JOURNAL OF
EMERGING INVESTIGATORS

Article

Mathematical modeling of plant community composition for urban greenery plans

Yueqi Fang[1], Weibo Zhang[1]
[1] Dalton Academy, The Affiliated High School of Peking University, Beijing, China

SUMMARY

Urban green space is critical for humans and other organisms living in cities. Researchers have noticed that urban green space planning has been focusing more on ornamental and public management purposes but less on its ecological purpose, which has a negative impact on eco-sustainability. We hypothesized that an urban greenery management plan with high eco-sustainability can be achieved by calculating the composition of a plant community through mathematical modeling. To test our hypothesis, we created a nonlinear programming function that generated a plant community composition aiming for high plant fitness, high biodiversity, and low managing cost. Green fields in a Beijing city park, comprising different combinations of plant species were used to optimize our model. Although the outputs of the optimized model do not match completely with the composition of testing fields, they illustrate possible patterns for plant community development. By applying this model, urban green space can be finely designed to become both ecologically and economically sustainable.

INTRODUCTION

Biodiversity is of crucial importance for an ecosystem (1). Studies show that increasing biodiversity in an ecosystem when the overall biodiversity is low can improve the stability of that ecosystem (1-3). Plant species diversity influences the biodiversity of other organisms such as birds, insects, and fungi (4-6). Plant species diversity is fundamental to the community's biodiversity as a whole (7,8). Therefore, it is crucial for green spots to have high plant species diversities. With the continuous advancement of China's urbanization process, the natural habitat in the rapidly expanding urban areas has decreased. The increased urban ecological burden threatens the urban ecosystem and the well-being of residents. Therefore, in recent years, the urban ecological environment construction has received widespread attention, and particularly, green space management have been strengthened. However, there are still many problems in urban greenery management that need to be addressed scientifically and economically, e.g., the green space planning and layout, green space composition, integration of urban and rural greening, and local biodiversity protection.

Several studies evaluated the species richness of artificial green spaces in Beijing city (9,10). Hou *et al.* indicated that the proportion of exotic plant species, mainly ornamental plants, accounted for 38.95% of the overall plant species in the sampled Beijing urban built-up areas (9). The proportion was more than three times higher than that of the exotic plant species in the nearest mountain area. In particular, a large number of exotic species were found of single genus and single family in the sampled windows, which made the characteristics of plants and their distribution areas deviate from the local natural conditions (9). It may thus lead to potential issues, ranging from an increased cost for maintenance of the local and introduced plant species to increased difficulties for protection of local species, ultimately affecting the stability of urban ecosystems.

Considering the various ecological and artificial factors an urban greenery management should be designed with scientifically clear and solid guidance. However, current urban greenery management guidance is ambiguous and insufficient to serve that purpose (11). Compared to other strategies, mathematical modeling strategy bears its advantage on assessing the biodiversity of a plant community quantitatively. Existing mathematical models used in mountain regions proved the feasibility of using this strategy in green space planning, although those models cannot be directly applied to urban regions because a higher proportion of trees are grown in mountains than in cities (12-15). This is likely because trees are considered to have higher ecological service value than herbs and shrubs, but there are very few natural spaces available for tree planting in cities (9). Nevertheless, many herbal species can be grown in cities, and these species would also significantly contribute to the city's biodiversity (9,16-17). It is important to develop a more precise command of planning urban green spots through mathematical modeling to achieve high eco-sustainability. By building a nonlinear programming model whose feasibility and accuracy was evaluated and optimized using testing fields in a Beijing city park, the plant compositions for urban greenery management plans with high biodiversity and low cost can be achieved.

RESULTS

Model Establishment

The ultimate goal of our model was to optimize biodiversity, plant fitness level, and managing costs together and thus generating a plan with high biodiversity and fitness, and low managing costs. To achieve this goal, we created a nonlinear programming model that optimized the comprehensive influence of plant fitness, biodiversity, and managing cost. We defined the plant fitness index (*ft*) as reflecting on how well all plants could grow on a specific piece of land, which ensures the basic assumption that the plant species selected by the model could prosper in the community. The biodiversity index (*BI*) described the biodiversity level of a community. We used the Simpson index as our BI. At the same time, because this research focused on the biodiversity of plant species, *BI* was calculated using the plants' importance values, which is the measure of how dominant each species is on a given field, reflecting on the amount of resources each species gains (13). The cost index (*ct*) inversely evaluates the total monetary cost of seeds/seedlings of the target field. The goal of our model was to have a maximum *ft*, *BI*, and *ct*.

The fitness of a plant describes the condition of this plant growing on a specific piece of land. In this way, we defined fitness as the adjacence of plant and land trait parameters. We selected light preference and moisture preference to evaluate the fitness of the plant species because they were crucial to plant growth as being key factors in photosynthesis. We obtained plant parameters from the Ellenberg index.

For land parameters, we defined the land light value (*LV*) as the light intensity level of that piece of land. The available light on a piece of urban land is associated with the buildings around it. Therefore, we defined *LV* of a piece of land as the proportion of the area that receives sunlight for a period of time. We used integral equations in the model to calculate the *LV* of a given area in a certain location with one, two, or three buildings around it. The result *LV* is an index in the interval of [0,1] which is positively related to the amount of light that land could receive (see **Materials and Methods**).

The moisture level (**MV**) of a city's land is associated with three factors, e.g., the local climate, the irrigation systems, and natural water sources nearby. None of the testing fields have irrigation equipment. The Type I testing fields are located within 10 meters of a river, and the Type III testing fields are located within 100 meters of a river. Based on the three factors of the testing fields, the *MV* was set as 0.4 for Type I, 0.2 for Type II, and 0.3 for Type III. However, to avoid generating extreme values (i.e., zeros) in calculating *ft* when *MV* is subtracted from the plant's moisture value (see Equation 2), the MV for Type I and II were slightly modified (i.e., adds or subtracts 0.01). In this way, the ultimate *MV* for Type I was set as 0.41, and the ultimate *MV* for Type II was set as 0.19.

Since both the plant trait parameters and the land trait parameters can be written as a binary vector, and all the elements in these vectors fit into the interval of [0,1], we used Euclidean distance to evaluate the likeliness between these vectors:

$$Distance = \sqrt{(a_i - a_j)^2 + (b_i - b_j)^2} \qquad \text{[Eqn 1]}$$

where a_i and a_j are the first elements of two different vectors, and bi and bj are the second elements of two vectors. Therefore, the fitness (*ft*), positively related to the plant species' fitness level, was calculated by:

$$ft = \sum_i^n \left(\frac{x_i * cvr_i}{A} * \frac{\sqrt{2} - \sqrt{(LI_i - LV)^2 + (WA_i - MV)^2}}{\sqrt{2}} \right) \qquad \text{[Eqn 2]}$$

in which *n* is the number of plant species, x_i is the population of the ith species, cvr_i is the unit coverage of the i^{th} species (i.e., the coverage of a single plant in dm^2), *A* is the total area of the target field (in dm^2), LI_i is the light value of the i^{th} species, WA_i is the moisture value of the ith species, *LV* is the land light value, and *MV* is the land moisture value.

In order to have our *ct* inversely related to the total cost of seed/seedlings of plants, we calculated it as:

$$ct = 1 - \frac{\sum_i^n x_i * c_i}{A * \max\left(\frac{\{c_i\}}{\{cvr_i\}}\right)} \qquad \text{[Eqn 3]}$$

in which x_i is the population of the i^{th} plant species, c_i is the unit cost of the i^{th} species (in CNY), *A* is the total area of the target field (dm^2), and $max(\frac{\{c_i\}}{\{cvr_i\}})$ is the highest unit cost-per-area among all plant species within the library (in CNY/dm^2). By doing so, the fraction in the equation will be constantly smaller than 1. By subtracting the fraction from 1, we ensured that the resulting *ct* will decrease as the total cost (i.e., the numerator in the equation) increases.

We used nonlinear programming as the main function of our model. Since the three parameters (i.e., *ft*, *BI*, and *ct*) are independent of each other, we constructed our objective function by multiplying these three indices. The resulting nonlinear programming model is:

$$Max\ Z = ft^n * \sqrt[n]{BI} * ct \qquad \text{[Eqn 4]}$$

$$Restrictions: \begin{cases} \sum_{i=1}^n x_i * cvr_i = A \\ x_i \geq 0 \end{cases}$$

where *n* is the weight of *ft* and *BI*, a real number in the range of [1,4] whose range is chosen for the simplicity of the model and also to ensure that the difference between the weight of *ft* and *BI* would not be too significant, and "*Max Z*" is a representation of the maximization objective function for the nonlinear programming. All three indexes in Equation 4 are associated with x_i – the population of the i^{th} species – which are the decision variables of this nonlinear programming. In the restrictions, cvr_i is the unit coverage of the ith species (in dm^2), and *A* is the total area of the target field (in dm^2). The first linear restriction ensures that the total coverage of all plant species will be equal to the land area, which means that the entire land will be covered by plants. The initial value for *n* in testing is 1. This nonlinear programming model will result in a row vector composed of all x_i that generates the maximum value of *Z*. The cross product of this vector with the coverage vector (i.e., a column vector composed of all cvr_i) is a plant composition vector that provides the suggested coverage of each plant species in the target field.

Due to the requirement of the R package Rdonlp2 (version 3042.11/r6080) (18) for nonlinear programming, we used 10 as the initial iteration value for all x_i, ensuring that the initial total coverage of the plant community will not exceed the field area and avoid local optimization. The overall input for this model is the light and moisture information of both the target field and plant species, and the overall output is the population of all plant species (i.e., $\{x_i\}$) that achieve a maximum value for the objective function. When applying this model, we sorted plants into three categories – arbors, shrubs, and herbs – based on their different niches. Arbors are separated from herbs and shrubs when establishing the objective function, and both share the same land area. This division is practical because shrubs and herbs are able to grow under most of the arbors.

To simplify the model testing process, we considered an ideal situation to restrain the number of variables involved in the equation. In an ideal situation, the plant community with the highest eco-sustainability should be comparable to the plant community in a well-growing wild area (i.e., we defined well-growing as the plant community does not have a great disturbance in the duration for its herbal populations to undergo a complete life cycle). These areas require no or little human management, and the seed and seedling costs could be neglected. Due to this neglection, the main function of the nonlinear programming was modified to Equation 5:

$$Max\ Z = ft^n * \sqrt[n]{BI} \qquad \text{[Eqn 5]}$$

Model Testing

In order to assess the feasibility and accuracy of the simplified model, we compared modeled and factual plant compositions in real urban green spaces mimicking well-growing wild plant communities. We chose Yuan Ming Yuan, a previous royal garden built in the Qing Dynasty and currently a historic park occupying a total area of 864.9 acres in the Haidian District of Beijing City, for its varied types of urban green space with scarce long-term human management. We selected a total of six testing fields in Yuan Ming Yuan with shrubs and herbs, the area (*A*) of which being roughly 750 dm^2 individually. The testing fields have not been cared for by humans for years, making them ideal representations of urban plant communities mimicking wild communities. The testing fields (q1-6) were classified into three groups based on their distinct light and moisture levels (**Table 1**). Shades in the light level of q1,2,5,6 were calculated with measured lengths and heights of the walls and buildings surrounding the respective fields (**Figure 1 and 2**). The field records of the plant compositions were described by the importance value of each species (**Figure 3**). There are shared characteristics of the testing fields within each field types. *Crepidiastrum*, *Ophiopogon*, *Oplismenus*, and *Viola* are genera shared by q1 and 2. Among the four genera, *Ophiopogon* and *Oplismenus* combined formed the dominant plant groups in these quadrats. *Amaranthus*, *Portulaca*, *Potentilla*, and *Setaria* are genera shared by q3 and 4, with *Setaria* dominating in both quadrats. *Glechoma* and *Setaria* are genera shared by q5 and 6, with *Glechoma* being the dominant genus.

To test the accuracy of the modeling strategy, we established a plant database including information of the plant type, height, coverage, light value, moisture value, and unit costs for the 16 plant species found in the six testing fields (**Table 2**). We fed database information into Equation 5 to generate one simulation result for each field type, e.g., q7 for field type I, q8 for field type II, and q9 for field type III (**Figure 4**). To train the mathematical model to generate the most satisfying proximity to factual results, we changed n values in the range of [1,4] (*n* = 1, 2, 3, 3.5, 4) in Equation 5 to modify the contribution of biodiversity and fitness for the simulating plant community, and we conducted Two-Way Indicator Species Analysis (TWINSPAN) between modeled results with the testing field records (**Figure 4**). Increasing the

Light & moisture level	Testing field
Light tree shade; moist *LV* = 0.75; *MV* = 0.41	q1, q2: Type I
No shade; dry *LV* = 1; *MV* = 0.19	q3, q4: Type II
Heavy shade; slightly moist *LV* = 0.28; *MV* = 0.30	q5, q6: Type III

Table 1: Testing fields and their corresponding light & moisture levels. q1-6 are the names of the six testing fields belonging to field types I-III based on light and moisture levels. *LV* represents the light value of each field and *MV* represents the moisture value of each field.

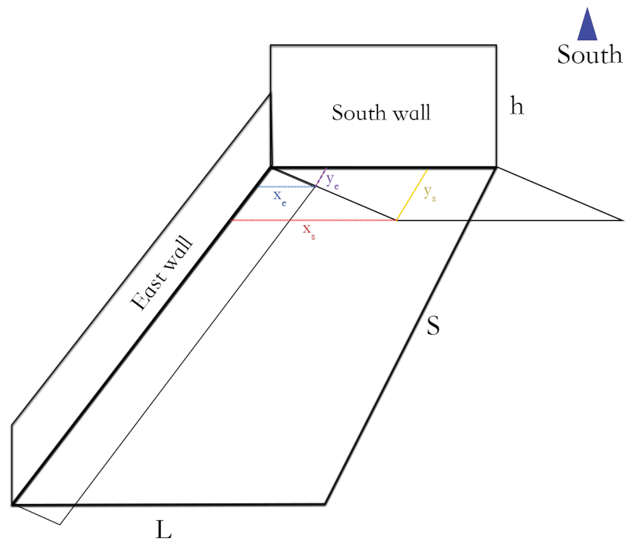

Figure 1: A land with two buildings. The center rectangle in the figure represents a piece of city land with an east wall and a south wall. x_e is the height of the east wall's shade (EWS), y_e is the distance between the EWS and the south wall, y_s is the height of the south wall's shade (SWS), x_s is the height of the SWS, *h* is the height of the wall, and *L* and *S* are measurements of the length of two sides of the land. All lengths are in decimeters.

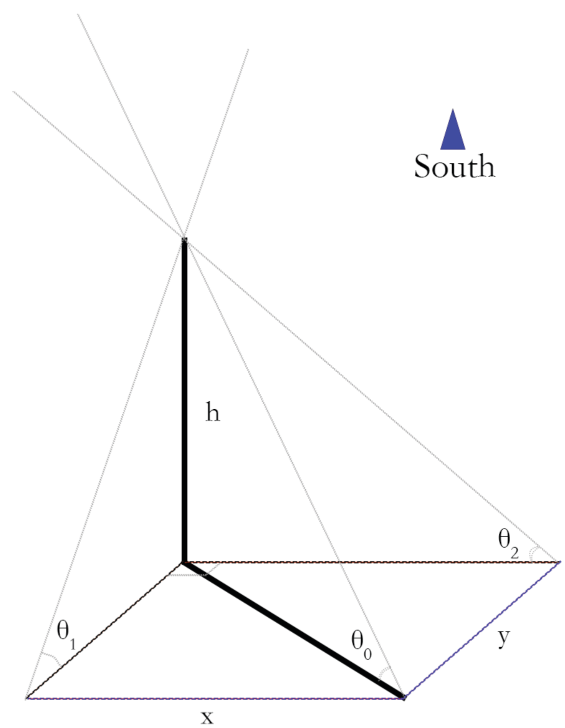

Figure 2: A rod and its shadow. The perpendicular bold line represents a rod, and the inclined bold line represents its shadow. The two thin, black dash lines are the length of the shadow's north-south (*y*) and east-west (*x*) components, respectively. θ_0 is the solar altitude at this moment, while θ_1 and θ_2 are the altitude angle of the shadow's two components. All lengths are in decimeters, and all angles are radians.

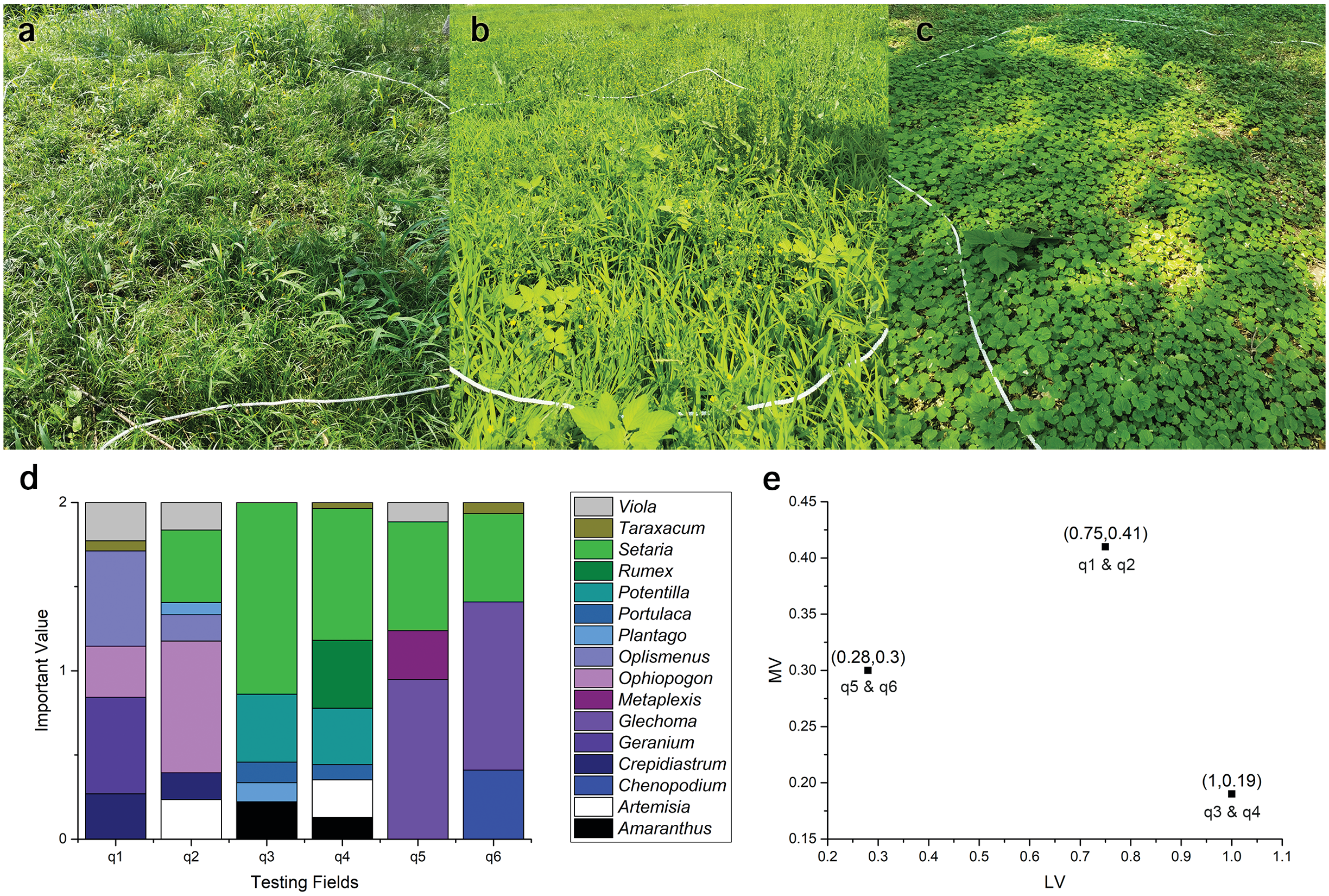

Figure 3: Plant community of the testing fields. Representative plant communities were pictured for **(a)** q2, **(b)** q4, and **(c)** q6, respectively. **(d)** The importance value of each plant species (identified to genera) in six testing fields collected from different quadrats in Yuan Ming Yuan. Importance values were calculated using the average height and coverages of each plant. **(e)** The *LV* and *MV* for each quadrat.

n value in Equation 5 increased the impact of the fitness index and lowered the impact of the biodiversity index. Among the five models adapted with *n* values, the modeled result with *n* = 3.5 showed the most proximal simulation with the factual results based on TWINSPAN clustering of their field types. Specifically, q9, the modeled result of field type III with *n* = 3.5 shared the same dominant genera *Glechoma* with the two field records of type III, namely q5 and q6 (**Figure 5**). Additionally, in this case, the two modeling quadrats both contained genera *Setaria*, *Potentilla*, and *Geranium* (**Figure 5**). Therefore, although q7 and q8 were inversely grouped with their corresponding types, *n* = 3.5 was still considered to be a suitable value. After the n value was determined using ideal and simplified scenarios, the initial equation which models common urban green space can thus be modified as:

$$Max\ Z = ft^{3.5} * \sqrt[3.5]{BI} * ct \qquad \text{[Eqn 6]}$$

DISCUSSION

We used the nonlinear mathematical modeling strategy to generate a plan to achieve high eco-sustainability of the urban plant community. This approach is different from current biodiversity indices (e.g., beta diversity indices) since the latter can only assess the biodiversity of an ecosystem, while our model can provide practical plans for managers to schedule a green spot. Mathematically, the objective function of our model is positively related to the plant fitness and biodiversity index, and negatively related to management costs. The equation of the model was optimized for a range of n values by comparing simulated plant compositions with field records in selected testing fields in Yuan Ming Yuan, a city park in Beijing.

We documented the plant communities on the six testing fields (**Figure 3d**). The shared characteristics in dominant plants observed in each field type suggest a strong growth advantage of certain plants under specific light and moisture conditions. However, no communities in the two testing fields per field type showed the same plant composition, suggesting that factors other than light and moisture could also influence a plant community's composition, and that there could be various kinds of plant composition under the same light and moisture conditions.

The best-fit model – with Equation 6 as its objective function – heavily weighs plant fitness to biodiversity. We reasoned that these uneven weights appear because plant fitness could be the precondition that decides whether a plant species could prosper on a piece of land, while community biodiversity is the outcome of plant species colonization. In another word, plant fitness determines whether a species could survive on a field in the first place, while biodiversity only

Genus	Type	Height (dm)	Coverage (dm^2)	Light value	Moisture value	Unit cost (CNY)
Amaranthus	herb	50	2	8	5	0.068
Artemisia	herb	90	2	8	5	0.004
Chenopodium	herb	90	1	7	5	0.07
Crepidiastrum	herb	60	1	7	6	0.083
Geranium	herb	55	1	7	4	0.068
Glechoma	herb	15	1	6	5	0.029
Metaplexis	herb	50	5	7	7	0.08
Ophiopogon	herb	60	2	7	5	0.06
Oplismenus	herb	35	2	7	4	0.001
Plantago	herb	25	1	8	5	0.0025
Portulaca	herb	13	1	7	4	0.059
Potentilla	herb	35	1	8	4	0.01
Rumex	herb	120	2	7	7	0.07
Setaria	herb	110	1	7	4	0.00072
Taraxacum	herb	12	1	8	6	0.003
Viola	herb	6.5	1	7	5	0.3

Table 2: Plant database for the mathematical model testing. Average height, coverage, Ellenberg indexes of light value and moisture value, and unit seed/seedling costs of each plant.

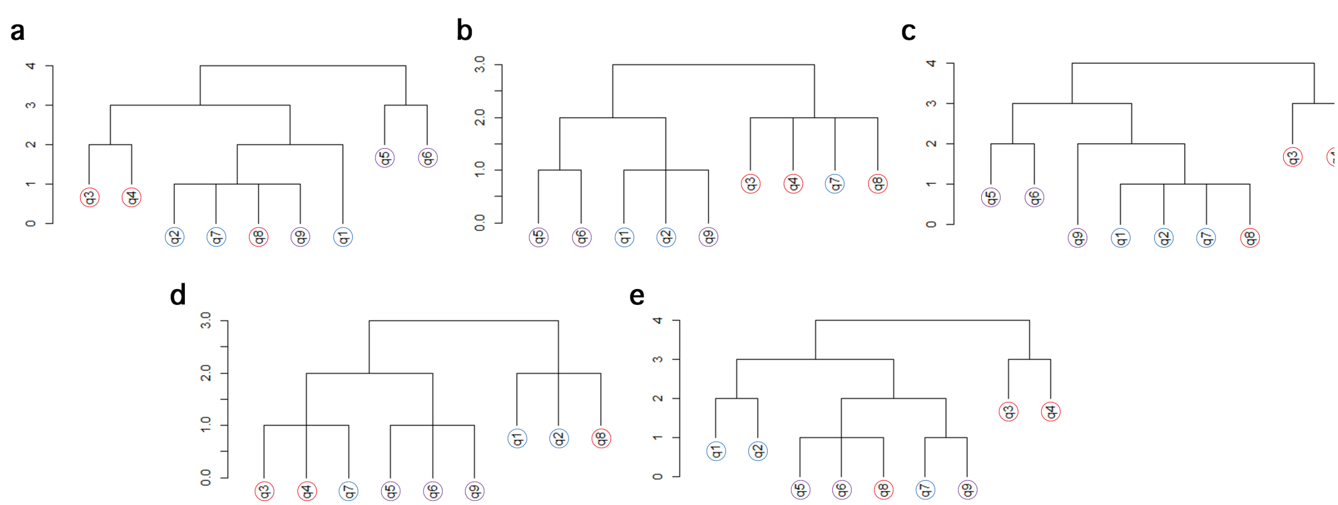

Figure 4: TWINSPAN evaluation for Equation 5 with different *n* values. The simulated plant community for field type I-III (q7-9) were clustered with plant communities in the six testing fields for similarity with adjusted n values in Equation 3, with n = 1(a), 2(b), 3(c), 3.5(d), 4(e), respectively. Colored circles highlight different land types (i.e., blue for type I, red for type II, and purple for type III). The y-axis in each graph is a qualitative evaluation of the proximity (i.e., the greater the value, the lower the proximity). Two-way indicator species analysis was applied to group quadrats by similarity.

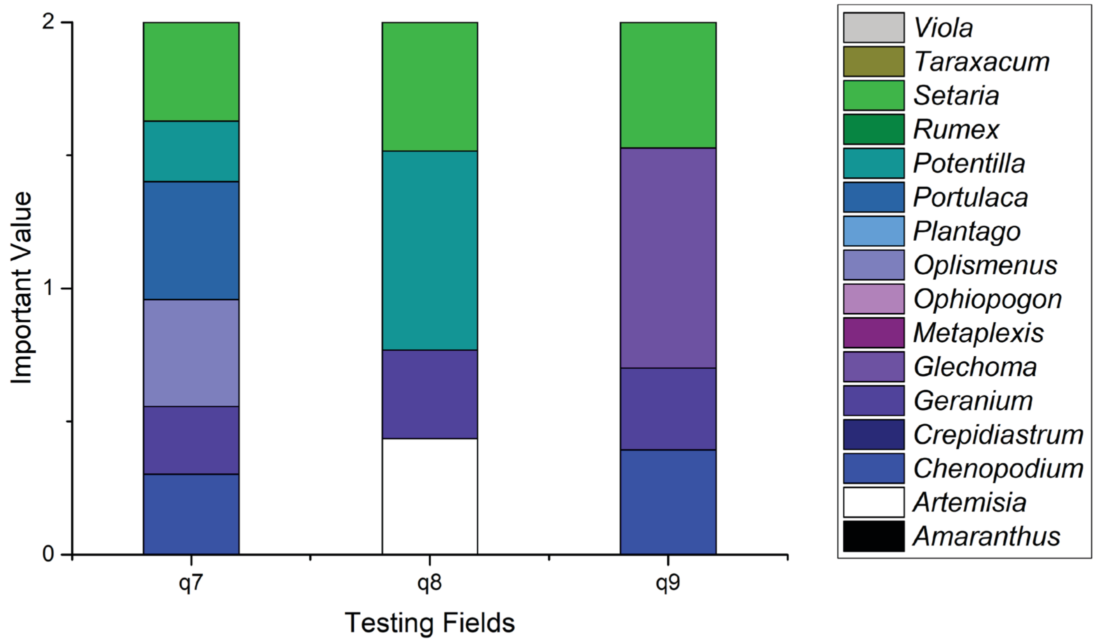

Figure 5: Modeling results using Equation 5, *n* = 3.5. The importance value of each plant species (identified to genera) on three model fields (q7,8,9). Data were drawn from Equation 5, *n* = 3.5.

describes how well this community can endure disturbances (4). Therefore, plant fitness is much more important than biodiversity in terms of establishing a community.

It is worth noting that, the TWINSPAN result did not show an absolute matching between the mathematical modeling result and the natural plant compositions. Although quadrats in Yuan Ming Yuan have been neglected by greenery management for years, most of them were cultivated by humans in earlier times. It is possible that managers in the past had planted popular plants such as *Ophiopogon* and *Setaria*, whose descendants still flourish in the field, occupying these testing fields such as q2, q3, and q4. Therefore, the testing fields were not completely mimicking wild circumstances. Another possible reason for the incomplete matching is that there are numerous possibilities for a plant community to develop in a given field. The modeling result of Equation 5 only shows one possibility that the plant community could grow in the field by considering plant fitness and biodiversity, but it is not exclusive. Quadrats in Yuan Ming Yuan that did not match the modeling prediction show other feasible patterns for plant community development.

What plant species should be included in the plant database to train this model is also worth discussing. We trained this model with a plant database (**Table 2**) that only includes all the existing plant species in the testing fields. Therefore, both the database and the modeling result may not be applicable to other fields in Beijing. To apply this model to other urban green spaces in Beijing, a database with a complete set of plant species in Beijing should be established. Furthermore, to fit for the various purposes of urban green spaces, the city's plant database should include more specific categorizations in addition to the existing parameters. For example, plant species should be specified as local or exotic if local preservation is required. It is also noteworthy that, although arbors did not appear in the testing fields in this simplified model of Equation 5, it is an essential plant type in urban greenery (19). However, the plant trait data for arbors are rarely and ambiguously described in research studies and the official profession guidance handbook (11). Therefore, the plant light value (*LI*) and the plant moisture value (*WA*) for arbors should be collected for further optimization of the modeling. This could be achieved through either ecological study or transforming the literal descriptions for arbor preferences in botanic handbooks (17, 20).

In addition to optimization of the plant database, the model equation itself can be further optimized. Since the weight of index *ct* has not been tested in the current model, modifications on *ct* should also be conducted in further research. Due to the limits of time and resources, we were unable to conduct a long-term field study to test the accuracy of the modeling result. Future research could plant several testing fields based on the modeling result followed by long-term monitoring of the plant growth and the total cost to judge the credibility of this model. Last but not least, in order to achieve high eco-sustainability, our model focuses the attention on the community biodiversity, plant fitness, and managing costs, and neglects social value of green spots. But social factors such as ornamental value, cultural value, human affinity should also be digitized and incorporated into the model in case a balance between sustainability and residential pleasure is aimed for.

Overall, our study established a mathematical model that sets a working framework for intellect urban green management. Further optimization and iteration of the model, such as to parameterize more factors, to optimize the parameterization of plant and land traits and the calculation of fitness index, to test other types of mathematical models, and to assess the credibility of the model using field studies, can be performed on a needed basis. All efforts would lead to improved urban ecosystems, and therefore, an improved environment.

MATERIALS AND METHODS

Plant Traits Parameterization

Plant light value (*LI*) and plant moisture value (*WA*) were obtained using the Ellenberg index and subsequently minimized by 10 (20-22).

Biodiversity Index (BI)

The importance values (IV) of plants were calculated based on the method described in the literature (13). For arbor species, the importance values were calculated by:

$$IV = \frac{RC + RA}{200} \quad \text{[Eqn 7]}$$

in which the relative coverage (*RC*) describes the percentage of surface covered by one species and the relative abundance (*RA*) describes the rarity of one species compared to others.

For herbs and shrubs, the *IV* was calculated by (13):

$$IV = \frac{RC + RH}{200} \quad \text{[Eqn 8]}$$

in which the relative height (*RH*) is calculated based on the literature (13).

The Simpson index (*D*) is calculated by (13):

$$D = \sum_{i=1}^{n} \left(\frac{x_i}{N}\right)^2 \quad \text{[Eqn 9]}$$

in which x_i is the importance value of species *i*, *n* is the number of plant species, and *N* is the sum of all *IV*. When used in the model, the returning value of the biodiversity index is:

$$BI = 1 - D \quad \text{[Eqn 10]}$$

In the model, *BI* is calculated using the vegan (version 2.6-2) function in the R package (24).

Land Traits Parameterization

In developing the Light land value (LV) calculation, we made the following assumptions: 1) the land shape is assumed to be a perfect rectangle; 2) the lands are barren (no blockings within the land); 3) two parallel sides of the land are paralleled to a meridian line; 4) buildings beside a piece of land are perfect cuboids; 5) buildings beside a piece of land are right next to the sides of the land; 6) clouds are negligible; 7) the light intensity level of each day is proportional to that of the summer solstice; 8) the time interval between sunrise and noon is equal to that between noon and sun-set; 9) the sun is moving at a constant angular speed; 10) the building's height is shorter than the land's length and width; 11) the land is on the northern hemisphere. Additionally, the terminology we defined during the calculation is presented in **Table 3**.

The light intensity level is defined as multiplying the area that receives sunlight by the length of time the sunshine reaches it. The area that is covered by the shadows can be calculated using:

$$A = \frac{1}{2} * (2L - x_s) * y_s + \frac{1}{2} * (2S - y_e) * x_e \quad \text{[Eqn 11]}$$

where x_e is the height of the east wall's shade (EWS), y_e is the distance between the EWS and the south wall, y_s is the height of the south wall's shade (SWS), x_s is the height of the SWS, *L* is the measurement of the east-west length of two sides of the land, and *S* is the measurement of the north-south length of two sides of the land.

The two bold lines represent a vertical rod and its shadow at a certain time. The two thin, black dash lines are the length of the shadow's north-south (*y*) and east-west (*x*) components, respectively. The coordinate of the shadow's terminal point is (*x*,*y*). θ_0 is the solar altitude at this moment, while θ_1 and θ_2 are the altitude angle of *y* and *x* respectively, where θ_1 is latitude of the target land, θ_2 is the equivalent of ωt, and *x* and *y* can be calculated using the equations:

$$x = \frac{h}{tan(\theta_2)} \quad \text{[Eqn 12]}$$

$$y = \frac{h}{tan(\theta_1)} \quad \text{[Eqn 13]}$$

Two conditions in calculating the light intensity level of a land with one building blocking the sun are shown below:

$$L_{i(South)} = \frac{\int_0^r A_{(South-p)} \cdot dt + \int_r^p A_{(South-a)} \cdot dt}{S \cdot L \cdot p} \quad \text{[Eqn 14]}$$

$$L_{i(W-E)} = \frac{\int_0^r A_{(W-E-p)} \cdot dt + \int_r^p A_{(W-E-a)} \cdot dt}{2 \cdot S \cdot L \cdot p} + 0.5 \quad \text{[Eqn 15]}$$

Amongst the two equations,

$$r = \frac{arctan\,(\frac{h}{L})}{\omega} \quad \text{[Eqn 16]}$$

$$A_{(South-p)} = L * S - \frac{1}{2} * L^2 * \frac{\tan(\omega t)}{\tan(\alpha)} \quad \text{[Eqn 17]}$$

$$A_{(South-a)} = L * S - \left(2L - \frac{h}{tan(\omega t)}\right) * \frac{h}{\tan(\alpha)} * \frac{1}{2} \quad \text{[Eqn 18]}$$

$$A_{(W-E-p)} = \frac{1}{2} * L^2 * \frac{\tan(\omega t)}{\tan(\alpha)} \quad \text{[Eqn 19]}$$

$$A_{(W-E-a)} = L * S - \left(2S - \frac{h}{\tan(\alpha)}\right) * \frac{h}{\tan(\omega t)} * \frac{1}{2} \quad \text{[Eqn 20]}$$

Three conditions in calculating the light intensity level of a land with more than one building blocking the sun are shown below:

$$L_{i(corner)} = \frac{\int_r^p A_{(corner)} \cdot dt}{2 \cdot S \cdot L \cdot p} + \frac{\int_r^p A_{(South-a)} \cdot dt}{2 \cdot S \cdot L \cdot p} \quad \text{[Eqn 21]}$$

$$L_{i(parallel)} = \frac{\int_0^r A_{(E-p)} \cdot dt + \int_r^p A_{(E-a)} \cdot dt}{2 \cdot S \cdot L \cdot p} + \frac{\int_0^r A_{(W-p)} \cdot dt + \int_r^p A_{(W-a)} \cdot dt}{2 \cdot S \cdot L \cdot p} \quad \text{[Eqn 22]}$$

$$L_{i(Three)} = \frac{\int_r^p A_{(E-S)} \cdot dt}{2 \cdot S \cdot L \cdot p} + \frac{\int_{r_w}^p A_{(W-S)} \cdot dt}{2 \cdot S \cdot L \cdot p} \quad \text{[Eqn 23]}$$

Amongst the three equations,

$$r_w = \frac{arctan\,(\frac{h}{L})}{\omega} \quad \text{[Eqn 24]}$$

$$A_{(corner)} = A_{(South-a)} + A_{(W-E)} - L * S \quad \text{[Eqn 25]}$$

Term	Meaning
t	Time (h)
S	The north-south side length of the land (dm)
L	The east-west side length of the land (dm)
h	A building's height (dm)
$L_{i(South)}$	The light intensity level for lands with a south building
$L_{i(W-E)}$	The light intensity level for lands with an east (or west) building
$L_{i(corner)}$	The light intensity level for lands with two attached, perpendicular buildings
$L_{i(parallel)}$	The light intensity level for lands with two parallel buildings
$L_{i(Three)}$	The light intensity level for lands with three buildings
r	The moment that light first reaches the west side of the land
r_w	The moment that the last light reaches the east side of the land
$A_{(South-p)}$	The area function for buildings on the south before moment r (dm^2)
$A_{(South-a)}$	The area function for buildings on the south after moment r (dm^2)
$A_{(W-E-p)}$	The area function for buildings either on the west or east before moment r (dm^2)
$A_{(W-E-a)}$	The area function for buildings either on the west or east after moment r (dm^2)
$A_{(corner)}$	The area function for two attached, perpendicular buildings– one north-south oriented, the other east-west oriented (dm^2)
$A_{(E-p)}$	The area function for the east building of two parallel buildings at west and east edges of the land before moment r (dm^2)
$A_{(E-a)}$	The area function for the east building of two parallel buildings at west and east edges of the land after moment r (dm^2)
$A_{(W-p)}$	The area function for the west building of two parallel buildings at west and east edges of the land before moment r (dm^2)
$A_{(W-a)}$	The area function for the west building of two parallel buildings at west and east edges of the land after moment r (dm^2)
$A_{(E-S)}$	The area function for east and south building (in the morning) (dm^2)
$A_{(W-S)}$	The area function for west and south wall (in the afternoon) (dm^2)
p	The time interval between sunrise and noon (12:00) (h)
α	The altitude of the sun at the noon of the summer solstice (rad) $$\alpha = \frac{\pi}{2} - (la - \frac{47\pi}{360})$$
la	The latitude of that area (°)
ω	The angular speed of the sun (rad) $$\omega = \frac{\pi}{2p}$$

Table 3: Table 3. Terminology for *LV* calculation. Calculation methods for some terms are also presented.

$$A_{(E-p)} = A_{(W-E-p)};\ A_{(E-a)} = A_{(W-E-a)} \quad \text{[Eqn 26]}$$

$$A_{(W-p)} = A_{(W-E-p)};\ A_{(W-a)} = A_{(W-E-a)} \quad \text{[Eqn 27]}$$

$$A_{(E-S)} = A_{(South-a)} + A_{(W-E-a)} - L * S \quad \text{[Eqn 28]}$$

$$A_{(W-S)} = A_{(South-a)} + A_{(W-E-a)} - L * S \quad \text{[Eqn 29]}$$

Land Moisture Value (MV)

We set the climate parameter (*CP*) as 0.2 for regions with annual precipitation of more than 800 mm, 0 for regions with annual precipitation within a range of 400 to 800 mm, and -0.2 for regions with annual precipitation below 400 mm. We set the irrigation equipment parameter (*IEP*) as 0.9 (for lands with automatic irrigation equipment), 0.5 (for lands with irrigation equipment that need manual work), and 0.2 (for lands with no irrigation equipment). We set the natural water sources parameter (*NWP*) as 0.3 for wetlands and marshes, 0.2 for lands with a water source (i.e., lakes, ponds, rivers, etc.) within 10 meters reach, 0.1 for lands with a water source within 100 meters reach, and 0 for lands with no adjacent water sources.

The *MV* of a piece of land was calculated as:

$$MV = CP + IEP + NWP \quad \text{[Eqn 30]}$$

R Language Programming

We established the mathematical model based on the R language (see **Appendix 1** for the codes). The two packages that were used in this model are *vegan* (version 2.6-2) for biodiversity index calculation and *Rdonlp2* (version 3042.11/r6080) for nonlinear programming (18,24).

Plant Database

All the plant traits data for the 16 plant species used in this study were acquired from a botanic website (www.iplant.cn/) and two handbooks except the unit coverages (17, 20, 24). The unit cost of seeds/seedlings was acquired from TaoBao **(Table 2)**.

Quadrat Survey

Six quadrats, each with an area roughly about 750 dm^2, were selected in Yuan Ming Yuan. These quadrats were enclosed by a thin rope during the survey to discriminate the boundary (**Figure 3a-c**). Plant coverage data were drawn from this process. After identifying and recording each plant species (genera), the coverage of plant species that grow in colonies was calculated by measuring the length and width of the area each species covered followed by the calculation of the covered area. Individual plants were counted by measuring the diameter of their shadow area, then calculating each plant's unit coverage.

Two-Way Indicator Species Analysis

Two-way indicator species analysis (TWINSPAN) was applied to evaluate the similarities between the testing fields and their corresponding mathematical modeling results. The R code for TWINSPAN programming was acquired from GitHub (26).

ACKNOWLEDGMENTS

The study was supported by Beijing Education Science 13th Five-Year Plan 2020 General Project CDDB2020209 entitled with " The Study of Implementation Strategies and Effects of In-Life-Problem-Based High School Biology Inquiry Experiments". We would like to thank Mr. Fang Wei for helping coordinate the R program codes and Mrs. Ling Shuyun for accompanying us through the ecological survey. We would also like to thank Mr. Lu Yizhou for supporting the mathematical modeling process, Ms. Fan Jiatong for providing us with guidance on writing our paper, Dr. Zeng Yao for inspiring the topic selection, Dr. Li Geng for his instruction in plotting our data, and to Mr. Alex Sicurella for his advice in revising our paper.

Received: August 18, 2022
Accepted: November 17, 2022
Published: July 5, 2023

REFERENCES

1. Wang, Guohong. “Biodiversity and Ecosystem Stability Review.” *Biodiversity Science,* vol. 10, no. 1, 2002, pp. 126–34.
2. Yu, Dandan, *et al.* “Indicator systems and methods for evaluating biodiversity and ecosystem services.” *Acta Ecologica Sinica,* vol. 37, no. 2, 2017, pp. 349–57. doi.org/10.5846/stxb201611092272.
3. Pennekamp, Frank, *et al.* “Biodiversity Increases and Decreases Ecosystem Stability.” *Nature,* vol. 563, no. 7729, Nov. 2018, pp. 109–12. doi.org/10.1038/s41586-018-0627-8.
4. Fu, Jinggen. “Biodiversity and Environment in Cities—A Case Study of Birds in Urban and Suburban Areas of Japan.” *Journal of Human Settlements in West China,* vol. 34, no. 3, 2019, pp. 8–18, doi.org/10.13791/j.cnki.hsfwest.20190302.
5. Sayer, Emma J., and Ralf Schäfer. “The Diversity of Decay.” *ELife,* vol. 9, Aug. 2020, p. e60375. doi.org/10.7554/eLife.60375.
6. Zhao, Yezhou, *et al.* “The Role of Leguminous Plants and Rhizobium in Ecological Environment.” A*gro-Environment & Development,* vol. 2013, no. 4, pp. 7–12, doi.org/10.3969/j.issn.1005-4944.2013.04.002.
7. Qian, Hong. “Relationships between Plant and Animal Species Richness at a Regional Scale in China.” *Conservation Biology,* vol. 21, no. 4, Aug. 2007, pp. 937–44. doi.org/10.1111/j.1523-1739.2007.00692.x.
8. Wu, Panlong, *et al.* “Effects of Plant Diversity, Vegetation Composition, and Habitat Type on Different Functional Trait Groups of Wild Bees in Rural Beijing.” *Journal of Insect Science,* vol. 18, no. 4, July 2018. doi.org/10.1093/jisesa/iey065.
9. Hou, Feibing, *et al.* “Composition and Distribution of Green Spaces and Plant Species in Urban-Rural Mosaic Areas of Beijing.” *Acta Ecologica Sinica,* vol. 36, no. 19, 2016, pp. 6256–65.
10. Li Xiaohong, *et al.* “Temporal and spatial characteristics of spontaneous plant composition and diversity in a Beijing urban park.” *Acta Ecologica Sinica,* vol. 38, no. 2, 2018. doi.org/10.5846/stxb201610272193.
11. DB11. *Technical Regulations of Establishing and Maintaining on Flower Meadow Landscape.* Beijing

Market Supervision Administration, 1 Jan. 2021.
12. Capitán, José A., et al. “Competitive Dominance in Plant Communities: Modeling Approaches and Theoretical Predictions.” *Journal of Theoretical Biology,* vol. 502, Oct. 2020, p. 110349. doi.org/10.1016/j.jtbi.2020.110349.
13. Zhang, Jintun. *Shu Liang* Sheng Tai Xue. 3rd, Sciencep, 2018.
14. Zhao, Meifang, *et al.* “Plant Phenological Modeling and Its Application in Global Climate Change Research: Overview and Future Challenges.” *Environmental Reviews,* vol. 21, no. 1, Mar. 2013, pp. 1–14. doi.org/10.1139/er-2012-0036.
15. Vanlisuta, C., and S. Prombanpong. “A Mathematical Approach for Afforestation to Mitigate Climate Change.” *International Journal of Innovation,* vol. 4, no. 1, 2013, p. 6.
16. Wang, Boya, *et al.* “Assessment of Ecosystem Service Quality and Its Correlation with Landscape Patterns in Haidian District, Beijing.” *International Journal of Environmental Research and Public Health,* vol. 16, no. 7, Apr. 2019, p. 1248. doi.org/10.3390/ijerph16071248.
17. Xiong, Youqing, and Chunling Li, editors. *Beijing Native Plant.* China Forestry Publishing House, 2015.
18. code PShwtos, version SShtdffttAC, allocation KSChadm, objecive CBhap, pointer cae, interface RThwtoR, interface. DWhwtcR (2017). _Rdonlp2: Rmetrics - An R extension library to use Peter Spelluci's DONLP2 from R_. R package version 3042.11/r6080, R-Forge.R-project.org/projects/rmetrics/.
19. DB11. *Technical Guideline for Urban Forest Construction.* 2019.
20. Xu, Huichun, editor. *Shumu Jingdian Tujian.* Jilin Science and Technology Press, 2013, www.jlstp.cn/.
21. Hill, M. O., editor. *Ellenberg’s Indicator Values for British Plants.* Center for Ecology and Hydrology, Natural Environment research Council, 1999.
22. Metcalfe, Helen, *et al.* “Using Functional Traits to Model Annual Plant Community Dynamics.” *Ecology,* vol. 101, no. 11, Nov. 2020. doi.org/10.1002/ecy.3167.
23. *Přírodovědecká Fakulta Muni,* www.sci.muni.cz/botany/juice/ELLENB.TXT.
24. “Iplant.” *Iplant,* Institute of Botany, the Chinese Academy of Sciences, www.iplant.cn/.
25. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2022). _vegan: Community Ecology Package_. R package version 2.6-2, CRAN.R-project.org/package=vegan.
26. Oksanen, Jari. (2021). *twinspan.* Retrieved from github.com/jarioksa/twinspan/blob/master/NAMESPACE

Copyright: © 2023 Fang and Zhang. All JEI articles are distributed under the attribution non-commercial, no derivative license (http://creativecommons.org/licenses/by-nc-nd/3.0/). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.

Appendix 1. R language programming codes

```
library(vegan)
library(Rdonlp2)
d.species <- read.csv("species listy.csv",header=TRUE,stringsAsFactors=FALSE)
#"species listy.csv" is the document of plant database shown in Table 2
Scientific<-d.species[,"Scientific"]
Type<-d.species[,"Type"]
Height <- d.species[,"Height"]
Coverage<-d.species[,"Coverage"]
Light_v<-d.species[,"Light_v"]
Moisture_v<-d.species[,"Moisture_v"]
Cost<-d.species[,"Cost"]

#Preparation functions
fc <-function(s,ns){ #turn original data into vectors
  n<-0
  for (i in s){
    n<-n+1
    a<-strsplit(i,split=",")
    a<-as.vector(a)
    ns[n]<-a
  }
  return(ns)
}
fn <-function(s,element){ #find the place of an element in a list/vector
  n<-0
  for (i in s){
    n<-n+1
```

```
    if (i==element)
      return (n)
  }
}

#Categorize plant species into arbor, shrub, and herb
Arbor_s<-c();Shrub_s<-c();Herb_s<-c();
k=0
for (i in Type){
  k=k+1
  q=Scientific[[k]]
  if (i=="arbor"){
    Arbor_s <-c(Arbor_s,q)
  }
  else if (i=="shrub"){
    Shrub_s <-c(Shrub_s,q)
  }
  else if (i=="herb"){
    Herb_s <-c(Herb_s,q)
  }
}

#Acquire the quantity of each type
plant<- c(Scientific)
total<-length(plant)
arbor<-length(Arbor_s)
shrub<-length(Shrub_s)
herb<-length(Herb_s)
```

```
#Land parameterization
dm<-c(750);d_sh<-dm #Land area (in dm^2)
dd1=4.1;dd2=1.9;dd3=3 #The aridity of each field type

#Land information: light value
h=50 #Height of the adjacent building (in dm)
S=30 #The east-west side length of the adjacent building (in dm)
L=25 #The north-south side length of the adjacent building (in dm)
al=2/9 #The latitude of Beijing
alpha=pi/2-(al-pi*47/360) #The altitude of the sun at the noon of the summer solstice
w=pi/12 #The angular speed of the sun
#The area function for different types of walls (see Materials and Methods)
Asouthp <-function(x){
  L*S-0.5*(L^2)*tan(w*x)/tan(alpha)
}
Asoutha <-function(x){
  L*S-0.5*(2*L-h/tan(w*x))*h/tan(alpha)
}
Acorner <-function(x){
  L*S-0.5*((2*L-h/(tan(w*x)))*h/tan(alpha)+(2*S-h/tan(alpha))*h/tan(w*x))
}
c=atan(h/L)/w
p=6-10^(-12)
q=10^(-12)
Ec=(integrate(Acorner,c,p))
Es=(integrate(Asoutha,c,p))
Esp=(integrate(Asouthp,q,c))
```

```
LV1=(Esp[[1]])/(S*L*6) #land light value for field type I

LV2=1 #land light value for field type II

LV3=(Ec[[1]]+Es[[1]])/(2*S*L*6) #land light value for field type III



#Nonlinear programming

HS_s<-c(Herb_s,Shrub_s); #combining herb and shrub species

cod_hs<-c() #Code for herb/shrub species

for (i in Herb_s){

  t<-fn(Scientific,i)

  cod_hs<-c(cod_hs,t)

}

for (i in Shrub_s){

  t<-fn(Scientific,i)

  cod_hs<-c(cod_hs,t)

}

cvr_hs<-c() #Coverage for each herb/shrub species

ht_hs<-c() #Average height for each herb/shrub species

for (i in cod_hs){

  t=Coverage[i]

  p=as.numeric(Height[i])

  cvr_hs<-c(cvr_hs,t)

  ht_hs<-c(ht_hs,p)

}


#Modeling for the field type I

Test1 = function(x){ #Objective function

  x<-abs(x)
```

```
  rh<-ht_hs*x/sum(ht_hs)*100
  rd<-cvr_hs*x/sum(cvr_hs*x)*100
  p<-(rh+rd)/200 #calculate the importance value for each species
  sip=diversity(p,"simpson") #calculate the Simpson biodiversity index of the plant community
  b<-0
  ct<-0
  ft<-0
  for (i in cod_hs){
    b=b+1
    li=as.numeric(Light_v[[i]])/10
    wa=as.numeric(Moisture_v[[i]])/10
    ft=ft+x[b]*cvr_hs[b]/d_sh*(sqrt(2)-sqrt((li-LV1)^2+(wa-dd1/10)^2))/sqrt(2)  #calculating  the
fitness index
    ct=ct+x[b]*Cost[[i]]
  }
  ct=1-ct/d_sh/(max(Cost[cod_hs]/Coverage[cod_hs])) #calculating the cost index
  mi=(ft^3.5)*sip^(3.5) #n=3.5
  return(-mi)
}
hs<-herb+shrub
p1 = rep(10,hs) #set initial iteration values
par1.l = rep(0,hs); #set lower limits
par1.u = rep(10000,hs); #set upper limits
A_1 = matrix(cvr_hs,1,byrow=TRUE); #set linear restrictions
lin1.l = c(d_sh); lin1.u = c(d_sh); #set linear limits (both equals to the total area)
#Nonlinear programming
T1 = donlp2(p1,Test1,par.u=par1.u,par.l=par1.l,A_1,lin.l=lin1.l,lin.u=lin1.u)
#testing
```

```
pt1 = function(x){
  x<-abs(x)
  s<-0
  for (i in 1:length(x)){
    if(x[i]>=1){
      s=s+ht_hs[i]
    }
  }
  rh<-c(1:length(x))
  for (i in 1:length(x)){
    if (x[i]>=1){
      rh[i]<-ht_hs[i]/s*100
    }
    else{
      rh[i]<-0
    }
  }
  rd<-cvr_hs*x/sum(cvr_hs*x)*100
  p<-(rh+rd)/200 #calculating the importance value
  return(p)
}

q1<-pt1(T1$par)*2 #output the importance value of each species

#Modeling for the field type II
Test2 = function(x){
  x<-abs(x)
  rh<-ht_hs*x/sum(ht_hs)*100
```

```
  rd<-cvr_hs*x/sum(cvr_hs*x)*100
  p<-(rh+rd)/200
  sip=diversity(p,"simpson")
  b<-0
  ct<-0
  ft<-0
  for (i in cod_hs){
    b=b+1
    li=as.numeric(Light_v[[i]])/10
    wa=as.numeric(Moisture_v[[i]])/10
    ft=ft+x[b]*cvr_hs[b]/d_sh*(sqrt(2)-sqrt((li-LV2)^2+(wa-dd2/10)^2))/sqrt(2)
    ct=ct+x[b]*Cost[[i]]
  }
  ct=1-ct/d_sh/(max(Cost[cod_hs]/Coverage[cod_hs]))
  mi=(ft^3.5)*sip^(3.5)
  return(-mi)
}
hs<-herb+shrub
p2 = rep(10,hs)
par2.l = rep(0,hs);
par2.u = rep(10000,hs)
A_2 = matrix(cvr_hs,1,byrow=TRUE)
lin2.l = c(d_sh); lin2.u = c(d_sh)
#Nonlinear programming
T2 = donlp2(p2,Test2,par.u=par2.u,par.l=par2.l,A_2,lin.l=lin2.l,lin.u=lin2.u)
#testing
pt2 = function(x){
  x<-abs(x)
```

```
  s<-0
  for (i in 1:length(x)){
    if(x[i]>=1){
      s=s+ht_hs[i]
    }
  }
  rh<-c(1:length(x))
  for (i in 1:length(x)){
    if (x[i]>=1){
      rh[i]<-ht_hs[i]/s*100
    }
    else{
      rh[i]<-0
    }
  }
  rd<-cvr_hs*x/sum(cvr_hs*x)*100
  p<-(rh+rd)/200
  return(p)
}
q2<-pt2(T2$par)*2

#Modeling for the field type III
Test3 = function(x){
  x<-abs(x)
  rh<-ht_hs*x/sum(ht_hs)*100
  rd<-cvr_hs*x/sum(cvr_hs*x)*100
  p<-(rh+rd)/200
  sip=diversity(p,"simpson")
```

```
  b<-0
  ct<-0
  ft<-0
  for (i in cod_hs){
    b=b+1
    li=as.numeric(Light_v[[i]])/10
    wa=as.numeric(Moisture_v[[i]])/10
    ft=ft+x[b]*cvr_hs[b]/d_sh*(sqrt(2)-sqrt((li-LV3)^2+(wa-dd3/10)^2))/sqrt(2)
    ct=ct+x[b]*Cost[[i]]
  }
  ct=1-ct/d_sh/(max(Cost[cod_hs]/Coverage[cod_hs]))
  mi=(ft^3.5)*sip^(3.5)
  return(-mi)
}
hs<-herb+shrub
p3 = rep(10,hs)
par3.l = rep(0,hs);
par3.u = rep(10000,hs)
A_3 = matrix(cvr_hs,1,byrow=TRUE)
lin3.l = c(d_sh); lin3.u = c(d_sh)
#Nonlinear programming
T3 = donlp2(p3,Test3,par.u=par3.u,par.l=par3.l,A_3,lin.l=lin3.l,lin.u=lin3.u)
#testing
pt3 = function(x){
  x<-abs(x)
  s<-0
  for (i in 1:length(x)){
    if(x[i]>=1){
```

```
      s=s+ht_hs[i]
    }
  }
  rh<-c(1:length(x))
  for (i in 1:length(x)){
    if (x[i]>=1){
      rh[i]<-ht_hs[i]/s*100
    }
    else{
      rh[i]<-0
    }
  }
  rd<-cvr_hs*x/sum(cvr_hs*x)*100
  p<-(rh+rd)/200
  return(p)
}
q3<-pt3(T3$par)*2
```