of tall trees, especially the verges. The fact that verges and monospecific stands had taller trees could be attributed to the common occurrence of *C. citriodora*. This species could easily attain > 20 m in urban locations (Jim, 1990). The effect of such large trees on car crash probability requires more research within the local context. A large tree canopy may catch drivers' attention and raise their awareness of avoiding trees as dangerous objects (Coppola and Golombek 2018). But large tree size was shown to raise collision risk (Bucsuházy et al. 2022). Along other highways with relatively long establishment history, similar situations may occur as mature specimens of a few species could be commonly found in the stands. A huge tree stock should be carefully and cautiously managed. Any tiny mistakes could be magnified. For instance, if a species with brittle wood is planted along a highway, numerous manhours would be spent on pruning or removing risky trees to remedy such avoidable mistake. Moreover, a hierarchical decision-making system can be implemented when making greening-related decisions. Moving up the hierarchy, there would be a greater importance and a larger scale of actions. Higher-level and more comprehensive arboricultural consultancy services would be sought for decisions at the higher ranks in the hierarchy.

3.2. Species richness and biodiversity

Overall, species richness was very low for all sites surveyed in this research. Only 23 tree species were identified (Fig. 4). When all verges and slopes were considered, the mean value at 1.9 species and the median value at 1.0 species (Fig. 3c). Acknowledging that the median was equal to the possible minimum value, monoculture was a commonplace for the surveyed stands. For the verges, the mean and median species richness values were 1.1 and 1.0 species respectively. The special shape of boxplot for the verges somehow reflected the overwhelming occurrence of *C. citriodora* monocultures.

Species richness was mainly made up by slopes because the mean and median species richness were 2.7 and 2.0 species respectively. In fact, slopes had significantly higher mean and median species richness than verges, as demonstrated by *t*-test ($t = 2.13$, $p < 0.05$) and Mann-Whitney U test ($U = 2{,}152$, $p < 0.001$). By inspecting the species composition charts, obviously more species were found in the stands on slopes (Fig. 4). Furthermore, by focusing only on mixed-species stands, the mean and median richness were 3.6 and 3.0 species.

Biodiversity indices were computed. As the calculations of H, D and J were related to species richness, their distribution patterns shared some similarities (Figs. 3c, 3d, 3e, 3f). For all stands, the mean value of H, D and J were 0.15, 0.83 and 0.12 respectively. More extreme values were found on verges, being 0.02, 0.98 and 0.02 correspondingly. The influence of verges on the overall data set was obvious, as both shared the identical median values of the biodiversity indices ($H = 0$, $D = 1$, $J = 1$). Verges contributed to the poor biodiversity of all sites surveyed for this research. Highly significant habitat differences were found in *t*-tests and Mann-Whitney *U* tests comparing the slope and verges. But for the slopes, the mean values were $H = 0.28$, $D = 0.69$, $J = 0.21$, and median values were $H = 0.27$, $D = 0.80$, $J = 0.24$. All differences were confirmed at $p < 0.001$. The composition of slope trees was more diverse ($t = 6.81$, $U = 2{,}168$), less dominated by certain species ($t = -5.65$, $U = 605$) and more even ($t = 6.71$, $U = 2{,}155$) than the trees on the verges. The significance of the habitat-related difference in biodiversity indices was partly due to the extreme values of the verges. Still, the surveyed stands in the present research were characterised by low diversity and evenness, and high dominance overall.

The fact that monospecific stands having $H = 0$ and $J = 0$ was self-explanatory due to the lack of diversity. Also, the complete dominance by a single species, mostly *C. citriodora*, was reflected by $D = 1$. But when mixed-species stands were separated for analysis, the mean values ($H = 0.42$, $D = 0.53$, $J = 0.33$) and median values ($H = 0.44$, $D = 0.43$, $J = 0.34$) were comparable. Generally, most mixed-species stands were found on slopes.

Among the 23 tree species was found, 10 species were regarded as common in Hong Kong (Jim 2008). For transport land use, previous studies also found limited species richness. For example, 11 species were found in verges of South Africa (O'farrell and Milton 2006). It would be easy to locate guidelines and materials for the management of common species. However, the management objectives of roadside green space may evolve with time. For instance, biodiversity improvement has been increasingly valued in a mission to optimse roadside habitats for the provision of ecosystem services (O'Sullivan et al. 2017). Monospecific stands may be converted into mixed-species stands (Juchheim et al. 2020). Nevertheless, clear objective(s) and a management plan must be formulated before putting biodiversity improvement plans in effect. Visual amenity, soil erosion control, noise interception, and various other ecosystem services could be ushered by suitable species selection and target-oriented maintenance. In case of transport land use with trees in great abundance, low-maintenance plantings may be desired in order to meet drivers' expectation (Booze-Daniels et al. 2000; Wolf 2006). If a more diverse selection of tree species is desired, corresponding management guidelines must be extended to cater for less common species. In this research, the species with relatively low frequencies could be surveyed for their health and vigour, which may serve as an indication of their suitability in future use in similar habitats. With sufficient empirical evidence, the tree stock of transport land use may be contain a greater diversity of species in future.

3.3. Associations among habitat type, monoculture status, and tree risk ratings

A highly significant association ($X^2 = 31.90$, $p < 0.001$) was found between habitat type and monoculture status. The primary cause of the association was the concentration of monocultures on verges, whose observed count (48) exceeded the expected (34) (Table 1a). Meanwhile, slopes having a mix of tree species showed a greater observed count (33) than the expected (19). Therefore, verges were associated with monoculture, but slopes with multiple species.

Table 1
Contingency table for the computation of associations between (a) monoculture status and habitat type, (b) monoculture status and tree risk rating, and (c) habitat type and tree risk rating. For the interpretation of Chi-square test results, the expected counts were provided in brackets.

(a)			**Habitat type**		
			Verge	Slope	Total
	Monoculture status	Monospecific	48 (34)	20 (34)	68
		Mixed-species	4 (18)	33 (19)	37
		Total	52	53	105
(b)			Tree risk rating		
			High	Moderate	Total
	Monoculture status	Monospecific	64 (54)	4 (14)	68
		Mixed-species	19 (29)	18 (8)	37
		Total	83	22	105
(c)			Tree risk rating		
			High	Moderate	Total
	Habitat type	Verge	52 (41)	0 (11)	52
		Slope	31 (42)	22 (11)	53
		Total	83	22	105

Another highly significant association was discovered between monoculture status and tree risk rating ($X^2 = 23.94$, $p < 0.001$) (Table 1b). The greatest divergence between observed and expected frequencies was found in the case of mixed-species stand with moderate tree risk, with $O = 18$ and $E = 8$ respectively. At the same time, monoculture had a greater observed count (64) than expected (54). In brief terms, monospecific tree stands were linked to high tree risks, while the tree risk level of mixed-species stand was associate with a moderate rating.

Last but not least, by extending the two significant associations as discussed above, an association between habitat type and tree risk rating was established with comparable significance ($X^2 = 24.86$, $p < 0.001$). (Table 1c). Slopes were associated with moderate tree risk, whose observed frequency (22) doubled the expected one (11). The tree risks of all verges were rated as "high", thus staging a total deviance from the expected count of moderate class (11). The associations elaborated in this section corroborated with each other.

The use of three chi-square tests acted as a triangulation which confirmed that the verges featured monospecific stands with high tree risks. This finding was once again caused by the prevalence of *C. citriodora* whose colossal size would bear serious implications on tree risk management. A target zone is the area which would be affected by the structural failure of a tree (Dunster et al. 2017). The target zone of the failure of a mature *C. citriodora* could stretch over two or three traffic lanes. Damages might be inflicted on vehicles, possibly causing injuries or casualties. Given the 100 km speed limit of San Tin Highway, upcoming vehicles might slam into the damaged vehicles. Verge. Tree risk assessors must closely monitor the health and structural conditions of trees, not just *C. citriodora* but also other large trees.

3.4. Predicting and detecting the presence of monospecific stands

Binomial logistic regression was used for predicting the occurrence of monospecific stands. In the initial model, there were only two significant predictors, namely maximum tree height ($p < 0.01$) and habitat type ($p < 0.05$). Tree abundance and risk rating were insignificant predictors. The log odds were exponentialised for more intuitive interpretation here (Table 2). A metre increase in maximum tree height would bring an increase of 1.22 times in the odds of the occurrence of monoculture. In terms of habitat type, the odds of a verge being monospecific was 5.34 times greater than that of a slope. In more general terms, a verge with tall trees would be expected to be monospecific, such as those with *C. citriodora* in this study.

Table 2
Estimates of log-odds of a unit change in the predictors in the initial logistic regression model for predicting whether a stand would be monospecific or mixed-species. For easier interpretation, the log-odds values were exponentialised. If the converted odds ratio was greater than one, a unit increase in the predictor would increase the probability of a stand being monospecific, vice versa. Predictors which were significant at α = 0.05 were underlined. Habitat and risk rating were categorical variables. The log odds of a stand being on a slope against a verge were presented. The log odds of a moderate risk rating against a high risk rating was displayed.

	Tree abundance	**Maximum tree height**	**Habitat (Slope)**	**Risk rating (Moderate)**
Log-odds	-0.36	0.20	-1.68	-0.93
Odds ratio	0.69	1.22	0.19	0.40

In the present study, boosting enhanced the performance of the binary logistic model to a certain extent (Table 3). With or without boosting, the models exhibited an identical accuracy at 0.81, and Mcnemar's Test p-values exceeding 0.05. However, the boosted model was more reliable as evinced by its higher kappa at $\kappa = 0.63$. In other words, a greater proportion of the correct predictions made by the boosted model could be attributed to the model's capabilities, instead of random chances. Consequently, the boosted model was preferred in predicting the occurrence of monoculture.

Table 3

Model evaluation metrics of the initial and the boosted binomial logistic model predicting whether a stand would be monospecific or mixed-species. As a binomial model, there were only two categorical outputs. Monospecific stand served as the positive class. Thus, model sensitivity represented correct prediction of monospecific stands, whereas specificity reflected correct prediction of mixed-species stands.

	Initial model	Boosted model
Accuracy	0.81	0.81
Kappa	0.51	0.63
Mcnemar's test p-value	0.07	0.13
Sensitivity	1.00	0.69
Specificity	0.44	1.00
Balanced accuracy	0.72	0.85

The accuracy of classifying a stand as monospecific and mixed-species was examined. The initial model was entirely correct in predicting the presence of monoculture, but had noticeable inaccuracies in predicting mixed-species sites (Table 3). This translated into a sensitivity and a specificity at 1.00 and 0.44 respectively. In contrast, the boosted model had remarkable accuracy in predicting mixed-species stands as shown by the specificity of 1.00, but less so for predicting monospecific stands as reflected by a sensitivity of 0.69. Still, in terms of balanced accuracy, the boosted model (0.84) surpassed the initial model (0.72). All in all, boosting, as an ensemble method, raised the predictive accuracy.

Before improving or replacing monocultures, a good detection mechanism would be necessary. The present findings corroborated past studies that variables related to forest structure could predict tree species diversity. For instance, the present study echoed the work by Hakkenberg et al. (2016) that maximum tree height acted as a strong predictor. However, an alternative perspective was provided due to a divergence in the direction of effect of maximum tree height. In past research, greater maximum tree height increased with species diversity (Hakkenberg et al. 2016) and species mixing (Pommerening and Uria-Diez 2017). Contrastingly, according to the modelling results in this research, maximum tree height increased with the probability of the presence of monoculture.

The possible underlying causes of such difference were identified. The effect of maximum tree height on species diversity should be interpreted in tandem with other variables so that a link between maximum tree height and species diversity could be established. No matter how variable the growth rates of different tree species in an urban forest were, tree height would increase with time. Thus, past studies catered for the effects of stand age and history (e.g. Hakkenberg et al. 2016). During forest stand development, new species might move in as a form of spontaneous vegetation growth, while the existing trees accumulated height increments. But such linkages explaining the positive relationship between maximum height and species diversity might not be applicable to the case of the present research.

However, for the present study, all stands were located along the same highway. Site establishment history or age posed no or just little confounding effects because the landscaping works were carried out in the same phase. Instead, as the monospecific stands were dominated by mature *C. citriodra*, whose maximum tree height values were likely to exceed those of mixed-species stands. As a result, the tie between the tall specimens of *C. citriodra* and monospecific stands contributed to the positive coefficient of maximum tree height in the model (Table 2). This research lacked other key tree dimensions such as diameter at breast height or crown spread. But if other dimensional parameters were included, a positive effect on the odds of the occurrence of monospecific stands would be expected in the context of this research.

Based on these relationships, if the species composition of urban forests was explained using habitat characteristics, adequate knowledge of the tree species would be important. In this research, maximum tree height was a significant predictor. In future, advanced techniques which could be used for the estimation of tree height, such as unmanned drones and satellite images can be utilised. Areas with exceptionally tall trees may be feature low species diversity. Tree height mapping could help identify priority zones for tree biodiversity enhancement. Although landscape plans, planting records, management logs, and tree removal records might be available, retrieving and consolidating old records and documents may be time-consuming and inefficient. This study demonstrated how logistic regression can help predict the species composition of urban forest stands.

4. Conclusion

This study was centred on the investigation of monospecific tree stands along a busy highway. 53 slopes and 52 verges were randomly selected along San Tin Highway in Hong Kong. Tree stands on each slope and verge were characterised by abundance, species, maximum height and biodiversity indices. A large number of trees (n = 7,209) was found on the surveyed sites, suggesting the enormous tree stock in transport land use. Significant differences were discovered in the structure and species composition of tree stands between two habitat types, namely verge and slope. Comprehensive results were obtained as both mean and median values were compared. The maximum tree height of the verges exceeded that of the slopes. But the slopes showed higher biodiversity as reflected by the related variables. The use of chi-square tests confirmed the significant association of the verges with higher tree risk and monoculture. The association was attributed to the prevalence of mature *C. citriodora*. The logistic regression model indicated that the likelihood of a monoculture would increase with maximum tree height and verge being the habitat. The divergence in the effects of the predictors against past research could be explained by the tree stock composition observed in this study. Not only did boosting increased the balanced accuracy of the regression model, but also model reliability. In future research, more habitat variables could be collected to improve the detection of monospecific tree stands. Other advanced techniques could be used in tandem to facilitate tree management in urban transport land use.

Declarations

Funding

This work was sponsored by a grant from Highways Department of the Hong Kong Special Administrative Region Government (Project Number: Hy(S)Q/031/2019).

Competing Interests

The third, fourth and fifth authors were employed by the Highways Department of the Hong Kong Special Administrative Region Government, which funded this research.

Availability of data and material

The data used for this study is available upon request.

Code availability

The statistical software, R, used in the statistical analyses for this study is an open-source, free software.

Author Contributions

All authors contributed to the study conception and design. Data analysis was performed by the first author. The manuscript was written by the first author and all authors commented on, read and approved the manuscript. We would like to express our gratitude to Gabriella Leung and Marcus Leung for their efforts in data collection.

Ethics approval

Not applicable

Consent to participate

Not applicable

Consent for publication

Not applicable

References

1. American National Standard (2017) Tree, shrub, and other woody plant management Part 9, Tree Risk Assessment a. Tree Failure (ANSI A300.9:2017). American National Standards Institute, New York

2. Booze-Daniels JN, Daniels WL, Schmidt RE, Krouse JM, Wright DL (2000) Establishment of low maintenance vegetation in highway corridors. Reclam Drastically Disturb Lands 41:887–920. https://doi.org/10.2134/agronmonogr41.c35
3. Bandara WARTW, Dissanayake CTM (2021) Most tolerant roadside tree species for urban settings in humid tropics based on Air Pollution Tolerance Index. Urban Clim 37:100848. https://doi.org/10.1016/j.uclim.2021.100848
4. Bucsuházy K, Zůvala R, Valentová V, Ambros J (2022) Factors related to severe single-vehicle tree crashes: In-depth crash study. PLoS ONE 17(1):e0248171. https://doi.org/10.1371/journal.pone.0248171
5. Casanelles-Abella J, Chauvier Y, Zellweger F, Villiger P, Frey D, Ginzler C, Moretti M, Pellissier L (2021) Applying predictive models to study the ecological properties of urban ecosystems: A case study in Zürich, Switzerland. Landsc Urban Plan 214:104137. https://doi.org/10.1016/j.landurbplan.2021.104137
6. Coppola N, Golombek Y (2018) Urban clear zones, street trees, and road safety. Res Transp Bus Manag 29:136–143. https://doi.org/10.1016/j.rtbm.2018.09.003
7. Divakara BN, Nikhitha CU, Mahmud MA, Nölke N, Tewari VP (2021) Tree Species Diversity in the Southern Transect Across the Rural–Urban Interface of Bengaluru. In: Hoffmann E, Buerkert A, von Cramon-Taubadel S, Umesh KB, Shivaraj PP, Vazhacharickal PJ (eds) The Rural-Urban Interface. Springer, Cham, pp 151–162
8. Dunster JA, Smiley ET, Matheny N, Lilly S (2017) Tree risk assessment manual, 2nd edn. International Society of Arboriculture, Champaign
9. Felton A, Nilsson U, Sonesson J, Felton AM, Roberge JM, Ranius T, Ahlström M, Bergh J, Björkman C, Boberg J, Drössler L, Fahlvik N, Gong P, Holmström E, Keskitalo ECH, Klapwijk MJ, Laudon H, Lundmark T, Niklasson M, Nordin A, Pettersson M, Stenlid J, Sténs A, Wallertz K (2016) Replacing monocultures with mixed-species stands: Ecosystem service implications of two production forest alternatives in Sweden. Ambio 45:124–139. https://doi.org/10.1007/s13280-015-0749-2
10. Fassnacht FE, Hartig F, Latifi H, Berger C, Hernández J, Corvalán P, Koch B (2014) Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. Remote Sens Environ 154:102–114. https://doi.org/10.1016/j.rse.2014.07.028
11. Hakkenberg CR, Song C, Peet RK, White PS (2016) Forest structure as a predictor of tree species diversity in the North Carolina Piedmont. J Veg Sci 27:1151–1163. https://doi.org/10.1111/jvs.12451
12. Hale DC, Morzillo AT (2020) Landscape characteristics and social factors influencing attitudes toward roadside vegetation management. Landsc Ecol 35(9):2029–2044. https://doi.org/10.1007/s10980-020-01078-6
13. Hong Kong Herbarium, South China Botanical Garden (2007) Flora of Hong Kong, vol 1. HKSARG Agriculture, Fisheries and Conservation Department, Hong Kong

14. Hong Kong Herbarium, South China Botanical Garden (2008) Flora of Hong Kong, Volume 2. HKSARG Agriculture, Fisheries and Conservation Department, Hong Kong
15. Hong Kong Herbarium, South China Botanical Garden (2009) Flora of Hong Kong, vol 3. HKSARG Agriculture, Fisheries and Conservation Department, Hong Kong
16. Hong Kong Herbarium, South China Botanical Garden (2011) Flora of Hong Kong, vol 4. HKSARG Agriculture, Fisheries and Conservation Department, Hong Kong
17. Hong Kong Herbarium (2012) Check List of Hong Kong Plants. HKSARG Agriculture, Fisheries and Conservation Department, Hong Kong
18. HKSARG Census and Statistic Department (2022) Population estimates. Census and Statistic Department. https://www.censtatd.gov.hk/en/scode150.html. Accessed 8 April 2022
19. HKSARG Development Bureau (2012) Allocation of space for quality greening on roads: TC(W) No.2/2012. HKSARG Development Bureau, Hong Kong
20. HKSARG Development Bureau (2018) Soft landscape provisions for highway structures: TC(W) No.1/2018. HKSARG Development Bureau, Hong Kong
21. HKSARG Development Bureau (2022) Guidelines for tree risk assessment and management arrangement. Development Bureau. https://www.greening.gov.hk/en/tree-care/tree-risk-assessment-and-management-arrangement. Accessed 8 April 2022
22. HKSARG Highways Department (2022) Road network. Highways Department https://www.hyd.gov.hk/en/road_and_railway/road_network. Accessed 8 April 2022
23. HKSARG Planning Department (2021) Land utilization in Hong Kong. Planning Department. https://www.pland.gov.hk/pland_en/info_serv/statistic. Accessed 8 April 2022
24. HKSARG Lands Department (2022) Determination of slope maintenance responsibility. Lands Department. https://www.landsd.gov.hk/en/land-mgt-enforce/slope-maintenance-responsibility. Accessed 8 April 2022
25. HKSARG Transport Department (2020) Road users' code. HKSARG Transport Department, Hong Kong
26. Huang X, Teng M, Zhou Z, Wang P, Dian Y, Wu C (2021) Linking naturalness and quality improvement of monoculture plantations in urban area: a case study in Wuhan city, China. Urban For Urban Green 59:126911. https://doi.org/10.1016/j.ufug.2020.126911
27. Jim CY (1989) The distribution and configuration of tree cover in urban Hong Kong. GeoJournal 18(2):175–188
28. Jim CY (1990) Trees in Hong Kong: species for landscape planting. Hong Kong University Press
29. Jim CY (2008) Multipurpose census methodology to assess urban forest structure in Hong Kong. Arboric Urban For 34:366–378. https://doi.org/10.1007/BF01207091
30. Juchheim J, Ehbrecht M, Schall P, Ammer C, Seidel D (2020) Effect of tree species mixing on stand structural complexity. Forestry 93:75–83. https://doi.org/10.1093/forestry/cpz046

31. Kaighn RJ, Yu SL (1996) Testing of roadside vegetation for highway runoff pollutant removal. Transp Res Rec 1523(1):116–123. https://doi.org/10.1177/0361198196152300114
32. Kassambara A(2020) ggpubr: ‘ggplot2’ based publication ready plots (Version 0.4-0). CRAN. https://CRAN.R-project.org/package=ggpubr. Accessed 26 March 2022
33. Kendal D, Dobbs C, Lohr VI (2014) Global patterns of diversity in the urban forest: Is there evidence to support the 10/20/30 rule? Urban For Urban Green 13:411–417. https://doi.org/10.1016/j.ufug.2014.04.004
34. Kuhn M (2008) Building Predictive Models in R Using the caret Package. J Stat Softw 28:1–26. https://doi.org/10.18637/jss.v028.i05
35. Kuhn M(2016) caret: Classification and regression training (Version 6.0–71). CRAN. https://CRAN.R-project.org/package=caret. Accessed 26 March 2022
36. Lang AC, Härdtle W, Bruelheide H, Geißler C, Nadrowski K, Schuldt A, Yu M, von Oheimb G (2010) Tree morphology responds to neighbourhood competition and slope in species-rich forests of subtropical China. For Ecol Manag 260:1708–1715. https://doi.org/10.1016/j.foreco.2010.08.015
37. Lee LSH(2022) Quantitative tools for the prediction of pavement damages associated with urban trees.Arboric Urban For48 (in press)
38. Lee LSH, Jim CY, Zhang H (2019) Tree density and diversity in Hong Kong’s public housing estates: From provision injustice to socio-ecological inclusiveness. Urban For Urban Green 46:126468. https://doi.org/10.1016/j.ufug.2019.126468
39. Lee LSH, Jim CY, Zhang H (2021) Serviceable tree volume: An alternative tool to assess ecosystem services provided by ornamental trees in urban forests. Urban For Urban Green 59:127003. https://doi.org/10.1016/j.ufug.2021.127003
40. Liu J, Slik F (2022) Are street trees friendly to biodiversity? Landsc Urban Plan 218:104304. https://doi.org/10.1016/j.landurbplan.2021.104304
41. Lugo-Pérez J, Sabat-Guérnica AM (2011) Structure and composition of woody plants in urban forest remnants with different adjacent land-use and slope aspect. Urban Ecosyst 14:45–58. https://doi.org/10.1007/s11252-010-0139-2
42. Ma B, Hauer RJ, Wei H, Koeser AK, Peterson W, Simons K, Timilsina N, Werner LP, Xu C (2020) An assessment of street tree diversity: findings and implications in the United States. Urban For Urban Green 56:126826. https://doi.org/10.1016/j.ufug.2020.126826
43. Marshall AJ, Grose MJ, Williams NS (2019) From little things: More than a third of public green space is road verge. Urban For Urban Green 44:126423. https://doi.org/10.1016/j.ufug.2019.126423
44. Modlingerová V, Száková J, Sysalová J, Tlustoš P (2012) The effect of intensive traffic on soil and vegetation risk element contents as affected by the distance from a highway. Plant Soil Environ 58(8):379–384
45. Mu Y, Lin W, Diao X, Zhang Z, Wang J, Lu Z, Guo W, Wang Y, Hu C, Zhao C (2022) Implementation of the visual aesthetic quality of slope forest autumn color change into the configuration of tree species. Sci Rep 12:1–19. https://doi.org/10.1038/s41598-021-04317-1

46. O’farrell PJ, Milton SJ (2006) Road verge and rangeland plant communities in the southern Karoo: exploring what influences diversity, dominance and cover. Biodivers Conserv 15:921–938. https://doi.org/10.1007/s10531-004-3102-9
47. O'Sullivan OS, Holt AR, Warren PH, Evans KL (2017) Optimising UK urban road verge contributions to biodiversity and ecosystem services with cost-effective management. J Environ Manage 191:162–171. https://doi.org/10.1016/j.jenvman.2016.12.062
48. Orlóci L, Stanek W (1980) Vegetation survey of the Alaska Highway, Yukon Territory: types and gradients. Vegetatio 41:1–56. https://doi.org/10.1007/BF00055301
49. Pommerening A, Uria-Diez J (2017) Do large forest trees tend towards high species mingling? Ecol Inf 42:139–147. https://doi.org/10.1016/j.ecoinf.2017.10.009
50. Pretzsch H, Biber P, Uhl E, Dahlhausen J, Rötzer T, Caldentey J, Koike T, van Con T, Chavanne A, Seifert T, du Toit B, Farnden C, Pauleit S (2015) Crown size and growing space requirement of common tree species in urban centres, parks, and forests. Urban For Urban Green 14:466–479. https://doi.org/10.1016/j.ufug.2015.04.006
51. RStudio Team (2019) RStudio: Integrated Development for R. RStudio, Inc. http://www.rstudio.com. Accessed 24 March 2022
52. Sever-Mutlu S, Selim C, Ün G (2017) Plant biodiversity of urban roadside trees in Antalya, Turkey. Kastamonu Univ J For Fac 17:80–87. https://doi.org/10.17475/kastorman.296501
53. Shataee S, Kalbi S, Fallah A, Pelz D (2012) Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. Int J Remote Sens 33:6254–6280. https://doi.org/10.1080/01431161.2012.682661
54. Smiley ET, Matheny NP, Lilly SJ (2011) Best management practices: Tree risk assessment. International Society of Arboriculture, Champaign
55. Soga M, Yamaura Y, Koike S, Gaston KJ (2014) Land sharing vs. land sparing: does the compact city reconcile urban development and biodiversity conservation? J Appl Ecol 51:1378–1386. https://doi.org/10.1111/1365-2664.12280
56. Song H, Jeon G, Lee S, Kim N, Park G, Lee B (2005) Vegetation structure and succession of highway cutting-slope area. J Korean Soc Environ Eng 8(6):69–79
57. Torgo L (2010) Data mining using R: Learning with case studies. CRC Press, Boca Raton
58. Tuszynski J(2021) caTools: Moving window statistics, GIF, Base64, ROC, AUC, etc (Version 1.18.2). CRAN. https://CRAN.R-project.org/package=caret. Accessed 26 March 2022
59. Volin E, Ellis A, Hirabayashi S, Maco S, Nowak DJ, Parent J, Fahey RT (2020) Assessing macro-scale patterns in urban tree canopy and inequality. Urban For Urban Green 55:126818. https://doi.org/10.1016/j.ufug.2020.126818
60. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, Grolemund AH, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the Tidyverse. J Open Source Softw 4(43):1686. https://doi.org/10.21105/joss.01686

61. Wolf KL (2006) Assessing public response to freeway roadsides: urban forestry and context-Sensitive solutions. Transp Res Rec 1984(1):102–111. https://doi.org/10.1177/0361198106198400110
62. Yang J, Luo X, Lu S, Yang Y, Yang J (2021a) Effects of compositional and configurational heterogeneity of the urban matrix on the species richness of woody plants in urban remnant forest patches. Landsc Ecol 37:619–632. https://doi.org/10.1007/s10980-021-01368-7
63. Yang J, Yang J, Xing D, Luo X, Lu S, Huang C, Hahs AK (2021b) Impacts of the remnant sizes, forest types, and landscape patterns of surrounding areas on woody plant diversity of urban remnant forest patches. Urban Ecosyst 24:345–354. https://doi.org/10.1007/s11252-020-01040-z
64. Zeng SL, Zhang TT, Gao Y, Ouyang ZT, Chen JK, Li B, Zhao B (2011) Effects of road age and distance on plant biodiversity: a case study in the Yellow River Delta of China. Plant Ecol 212:1213–1229. https://doi.org/10.1007/s11258-011-9899-x
65. Zheng J, Zang H, Yin S, Sun N, Zhu P, Han Y, Kan H, Liu C (2018) Modeling height-diameter relationship for artificial monoculture Metasequoia glyptostroboides in sub-tropic coastal megacity Shanghai, China. Urban For Urban Green 34:226–232. https://doi.org/10.1016/j.ufug.2018.06.006
66. Źróbek-Sokolnik A, Dynowski P, Źróbek S (2021) Preservation and Restoration of Roadside Tree Alleys in Line with Sustainable Development Principles—Mission (Im) possible? Sustainability 13:9635. https://doi.org/10.3390/su13179635

Figures

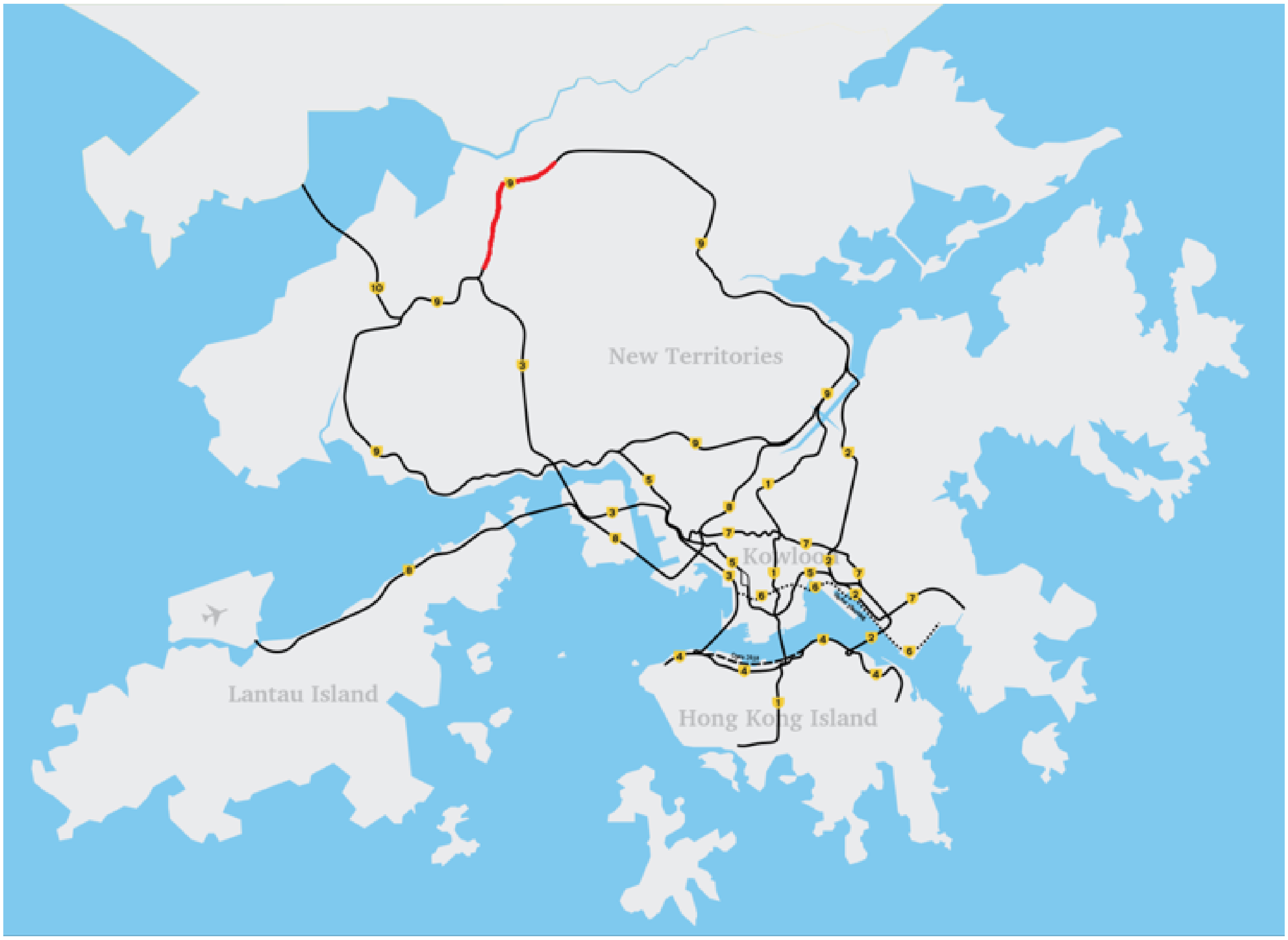

Figure 1

Hong Kong Strategic Route System. The location of this research, namely San Tin Highway, was highlighted in red. This figure was adapted from an open-source map (https://en.wikipedia.org/wiki/Route_9_(Hong_Kong))

Figure 2

Photographs showing an example of the (a) verges and (b) slopes selected in this research. The photographs were taken on 3rd April, 2021 and 31st May, 2021 respectively

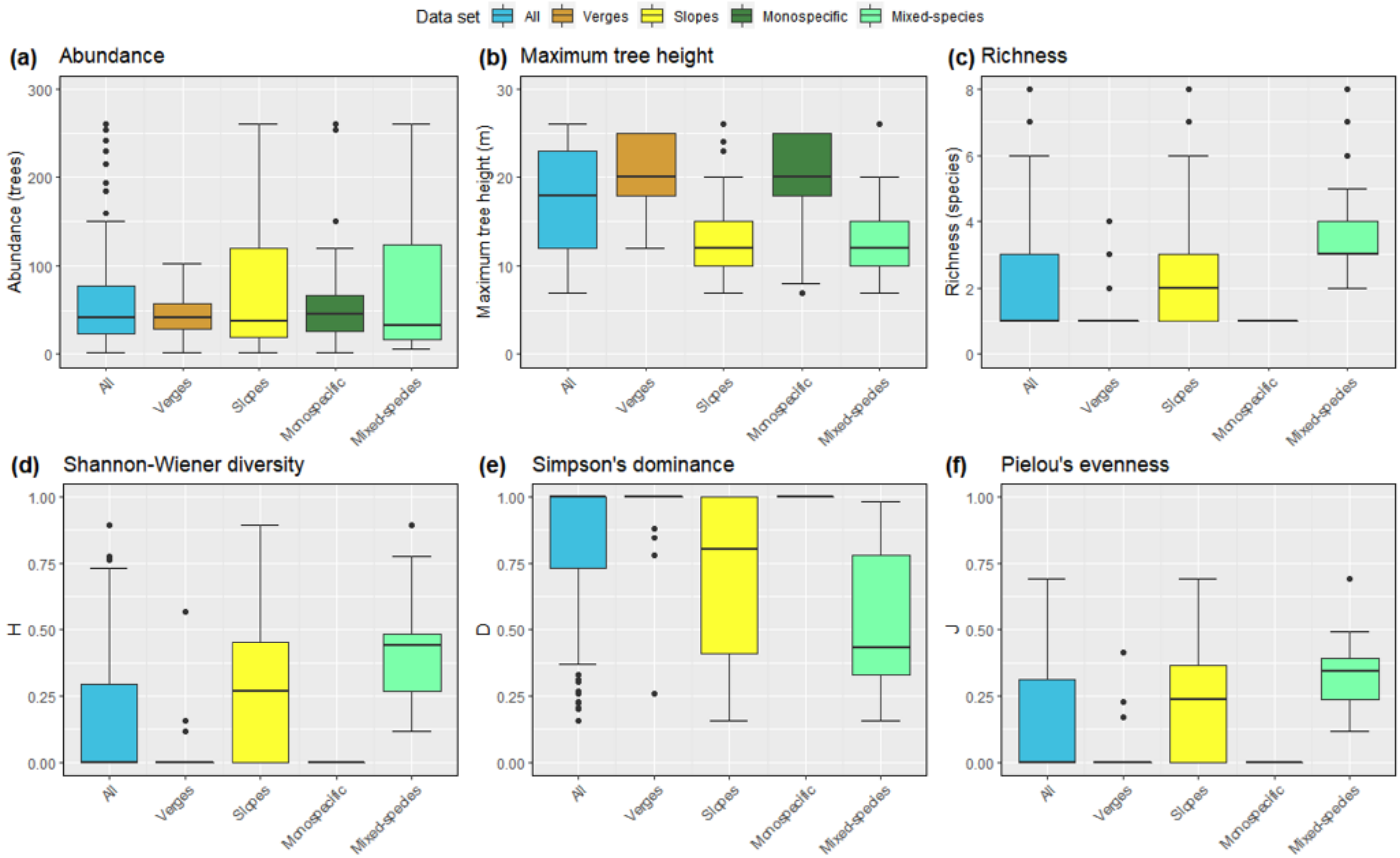

Figure 3

Boxplots showing the distribution of (a) abundance, (b) maximum tree height, (c) species richness, (d) Shannon-Wiener diversity, (e) Simpson's dominance, and (f) Pielou's evenness. In each graph, from left to right, the boxes were constructed using data from all stands (*n* = 105), verges (*n* = 52), slopes (*n* = 53), monospecific stands (*n* = 68), and mixed-species stands (*n* = 37). The box boundaries were defined by the inter-quartile range (IQR). The horizontal line inside each box, and the upper and lower whiskers indicated median, maximum and minimum values. Any values outside 1.5 × IQR were outliers and visualised by dots.

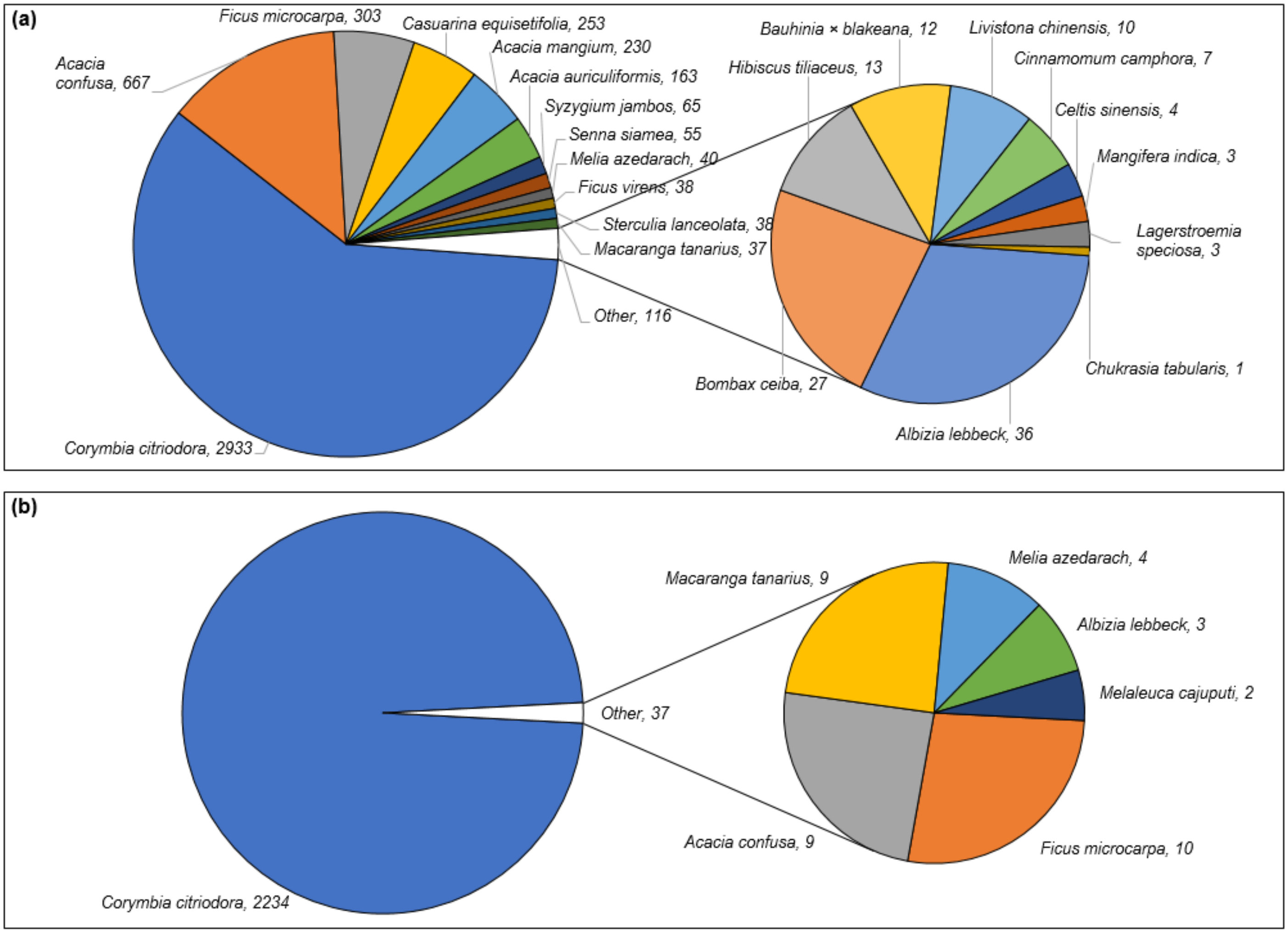

Figure 4

Pie-in-pie charts depicting the frequency of tree species of (a) the slopes and (b) the verges sampled in this research. There were 4,938 trees on the slopes versus 2,271 trees on the verges. The pies placed on the right were an extension of the slice resembling the species with relatively low frequency in main pie