

# Space-ecology set covering problem for modeling Daiyun Mountain Reserve, China

Chih-Wei Lin<sup>1,2,3</sup>, Jinfu Liu<sup>2,4,7</sup>, Jiahang Huang<sup>2,4</sup>, Huiguang Zhang<sup>5</sup>, Siren Lan<sup>4</sup>, Wei Hong<sup>4</sup> and Wenzhou Li<sup>6</sup>

<sup>1</sup> Forestry Post-doctoral Station of Fujian Agriculture and Forestry University, Fuzhou, China;

<sup>2</sup> College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou, China;

<sup>3</sup> College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, China;

<sup>4</sup> Strait Nature Reserve Research Center of Fujian Agriculture and Forestry University, Fuzhou, China;

<sup>5</sup> Fujian Forestry Survey and Planning Institute, Fuzhou, China;

<sup>6</sup> Administration of Daiyun Mountain National Nature Reserve of Fujian Province, Quanzhou, China.

<sup>7</sup> fjljf@126.com

**Abstract.** Site selection is an important issue in designing the nature reserve that has been studied over the years. However, a well-balanced relationship between preservation of biodiversity and site selection is still challenging. Unlike the existing methods, we consider three critical components, the spatial continuity, spatial compactness and ecological information to address the problem of designing the reserve. In this paper, we propose a new mathematical model of set covering problem called Space-ecology Set Covering Problem (SeSCP) for designing a reserve network. First, we generate the ecological information by forest resource investigation. Then, we split the landscape into elementary cells and calculate the ecological score of each cell. Next, we associate the ecological information with the spatial properties to select a set of cells to form a nature reserve for improving the ability of protecting the biodiversity. Two spatial constraints, continuity and compactability, are given in SeSCP. The continuity is to ensure that any selected site has to be connected with adjacent sites and the compactability is to minimize the perimeter of the selected sites. In computational experiments, we take Daiyun Mountain as a study area to demonstrate the feasibility and effectiveness of the proposed model.

## 1. Introduction

The biodiversity is a crucial component for maintaining the stability of ecosystems and is the basis of developing the resources. The abundant biodiversity can provide various resources for human such as the food and drugs. However, the destruction of habitat, the over-exploiting of resource, and the environmental pollution make the loss of biodiversity. To maintain the biodiversity, conservation biologists consider that establishing nature reserves is an effective approach [1], where designing nature reserves is the foundation of establishing nature reserves.

The nature reserve is the area with land or water, which is designed for protecting the biodiversity, including ecology, species, communities and other conservation characteristics. Many methods have been proposed for designing nature reserves within the last two decades, these methods can be classified into five forms: (1) numerical scoring, (2) minimum reserve set, (3) integer program, (4)



spatial problem, and (5) heuristic approach. The approaches of the first form are to select sites with the numerical scoring, which considers various criteria with different numerical scores, and the sites with the highest scores are recommended [2]. However, to maximize the numbers of species, the large numbers of sites are selected. The approaches of the second form aim at minimizing the number of selected sites by utilizing a variety of iterative procedures [3]. The numerical score is also considered and adjusted in each iterative step in the approaches of the second form. The problem of minimizing the reserve sites can be identified as a set covering problem (SCP). SCP is a systematic and efficient reserve selection model which can be formulated as the integer program (IP) and its basic objective is using the minimum number of sites to represent each species at least once [4]. However, the long-term persistence of all species cannot be ensured. Therefore, each species considers at least  $n$  times in the SCP [5]. The maximal covering problem (MCP) is another problem of IP, which related to SCP, and aims at maximizing the number of species in the selected reserve sites [6]. Although the constraints of species coverage are considered in various models, such as SCP and MCP, the spatial attributes are loss.

The habit fragmentation is related to the connectivity of the selected sites that is important for preserving the biodiversity in the reserve selection problem. Therefore, the spatial distribution is a key factor of species richness, such as spatial continuity and compactability, in designing nature reserve [7]. Six geometric principles are proposed for designing nature reserves including, reserve size, number of reserves, reserve connectivity, reserve shape and two reserve proximities [8]. These geometric principles consider the characteristics of the spatial contiguity and compactness in designing reserve. A number of researchers focus on the connectivity of selected sites in the field of biology, in which the optimization and the heuristic algorithms are proposed to solve the problem of continuity [9, 10]. The optimization algorithms find the optimal solution with scarce resources. However, it has a high computational cost compared with a heuristic algorithm in practice [11]. Although the computational and spatial costs of the heuristic algorithm are acceptable costs, the optimal solution cannot be guaranteed.

Daiyun Mountain national nature reserve is an important conservation of water resource and a typical forest system in the mountains of southeast coast of China, lacking reasonable layout and effective protection area. The ecological information is an important factor of forest ecosystems, and does not consider in the aforementioned methods. In this study, we propose an algorithm called Space-ecology Set Covering Problem (SeSCP) which associates two spatial properties (connectivity and compactness) and considers the ecological information of forest for designing the reserve.

The rest of this paper is organized as follows. First, we introduce the material and methods including the study area, the ecological information and the proposed model, Space-ecology Set Covering Problem (SeSCP). Next, we demonstrate the experimental results and discussion. Finally, we give the conclusions.

## 2. Material and methods

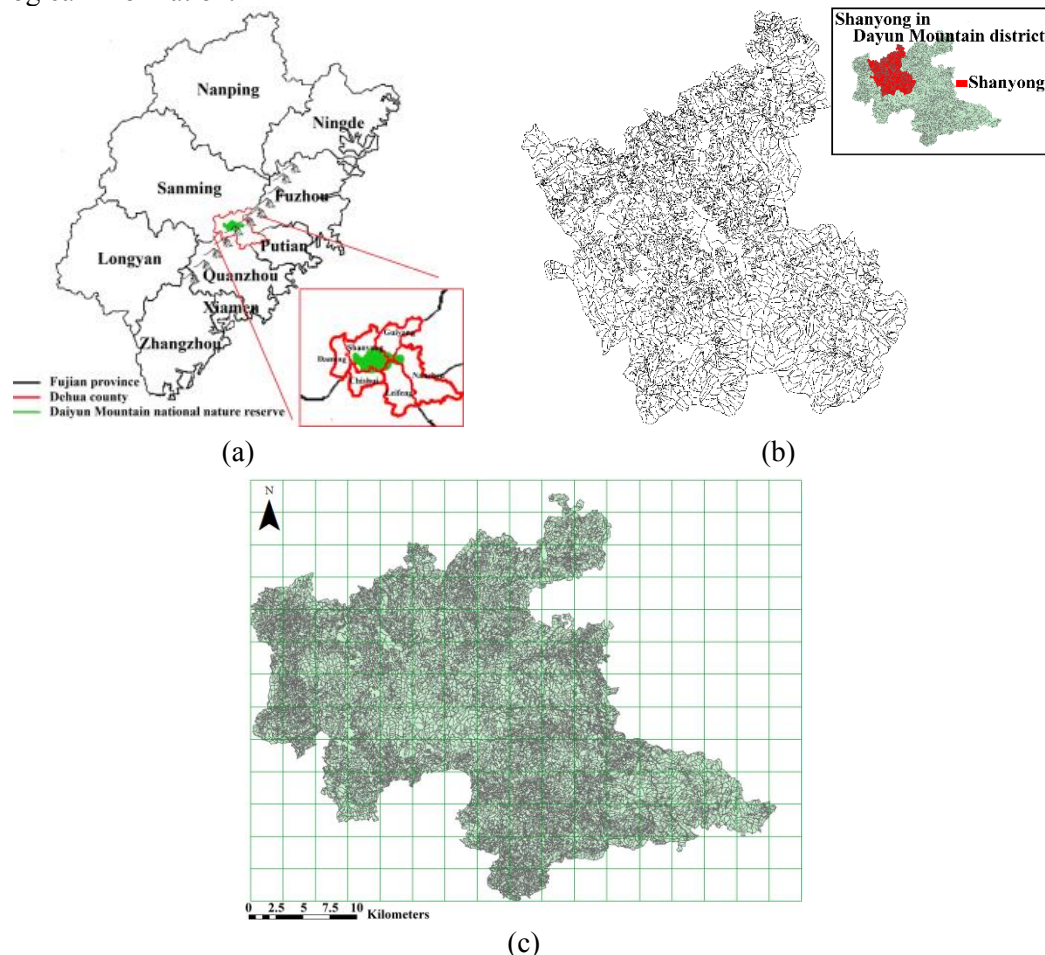
In this section, we first present the study area, Daiyun Mountain, China. Then, we represent the evaluation of the ecology of the study area. Next, we introduce the proposed set covering problem, Space-ecology Set Covering Problem (SeSCP), by considering the properties of ecology and space information.

### 2.1. Study area: Daiyun Mountain, China

The Daiyun Mountain national nature reserve is located in Dehua County that is an important conservation and cradle of water resources in central Fujian (figure 1a). This nature reserve is covered by a semi-natural forest, which has a complete forest ecosystem in the mountains, abundant species of plants and animals, and is the central of biodiversity in Fujian. This area is approximate to 93.4% and that covers 6 villages and towns with 18,762 sublots and has 13472.4  $hm^2$  of the area (figure 1a).

The integrated forest ecosystem includes animals, plants, microorganisms and abiotic environmental factors. However, it is difficult to collect the real data by forest resource investigation

for various species. Therefore, we only consider the information about plants with zones, province and nationally protected species, which are the second category investigation of forest resource to generate the ecological information.

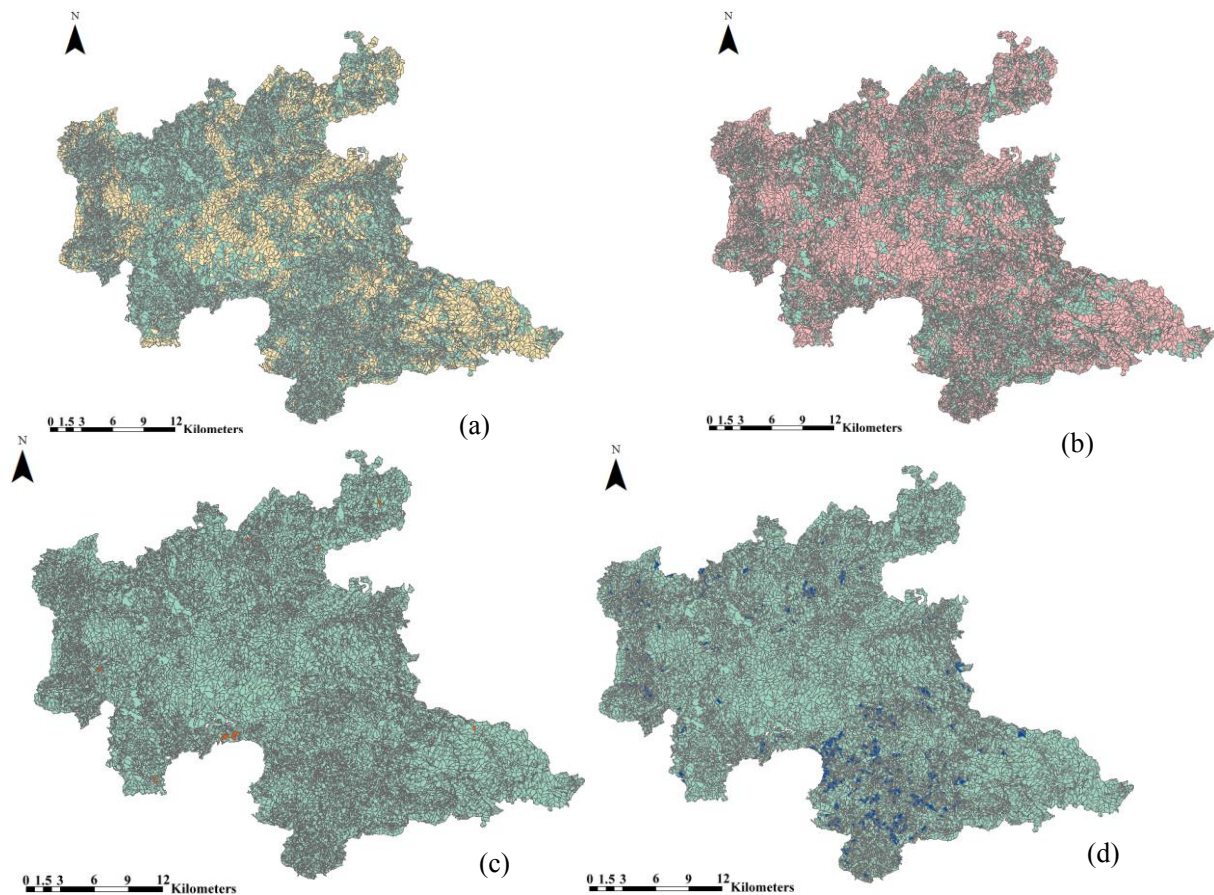


**Figure 1.** Study area - Daiyun Mountain district. (a) Daiyun Mountain nature reserve covers 6 villages in Fujian, (b) sublots of Daiyun Mountain district, (c) grid of Daiyun Mountain district. In figure 1a, the area of Daiyun Mountain is indicated as the red line and the Daiyun Mountain national nature reserve is indicated as the green area.

In this study, we select species according to the plants of zonality and protection of Daiyun Mountain for designing reserve. Its zonal plants are the montane broadleaved and coniferous forests (figures 2a, 2b). We select the *Pinus taiwanensis* Hayata, *Liquidambar formosana*, *Pinus massoniana*, *Cunninghamia lanceolata*, *Schima superba*, and other hardwood species to represent the zonal plants of Daiyun Mountain. We also consider the 1<sup>st</sup> class national protected species, *Fokienia hodginsii*, and the Fujian province protected species, *Cryptomeria japonica* var. *sinensis* in our study (figure 2c). To maintain the diversity, we further consider the alien species, *Eucalyptus robusta* and *Pinus taeda* (figure 2d).

## 2.2. Ecological Information

To protect the forest ecosystem, the ecological score  $e$  is considered. Due to the difficulties in field investigation, we use the second category investigation of forest resource to calculate the ecological score in this study. The subplot is the basic unit for evaluating the ecological score, which has an irregular shape. Figure 1b is an example of subplot of Shanyong in Daiyun Mountain district, and it has a numerous of sublots. To decrease the computation cost and to improve the feasibility, we separate the study area into 215 grids with  $3 \times 3 \text{ km}^2$  for each grid (figure 1c).



**Figure 2.** The distribution of species. (a) zonal species - montane broadleaved forests, (b) zonal species - coniferous forests, (c) The protected species, (d) alien species. The zonal species of montane broadleaved forests are indicated as yellow color and the zonal species of coniferous forests are indicated as red color. The protected species are indicated as orange color. The alien species are indicated as blue color.

Various properties are considered to evaluate the ecological scores for each subplot, includes the varieties of trees ( $VoT$ ), the stand closure ( $S_C$ ), the average height of trees ( $H_{avg}$ ), the mean diameter of trees ( $D_m$ ), the age composition of trees ( $C_{age}$ ), and the site quality ( $S_Q$ ). The ecological score of each subplot is evaluated according to the grading criteria as shown in table 1.

To enhance the readability, the items of land are abbreviated as:  $c$  &  $b$  is commercial and bamboo forests,  $o$  &  $s$  is open woodland and shrub land,  $s$  &  $u$  is the suitable and unplanned lands,  $u$  &  $n$  is the unutilized wasteland and non-forest land. The items of  $VoT$  are abbreviated as: *obvious dominating* is the obvious dominating species with multiple accompanying species, *dominating* is the dominating species with multiple accompanying species, *single dominating* is the single dominating species with single accompanying species, *no dominating* is no dominating species, *single* is the single species. The items of  $C_{age}$  are abbreviated as:  $C_{Mid\_age}$  is middle-aged stand,  $C_{Near\_mat}$  is near-mature forest,  $C_{Mat}$  is mature forest,  $C_{over\_mat}$  is over-mature forest, and  $C_{Young\_age}$  is young-aged stand.

The ecological score of each subplot is calculated by accumulating the score of each item, as shown in table 1. For example, one of the sublots  $i$  has the properties includes stand, dominating,  $S_C$  is 0.7~0.8,  $H_{avg}$  is 7~10,  $D_m$  is 8~10,  $C_{age}$  is  $C_{Mat}$ , and  $S_Q$  is fertile. The scores of each item are 10, 8, 8, 6, 4, 6, 6 and the total score of ecology of that subplot  $i$  is  $e_i = 10 + 8 + 8 + 6 + 4 + 6 + 6 = 48$ .

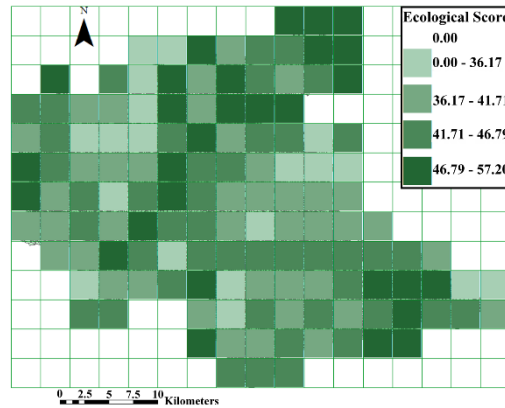
**Table 1.** Grading standard of ecological score.

Score	Land	VoT	$S_C$	$H_{avg/m}$	$D_{m/cm}$	$C_{age}$	$S_O$
10	stand	obvious dominating	0.8 over	16 over	20 over	$C_{Mid\_age}$	Most fertile
8	$c \& b$	dominating	0.7 ~ 0.8	10 ~ 16	14 ~ 20	$C_{Near\_mat}$	More fertile
6	$o \& s$	single dominating	0.6 ~ 0.7	7 ~ 10	10 ~ 14	$C_{Mat}$	fertile
4	$s \& u$	no dominating	0.4 ~ 0.6	4 ~ 7	8 ~ 10	$C_{over\_mat}$	Infertile
2	$u \& n$	single	0 ~ 0.4	0 ~ 4	0 ~ 8	$C_{Young\_age}$	-
0	non-forest	no species	0	0	0	-	-

However, numerous of sublots make the high computational cost in selecting sites. Therefore, we divide the study area, Daiyun mountain district, into  $n \times n$  grids to reduce the computational complexity (figure 1c). After separating, we re-calculate the ecological score for each grid based on the ecological score of each subplot. We utilized the ecological score and the area of each subplot to evaluate the ecological score of each grid, the formula can be formally defined as follow,

$$\varepsilon_i = \frac{e_1 s_1 + e_2 s_2 + \dots + e_n s_n}{s_1 + s_2 + \dots + s_n}, \quad (1)$$

where  $\varepsilon_i$  is the ecological score of each grid,  $e_n$  is the ecological score of each subplot, and  $s_n$  is the area of each subplot. Figure 3 shows the evaluated results of the ecological score of each grid in Daiyun Mountain district.



**Figure 3.** The ecological score of each grid with in Daiyun mountain district. The darker the color is, the more ecological score is.

### 2.3. Mathematical models

Designing nature reserve is considered as the problem of site selection. Two issues have to be addressed to ensure the biodiversity, continuity and compactability. In this study, we based on the network flow problem [12] and proposed Space-ecology Set Covering Problem (SeSCP), which associated the ecological information with the spatial properties to solve the problem of selecting sites.

SeSCP is designed as a bio-objective function as shown in Eq. 2 and the detail of the proposed model, SeSCP, is written as following,

$$\min c = \omega \cdot \sum_i U_i (1 - \varepsilon_i) + (1 - \omega) \cdot \left( 4 \sum_i U_i - \sum_{i,j \in N_i} Z_{i,j} \right), \quad (2)$$

$$s.t. \sum_i \delta_{s_i} U_i \geq k_s, \quad \forall s \in S, \quad (3)$$

$$\sum_{i=1}^n T_i = 1, \quad (4)$$

$$\sum_{j \in N_i} X_{j,i} \leq (M - 1) \cdot U_i \quad \forall i = 1, 2, \dots, n \quad (5)$$

$$\sum_{j \in N_i} X_{i,j} - \sum_{j \in N_i} X_{j,i} \geq U_i - M \cdot T_i \quad \text{for all } i = 1, 2, \dots, n, \quad (6)$$

$$Z_{i,j} \leq U_i \quad \forall i = 1, 2, \dots, n \quad j \in N_i, \quad (7)$$

$$Z_{i,j} \leq U_j \quad \forall i = 1, 2, \dots, n \quad j \in N_i, \quad (8)$$

$$Z_{i,j} \geq U_i + U_j - 1 \quad \forall i = 1, 2, \dots, n \quad j \in N_i, \quad (9)$$

$$T_i \leq U_i \quad \forall i = 1, 2, \dots, n, \quad (10)$$

$$U_i = \begin{cases} 1, & \text{if site } i \text{ is selected} \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$T_i = \begin{cases} 1, & \text{site } i \text{ is the sink point} \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

$$Z_{i,j} = \begin{cases} 1, & \text{site } i \text{ and } j \text{ are the sink point} \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

In this model, Eq. 2 is the bio-objective function, which minimizes the cost  $c$  by minimizing the selected site number with characteristics of the spatial continuity and compactness and maximizing the ecological score. In Eq. 2,  $\omega$  is the weight parameter for tri-objective,  $Z_{i,j}$  is a binary variable for defining the shared boundary between two sites,  $\sum_i U_i(1 - \varepsilon_i)$  is a objective, including maximizing

the ecological score and minimizing the selected site number and  $\left(4 \sum_i U_i - \sum_{i,j \in N_i} Z_{i,j}\right)$  is the objective

of the space constraint. Eq. 3 is a constraint, which imposes a restriction on the percentage of the species; for each species  $s$ , there is at least percentage,  $k_s$ , in the selected sites. The objective function Eq. 2 and constraint Eq. 3 built the proposed model, Space-ecology Set Covering Problem (SeSCP). Eq. 4 defines the sink point which is the only one in the designing network. Eq. 5 states the inflow specifications. Eq. 6 describes the relationship between outflow and inflow of a site, in which  $\sum_{j \in N_i} X_{i,j}$

presents the inflow and  $\sum_{j \in N_i} X_{j,i}$  represents the outflow; the difference between inflow and outflow

should be greater than or equal to  $U_i - M \cdot T_i$ . Eqs. 7, 8 and 9 are used to express the logic relationship between  $Z$  and  $U$ . Eq. 10 shows that a site is selected as the sink site ( $T_i = 1$ ) and it must be included in the reserve ( $U_i = 1$ ). However, a selected site ( $U_i = 1$ ) may or may not be a sink site ( $T_i = 1$  or  $T_i = 0$ ). For each adjacent site, if the sites  $i$  and  $j$  are selected,  $Z_{i,j}$  is set to 1. Otherwise, it is set to 0, as shown in Eq. 13.

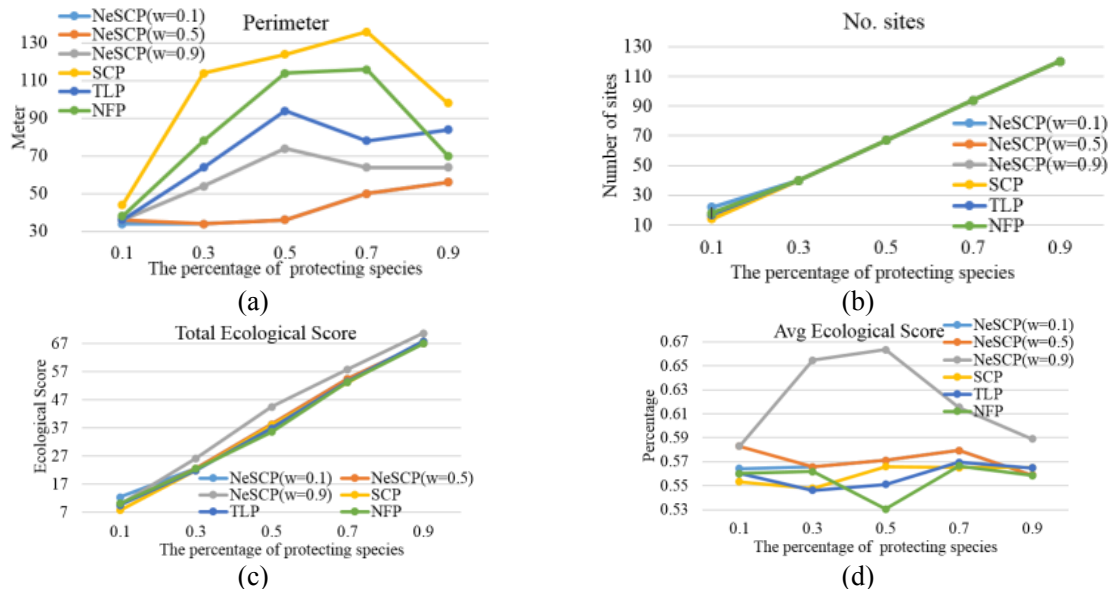
### 3. Results and discussion

The models described in Section 3 are applied to Daiyun Mountain in China for designing the conservation. To demonstrate the feasibility of the proposed model, we compare the proposed method with three popular site selection models, the species set covering problem (SSCP), the tail length problem (TLP) and the network flow problem (NFP).

We analyse the performance of each model by considering various percentages of each protecting species,  $k_s$ , which is set as 10%; 30%; 50%; 70%; 90%, with various weight  $\omega$ . In the experiment, the solver of branch and bound (B&B) algorithm is adopted to find the optimal solution of each model.

We evaluate the performance of each model with four indicators, perimeter, number of sites, total ecological score, and average ecological score and the experimental results are shown in figure 4. In

designing nature reserve, we prefer to have the tight plots which means the designed reserve has short perimeter. Although, various models have similar number of sites as shown in figure 4b, the proposed model has the shortest perimeter with various percentages  $\omega$  compare with other models in the same number of plots as shown in figure 4a.



**Figure 4.** The comparison results of each model. (a) perimeter; (b) number of sites; (c) total ecological score; (d) average ecological score.

Figures 4c and 4d show the total and average ecological scores of each model with various percentage of protecting species. In forest ecology system, the higher the ecological score has, the better the protection performance is. The proposed model with percentages  $\omega=0.9$  has better performance in both of total and average ecological score as shown in figures 4c and 4d.

**Table 2.** The visualization results of each model with various percentage of protecting species.































Model	Percentage of protecting species				
	0.1	0.3	0.5	0.7	0.9
SSCP					
TLP					
NFP					
SeSCP ( $\omega=0.1$ )					
SeSCP ( $\omega=0.5$ )					
SeSCP ( $\omega=0.9$ )					

Table 2 shows the visualization results of each model with various percentage of protecting species. Due to SSCP does not consider the spatial information, the characteristics of continuity and compactness are loss. Although TLP and FLP consider the spatial characteristic of continuity, the spatial property of compactness is missed.

#### 4. Conclusions

In this study, we proposed a new model, SeSCP, for solving the problem of selecting sites in the reserve network. SeSCP is based on the concept of the network flow problem which considers the property of spatial continuity and further considers the properties of spatial compactness and ecological information. The model avoids the problems of continuity and compactness compared with the other recent works, and also provides the ecological information for reserve network construction. The model aims to maximise the sum of the ecological values and to minimise the sum of the perimeter of the selected sites.

Three popular models including species set covering problem (SSCP), tail length problem (TLP), and network flow problem (NFP) are compared to demonstrate the performance of SeSCP. In spatial continuity, SSCP does not consider the spatial continuity and compactness; although TLP and NFP consider the characteristic of spatial continuity, the total and average ecological value are not the best. SeSCP considers the compactness property and ensures the selected site has the highest ecological value that is a tight reserve network.

In the future, we plan to integrate the remotely-sensed image into the ecological value which can efficiently reduce the manual resources investigation and consider the information of various species, such as animals, microorganism and funguses, to construct the complete ecosystem. Moreover, the proposed framework will further consider the core areas, corridors, buffer zones and nature development areas.

#### Acknowledgment

This work was supported by Fujian Provincial Department of Social and Technology, Natural Science Foundation of Fujian Province of China and Channel Post-doctoral Exchange Funding Scheme, under grants 2015Y0042 and 2016J01718.

#### References

- [1] Ehrlich and Paul R. 1992 Population biology of checkerspot butterflies and the preservation of global biodiversity *Oikos* **6(1)** 6-12.
- [2] Bedward M, Pressey R L and Nicholls A O 1991 Scores and score classes for evaluation criteria: a comparison based on the cost of reserving all natural features *Biol Conserv* **56(3)** 281-294
- [3] Pressey R L and Nicholls A O 1991 Reserve selection in the western division of new south wales: development of a new procedure based on land system mapping 1991 *Nature conservation: Cost effective biological surveys and data analysis* 98-108
- [4] Parga I C, Saiz J C M, Humphries C J and Williams P H 1996 Strengthening the natural and national park system of iberia to conserve vascular plants *Bot J Linn Soc* **121(3)** 189-205
- [5] Williams J C and ReVelle C S 1997 Applying mathematical programming to reserve selection *Environ Model Assess* **2(3)** 167-175
- [6] Camm J D, Norman S K, Polasky S and Solow A R 2002 Nature reserve site selection to maximize expected species covered *Oper Res* **50(6)** 946-955
- [7] Wang Y and Önal H 2015 Optimal design of compact and connected nature reserves for multiple species *Conserv Biol* **30(2)** 413-424
- [8] Diamond J M 1975 The island dilemma: lessons of modern biogeographic studies for the design of natural reserves *Biol Conserv* **7(2)** 129-146
- [9] Billionnet A 2012 Designing an optimal connected nature reserve *Appl. Math. Model* **36(5)** 2213-23
- [10] Jafari N, Nuse B L, Moore C T, Dilkina B and Hepinstall-Cymerman J 2017 Achieving full connectivity of sites in the multiperiod reserve network design problem *Computers and Operations Research* **81** 119-127
- [11] Önal H and Wang Y 2008 A graph theory approach for designing conservation reserve networks with minimal fragmentation *Networks* **51(2)** 142-152
- [12] Shirabe T 2009 Districting modeling with exact contiguity constraints *Environment and planning B Planning & design* **36(6)** 1053-1066