Predictions of the Potential Geographical Distribution and Quality of a *Gynostemma pentaphyllum* Base on the Fuzzy Matter Element Model in China

Zefang Zhao [1,2,3] (ID)**, Yanlong Guo [1], Haiyan Wei [1,*], Qiao Ran [1,2] and Wei Gu [3,4,*]**

[1] School of Geography and Tourism, Shaanxi Normal University, Xi'an 710119, China;
lh_zzf.nn@snnu.edu.cn (Z.Z.); guoyl1987@126.com (Y.G.); ranqiao@snnu.edu.cn (Q.R.)

[2] National Demonstration Center for Experimental Geography Education, Shaanxi Normal University,
Xi'an 710119, China

[3] National Engineering Laboratory for Resource Development of Endangered Crude Drugs in
Northwest of China, Shaanxi Normal University, Xi'an 710119, China

[4] College of Life Sciences, Shaanxi Normal University, Xi'an 710119, China

* Correspondence: weihy@snnu.edu.cn (H.W.); weigu@snnu.edu.cn (W.G.);
Tel.: +86-29-8531-0525 (H.W.); +86-29-8531-0266 (W.G.)

Abstract: Specific spatial information about medicinal plants is becoming an increasingly important part of their conservation. *Gynostemma pentaphyllum* (Thunb.) Makino is a traditional Chinese medical plant, and gypenosides is one of the main active components of *G. pentaphyllum*. In our research, many samples of *G. pentaphyllum* and the data of gypenosides content in these samples were collected from 43 sampling sites, and based on the Fuzzy Matter Element model (FME), the relationships between gypenosides content and 19 environmental variables were established. Then, the maximum entropy model was used to determine the relative importance of each environmental variable, and thus determine the most limiting habitat criteria. Finally, the weighted average method was applied to determine the potential distribution of *G. pentaphyllum* in China, which was based on the content of gypenosides. The results showed that the areas of marginally suitable and suitable habitats for *G. pentaphyllum* in China were approximately 1.2×10^6 km^2 and 0.3×10^6 km^2, respectively. The suitable habitats were mainly located in southern China, including Hunan, Hubei, Chongqing, Anhui, Jiangxi, Zhejiang, Shaanxi, and other regions. In conclusion, the FME model could accurately evaluate the habitat suitability of *G. pentaphyllum*, quantify the area of a suitable habitat, and analyze the spatial distribution.

Keywords: *Gynostemma pentaphyllum* (Thunb.) Makino; Fuzzy Matter Element model; gypenosides; GIS; membership function; habitat suitability

1. Introduction

Medicinal plants and their natural compounds play a significant role in the treatment of many diseases in many developing countries. The importance of medicinal plants is increasing rapidly due to their multiple benefits [1,2]. With the increasing demand of medicinal plants in domestic and international pharmaceutical markets and the influence of human activities, many medicinal plant species are severely threatened by over exploitation and habitat destruction [3]. Artificial cultivation is the most viable way of protecting the wild resources, while meeting the market demand. Meanwhile, successful artificial cultivation with a high quality of medicinal plants not only depends on good germplasm resources, but is also closely related to suitable environmental conditions [3,4]. Thus, determining the environmental requirements of the high quality of medicinal plants by correlating the

content of the effective components of the target species with the physical environment will facilitate more effective plant cultivation, whilst also being an important part of species conservation [5]. Additionally, cultivation is also based on the specific information on the habitat requirements of medicinal plants; we can map the distributions of the high quality of medicinal plants across landscapes and extrapolate these across space and time [3,4,6].

Gynostemma pentaphyllum (Thunb.) Makino is an herbaceous climbing plant in the gourd (god) family, which is a traditional Chinese medical plant. Its natural distribution is limited to damp areas under trees at an altitude that ranges between 200 and 3200 m. It is widely distributed in the south of China, Japan, Korea, India, Nepal, Malaysia, and other regions due to its strong ability to adapt to the environment [7]. During the growth process of *G. pentaphyllum* in the south of China, the period of sprouting leaves is from late March to early April. The leaf expansion period continues until the end of April, and is then followed by a peak-growth period. The blossom period starts between July and September. In mid-November, with the withering of the parts above the ground, the plant growth decreases [7–9]. In traditional Chinese medical science, the whole plant of *G. pentaphyllum* can be used as medicine. Modern pharmacology has confirmed that the gypenosides extracted from the *G. pentaphyllum* exhibits anti-tumor and anti-aging properties, as well as lowering cholesterol and enhancing immunity [8,9]. With a wide range of growth, high yield, and great market potential, *G. pentaphyllum* has been included in the List of National Protected Plants, and at the end of the last century, *G. pentaphyllum* became popular in Europe [10]. Currently, the research of *G. pentaphyllum* mainly focuses on pharmacological analyses, breeding germplasm resources, and population genetics [11–13]. However, only a few of them refer to the potential distribution area of this species, especially for high-quality *G. pentaphyllum*.

In recent years, most studies have focused on the plant species distribution and habitat suitability simulation, and statistical species distribution models (SDMs) are the main tools for such research [14]. Through building a correlation between species occurrence (or abundance) and environmental estimates, SDMs can predict distributions across landscapes. Over the past two decades, with technological advancements, numerous SDMs from a range of methods have become widely available for describing patterns and performing predictions [14]. Among these models, maximum entropy (MaxEnt) [15,16] is currently the most popular SDM, Bayesian networks (BN) have been increasingly used for the habitat suitability modelling of threatened species due to their potential to construct robust models with limited survey data [17], and the Fuzzy Matter Element model (FME) [3,4,6] can predict changes in the medicinal compositions of plants in different geographical conditions. When comparing the strengths and weaknesses of these three modelling techniques (Table 1), we found that the FME model was most suitable for this study.

In this study, the data of *G. pentaphyllum* samples at 43 sampling sites in China were collected and gypenosides was extracted as an index of the medicinal materials quality. Following this, we used a geographic information system (GIS) and FME model to estimate the spatial distribution of high-quality *G. pentaphyllum* in China. The objective was to construct maps of the potential distributions of high-quality *G. pentaphyllum*, and identify the key environmental factors influencing its ranges and the suitable habitat characteristics for *G. pentaphyllum* with a high content of gypenosides. Our results will supply advice for artificial planting and the choice of GAP (Good Agricultural Practices) base for *G. pentaphyllum*, establishing measures and proposals for the protection and sustainable utilization of this species.

Table 1. Strengths and weaknesses of MaxEnt, BN, and the FME model.

Model	Strengths and Weaknesses
MaxEnt [15,16]	**Strengths** MaxEnt prediction accuracy is always stable and reliable, even with incomplete data and small sample sizes. MaxEnt is a method that only needs presence data of species, and for environmental data both continuous and categorical data can be used as input variables. MaxEntis is free, and it has a user-friendly operational interface. **Weaknesses** MaxEnt cannot predict the change of plants' medicinal component contents in different geographical conditions.
BN [17,18]	**Strengths** The potential to use both expert knowledge and empirical data; an important advantage in cases of limited data availability. The suitability of BN models in a participatory modelling approach. Their explicit treatment of uncertainties. The availability of a variety of model validation tools. **Weaknesses** The directed acyclic graph (DAG) representing the dependencies between related nodes (variables) is the foundation of constructing BN models, however, for high quality of medicinal plants without certain criteria of habitat suitability, there is lacking sufficient expert knowledge or experimental data to build an authentic DAG. Their limited capacity to model systems mechanistically and dynamically. Exclusive use of discrete variables which often results in significant information loss.
FME [3,4,6]	**Strengths** FME model not only can predict the potential distribution of species, but also can predict the change of medicinal compositions of plant in different geographical conditions. Based on fuzzy theory, FME model determines the fuzzy membership functions according to the relationship between evaluation factors and active ingredient of plants. FME can standardize environmental variables with no prior knowledge as well as retain information about the original data. FME has good extrapolation performance even with incomplete data and small sample sizes. **Weaknesses** FME model not only needs presence data of species, but also need the data of effective components content for every samples site. The samples data in FME model were collected by field work; all the samples consumed extremely labor, material, and financial resources, especially in mountains or regions of complex terrain.

2. Materials and Methods

2.1. Species Data Collection

Sampling data of the species *G. pentaphyllum* were obtained from the research group of Prof. Yaping Xiao (Figure 1) at Shaanxi Normal University [13]. The research group collected the samples on the basis of relevant literature about the distribution of *G. pentaphyllum* throughout southern China, including the Shaanxi, Henan, Hubei, Hunan, Yunnan, Guizhou, Guangxi, Jiangxi, and Anhui provinces, from 2004 to 2010 (Figure 2). The sampling sites were selected on the basis of different environmental conditions when compared the traditional and original producing area of *G. pentaphyllum*. There were 43 sampling sites of *G. pentaphyllum* and more than 10 samples were collected from each site. The HPLC test was applied at each sampling site to examine the gypenosides content. We used the mean value of the gypenosides content of multiple samples at each sampling site to ensure that the data regarding the gypenosides content were representative for each region.

Figure 1. A picture of *G. pentaphyllum*. This picture was taken by Prof. Yaping Xiao in August 2016, in Pingli County, Shaanxi Province, China.
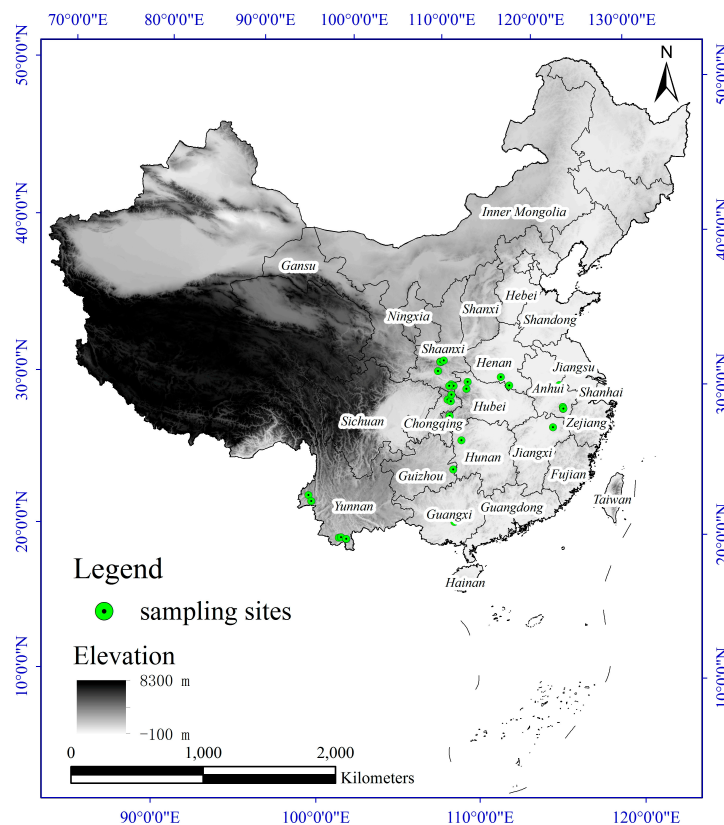

Figure 2. Study area and geographic locations of the *G. pentaphyllum* population distribution.

2.2. Collection and Selection of Environmental Variables

To determine the geographic distribution of suitable habitats for a target species, a set of environmental characteristics for this species must be defined [14,19–21]. Hence, three types of predictors, including 23 climatic variables, 10 soil variables, and three topographic variables, were chosen to simulate the distribution of high-quality *G. pentaphyllum*.

The climatic variables included two types of data: one was bioclimatic variables, which were downloaded from the WorldClim database at a resolution of 30″ (approximately 1 km^2). The bioclimatic variables included 19 bioclimatic variables, which were derived from the monthly temperature and

rainfall values. Previous studies have indicated that bioclimatic variables are useful for SDMs, and these variables are one of the most widely used environmental variables in the practice of SDMs [22,23]. The other types of climatic variables were downloaded from the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (RESDC) (http://www.resdc.cn), including an accumulated temperature above 0 °C (AAT0), accumulated temperature above 10 °C (AAT10), the aridity index (AI), and the moisture index (MI), at a resolution of 1 km^2. The climatic variables were derived from a spatial interpolation method which was based on the average value of weather station records over the last few years. Hence, these data can represent climatic conditions in the study area. The soil variables were obtained from the harmonized world soil database (HWSD) (http://www.fao.org/soils-portal/en) [24], at a resolution of approximately 1 km^2. The HWSD soil data incorporate the 1:1,000,000 scales Soil Map of China and other data sources, ensuring the accuracy and reliability of the data.

Topographic variables including the elevation, slope, and aspect are important factors for the distribution patterns of vegetation. In this study, the elevation variable was provided by RESDC at a resolution of 1 km^2. The slope and aspect variables were obtained via the ArcGIS spatial analysis function based on the elevation variables.

In this study, principal component analysis (PCA) and correlation analysis were used for reducing the influence of multi-collinearity of the environmental variables [25]. Furthermore, expert knowledge and biological characteristics were also considered for the selection of environmental variables. The detailed analysis procedure is given below:

Firstly, we used PCA to discern 23 climatic variables, and the results showed that the first three principal components obtained with the PCA explained a total of 93.12% of the variation. Based on the 15 climatic variables, those with the largest percentage of variance in the principal component loadings were screened out. Following this, we ascertained 12 climatic variables for model building (Table 2) after excluding some of the variables on the basis of the growth condition of *G. pentaphyllum* and a Pearson cross-correlation coefficient value of >±0.7.

Secondly, the soil variables, including Topsoil organic carbon (TOC), Topsoil pH (H$_2$O) (TPH), Topsoil USDA Texture Classification (TTEX), and Subsoil pH (H$_2$O) (S_PH) for model building, were selected on the basis of previous literature and the biological characteristics of *G. pentaphyllum*. Moreover, the three topographic variables were added, resulting in a total of 19 environmental variables for model building analysis, as shown in (Table 2).

Table 2. Explanatory variables used to model the distribution of *G. pentaphyllum*.

Geographical Environment Index	Code	Name
	AAT10	Accumulated temperature above 10 °C
	AI	Aridity index
	MI	Moisture index
Climatic factor	Bio1	Annual mean temperature
	Bio5	Max temperature of warmest month
	Bio6	Min temperature of coldest month
	Bio7	Temperature annual range
	Bio8	Mean temperature of wettest quarter
	Bio11	Mean temperature of coldest quarter
	Bio12	Annual precipitation
	Bio14	Precipitation of driest month
	Bio16	Precipitation of wettest quarter
Soil factor	TOC	Topsoil organic carbon
	TPH	Topsoil pH(H$_2$O)
	TTEX	Topsoil USDA Texture Classification
	S_PH	Subsoil pH(H$_2$O)
Topographical factor	ASL	Elevation above sea level
	SLOP	Slope
	ASPE	Aspect

2.3. Model Evaluation

2.3.1. Fuzzy Matter Element Model

Fuzzy matter-element theory was founded in 1980s by the Chinese mathematician Cai [26], and is a combination of Fuzzy mathematics theory and matter-element theory. With reference to matter-element theory, in order to provide explicit conceptualizations for the practical problems which cannot be described by precise mathematical language, a correlation function was used to produce a quantitative description of the process of quantitative and qualitative change [6]. The matter-element (*R*) consists of three fundamental elements, which can be defined as equation (Equation (1)):

$$R = (N, C, V) \tag{1}$$

where *N* is the matter name, *C* represents the matter characteristics, and *V* represents the values of matter characteristics.

In the Equation (1), if *V* is a fuzzy set, then *R* should be defined as a fuzzy matter element. Furthermore, we defined R_{nm} (Equation (2)) as a compound matter element which contains *n* dimensions of *m* matter elements.

$$R_{nm} = \begin{bmatrix} & M_1 & M_2 & \cdots & M_m \\ C_1 & \mu_{11} & \mu_{12} & \cdots & \mu_{1m} \\ C_2 & \mu_{21} & \mu_{22} & \cdots & \mu_{2m} \\ & & & \cdots & \\ C_n & \mu_{n1} & \mu_{n2} & \cdots & \mu_{nm} \end{bmatrix} \tag{2}$$

where μ_{ij} ($i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, m$) is the value of the *j*th matter characteristics of the *i*th matter elements.

2.3.2. Establishing the Fuzzy Correlation Function

For high-quality *G. pentaphyllum*, there is no prior knowledge about its explicit habitat requirements, and thus, it can be considered as a fuzzy matter element. The suitability of each environmental factor to the matter-element is fuzzy, so it is necessary to establish a fuzzy membership function to standardize the variables and correlative transformation. In this study, to ensure consistency between the sampling data and the environmental data, we extracted environmental variable values from the sampling sites, producing environmental data sequences for model training [4]. Using the statistical association between the data of gypenosides and the value of environmental variables from the samples sites, we built fuzzy membership for every environmental variable to quantify the relation between the value of environmental variables and habitat suitability.

During the modelling process, 80% of the sample sites data were used for model calibration, and the remaining sample sites data were used to test the predictive ability of the model. According to the data characteristics and biological characteristics of *G. pentaphyllum*, the Gaussian membership function (Equation (3)) and the Trapezoidal membership function (Equations (4) and (5)) were used to normalize each environmental variable, and we used Fitting degree (R-square) to control the fitting error (Table 3).

$$A(x) = e^{-\left(\frac{x-a}{\sigma}\right)^2}, \ -\infty < x < \infty \tag{3}$$

where *x* is the independent variable, and *a*, σ are the parameters.

$$A(x) = \begin{cases} 1, x < a \\ \dfrac{b-x}{x-a}, a \le x \le b \\ 0, b < x \end{cases} \tag{4}$$

$$A(x) = \begin{cases} 0, x < a \\ \dfrac{x-a}{b-a}, a \leq x \leq b \\ 1, b < x \end{cases} \tag{5}$$

In Equations (4) and (5), where x is the independent variable, a, b are the parameters.

For the Topsoil USDA Texture Classification (TTEX), the grade value has been determined by means of pairwise comparison methods [3] (Table 4). The membership function illustrates how the grade of membership changes as each environmental variable changes. For this reason, the optimum and threshold values for the environmental variables were calculated within the distribution area of high-quality *G. pentaphyllum*.

2.3.3. Correlative transformation

According to the fuzzy membership function of Tables 3 and 4, the correlation coefficient is constructed, presenting the fuzzy matter element R_ζ:

$$R_\xi = \begin{bmatrix} & M_1 & M_2 & \cdots & M_m \\ C_1 & \xi_{11} & \xi_{12} & \cdots & \xi_{1m} \\ C_2 & \xi_{21} & \xi_{22} & \cdots & \xi_{2m} \\ & & & \cdots & \\ C_n & \xi_{n1} & \xi_{n2} & \cdots & \xi_{nm} \end{bmatrix} \tag{6}$$

where $\xi_{ij} = \mu_{ij}$ ($i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, m$), which represents the value of the ith sample site membership grade to the jth environmental variables after normalization by the fuzzy membership function.

Table 3. Membership function of the evaluation factors and fitting degree.

Geographical Environment Index	Membership Function	Fitting Degree (%)
AAT10	Gaussian-type	83.89
AI	Gaussian-type	77.70
MI	L trapezoidal	83.36
Bio1	Gaussian-type	72.59
Bio5	Gaussian-type	81.60
Bio6	Subsection Gaussian-type	83.52
Bio7	Gaussian-type	70.00
Bio8	Gaussian-type	70.92
Bio11	Gaussian-type	91.11
Bio12	Gaussian-type	75.73
Bio14	Subsection Gaussian-type	76.42
Bio16	Gaussian-type	70.92
TOC	L trapezoidal	69.39
TPH	Gaussian-type	77.34
S_PH	Gaussian-type	75.29
ASL	Gaussian-type	80.24
SLOP	D trapezoidal	85.64
ASPE	Subsection Gaussian-type	64.20

Table 4. Fitting degree of TTEX.

TTEX Type	Fitting Degree
silt loam	1
loam	0.76
sandy clay loam	0.73
sandy loam	0.58
others	0

2.3.4. Determine the Weights of Environmental Variables

To eliminate the subjective influence, the maximum entropy method (MEM) was used to determine the weights of these environmental variables. In this research, information entropy is defined by Equation (7) [27,28]:

$$E = -\sum_{i=1}^{n}\sum_{j=1}^{m}\mu_{ij}ln\mu_{ij} \tag{7}$$

where μ_{ij} is the same as Equation (6), m is the number of training samples, and n is the number of environmental variables.

Then, the entropy value of the jth indicator can be calculated using Equation (8):

$$e_j = -k\sum_{j=1}^{m}\frac{\mu_{ij}}{\mu_i}ln\frac{\mu_{ij}}{\mu_i} \tag{8}$$

where $u_i = \sum_{j=1}^{m} u_{ij}$, $k = \frac{1}{lnm}$.

The weighting value of the jth indicator is defined as Equation (9):

$$w_j = \frac{1}{n-E}\left[1 - e(\mu_{ij})\right] \tag{9}$$

2.3.5. Calculation of Habitat Suitability Evaluation Index

R_k is the matter element which has m matter characteristics and R_w is the information entropy weight vector. Then, the habitat suitability evaluation index is defined as Equation (10):

$$R_k = R_w \qquad R_\xi = \begin{bmatrix} & M_1 & M_2 & \cdots & M_m \\ K_j & K_1 & K_2 & \cdots & K_m \end{bmatrix} \tag{10}$$

where $K_i = \sum_{i=1}^{n} W_j\xi_{ij}$ represents the fuzzy operator in the formula uses (, +), and K_i refers to the value of the habitat suitability evaluation index in the ith evaluation unit, W_j is the weighting value of the jth environmental variables, and $\xi_{ij} = \mu_{ij}$ is the same as Equation (6). The range of K_i is 0–1, and a higher value means that the evaluation units are more suitable for the growth of *G. pentaphyllum* and the synthesis and accumulation of gypenosides.

2.3.6. Model Validation

The accuracy of the models is evaluated using the root mean square error (RMSE) [29], which is defined as Equation (11):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - Q_i)^2} \tag{11}$$

where N represents the amount of test data, P_i is the prediction value of the membership of the ith test sample, and Q_i is the actual value of the membership of the ith test sample. RMSE is a widely used evaluation index, where the smaller the RMSE is, the higher the model precision [30].

2.3.7. Mapped the Distribution of Habitat Suitability

For further analysis, with ArcGIS 9.3 (ESRI, Redlands, CA, USA) support, we classified the model results into three classes by the value of the habitat suitability evaluation index: unsuitable habitats ($K_i < 0.3$), suitable habitats ($0.3 \leq K_i < 0.6$), and highly suitable habitats ($K_i \geq 0.6$). We mapped the potential distribution of habitat suitability of *G. pentaphyllum* in China (Figure 3) calculated the primary area of habitat suitability distribution of *G. pentaphyllum* in different provinces (Figure 4).
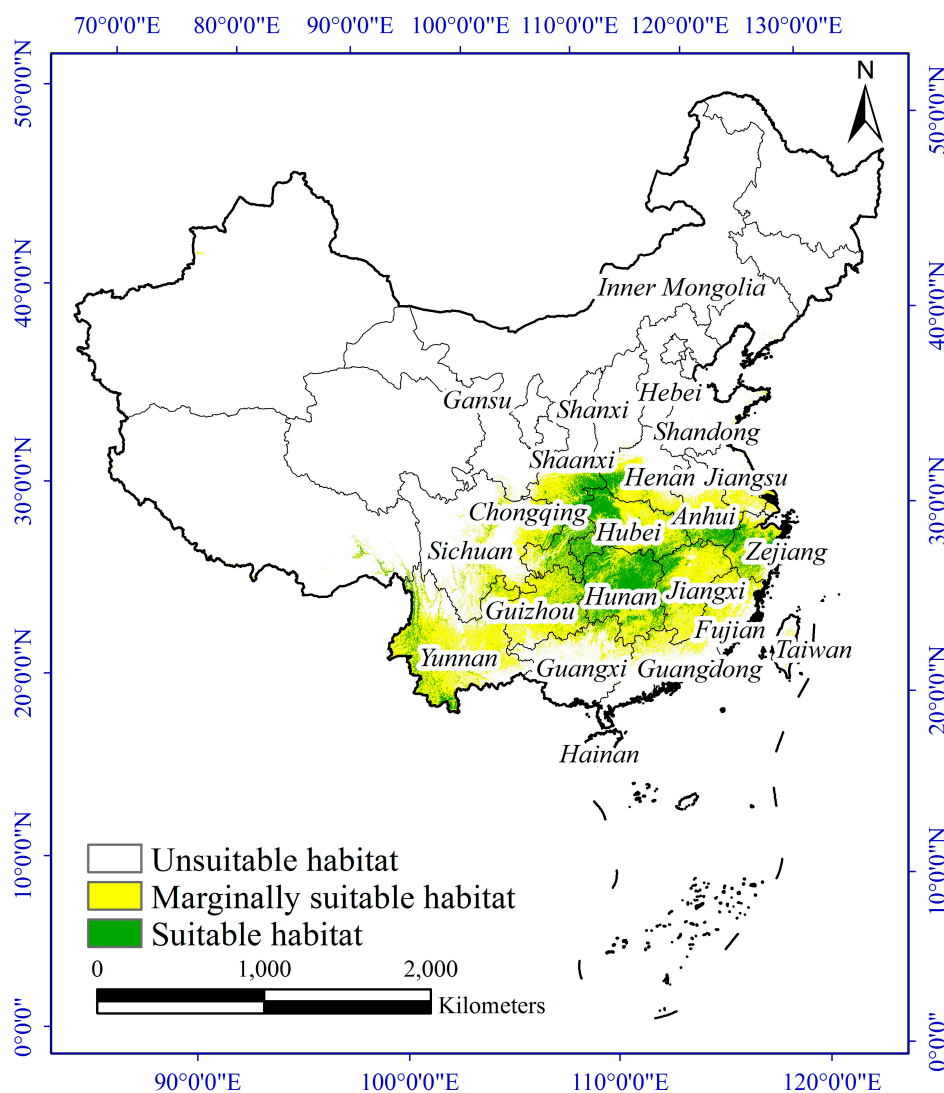
Figure 3. Distribution map of the potential suitability of *G. pentaphyllum*.
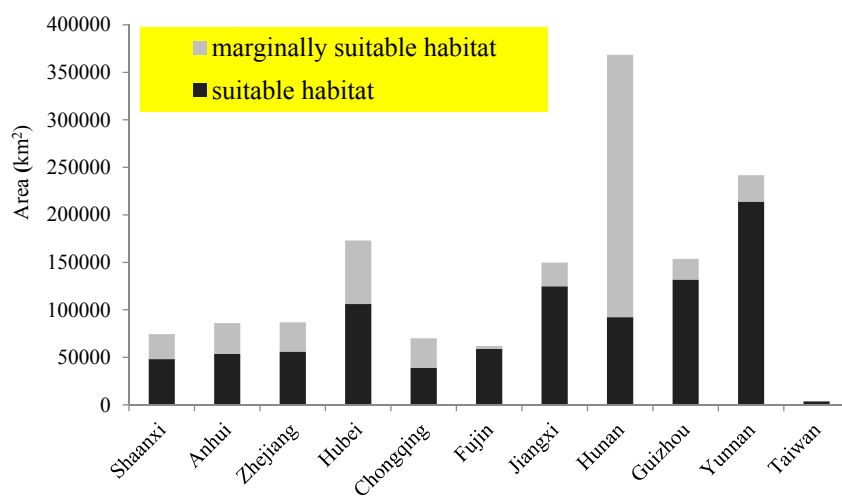

Figure 4. The primary areas of habitat suitability distribution of *G. pentaphyllum* in different provinces.

3. Results

3.1. The Determination of the Fuzzy Membership Functions and Model Validation

The membership function was established to explore the relationships between factors and membership degrees by cure-fitting and drawing scatter diagrams [4]. However, the interaction between the content of gypenosides and the value of environmental variables was quite different. According to the shape characteristics of the fitting curve and the graphical features of different types of membership functions, we chose different fuzzy membership functions for every environment variable. For most environmental variables, the suitable range for the growth of *G. pentaphyllum* has upper and lower limits, so we chose Gaussian-type functions to demonstrate these relationships. As for the environmental variables, which only have lower limits, we chose the L trapezoidal function as the fuzzy membership function. For environmental variables, which only have upper limits, we selected the R trapezoidal function as the fuzzy membership function. The value of the average fitting degree of all the membership functions in this study was 77.21%, which varied from 64.20% to 91.11% (Table 3). Moreover, the RMSE of this model was 9.75%. Usually, if RMSE < 10%, the simulation is considered excellent [29]. Thus, the model results could be considered satisfactory.

According to the fuzzy membership function, we defined the optimum and threshold values for every environmental variable. When the membership degree reaches its maximum, the value of the environmental variables is the optimum variables, and when the membership degree >0.3, the range of environmental variables is the threshold variables (Table 5).

Table 5. The optimal value, threshold value, and weights of each environment index.

Geographical Environment Index	Range and Units	Optimum Value	Threshold Value	Weights (%)
AAT10	0–10460 (d·°C)	5300	3500–7158	8.07
AI	0–119	7.7	5.2–10.2	5.04
MI	0–259	253	150–259	4.09
Bio1	–22.7–26 (°C)	>17	>9.8	7.56
Bio5	–8–41.4 (°C)	32.8	23.3–34	4.54
Bio6	–35–5.4 (°C)	3.25	0.14–4.2	10.10
Bio7	13.7–61.5 (°C)	32.8	20.08–36	3.99
Bio8	–15–31.9 (°C)	14.9	16.7–27.2	6.65
Bio11	–30.1–22.2 (°C)	7.6	0–14.3	7.33
Bio12	12–5054 (mm)	1010	850–1360	4.96
Bio14	0–229 (mm)	120	34–170	3.84
Bio16	5–3140 (mm)	530	350.2–910.7	7.25
TOC	0–3.02 (%)	0.95	0.6–1.5	6.67
TPH	4.7–8	6.8	5.5–7.6	5.94
TTEX	9 types: clay(light); silty clay loam, clay loam, silt loam, loam; sandy clay loam; sandy loam; loamy sand, sand	loam	silt loam loam sandy clay loam sandy loam	1.74
S_PH	4.5–8.1	6.3	5.5–8.1	1.01
ASL	0–8300 (m)	330	60–1670	5.40
SLOP	0–48 (°)	<5	<20	1.01
ASPE	0–359.95 (°)	290	0–359.95	4.81

3.2. Distribution of Suitable Habitats

The model results indicated that the high-quality *G. pentaphyllum* was mainly distributed in southern China (Figure 3), including Hunan, Hubei, Chongqing, Anhui, Jiangxi, Zhejiang, Shaanxi, and Taiwan, etc. (Table 6). According to the results of the habitat suitability grades distribution, the marginally suitable habitat of *G. pentaphyllum* is approximately 1.19×10^6 km^2, and the suitable habitat is only 0.42×10^6 km^2, which accounted for 4.38% of the land area of China. The suitable habitats are mainly located in southeastern Shaanxi, southwestern Henan, southern Anhui, western Zhejiang,

eastern Chongqing, northwest Hubei, central and northern Hunan, and their surrounding areas. The largest area of suitable habitat occurs from Shaanxi to Hunan, covering an area of approximately 0.32×10^6 km^2, accounting for 76% of the suitable habitats of *G. pentaphyllum* in China. These areas should be the first choice for the artificial planting base location of *G. pentaphyllum*. The marginally suitable habitat area covers the Qinba mountain area, the middle and lower reaches of the Yangtze River, the Yunnan-Guizhou Plateau, the northern part of the Pearl River Basin, and the middle part of the Taiwan province.

The model indicated that, when the content of gypenosides reached its maximum, the characteristic of the habitat should be as follows: 3.25 °C for Bio6 (Minimum temperature of the coldest month), 5300 d·°C for AAT10 (Accumulated temperature above 10 °C), more than 17 °C for Bio1 (Annual mean air temperature), 7.6 °C for Bio11 (Mean temperature of the coldest quarter), 530 mm for Bio16 (Precipitation of the wettest quarter), 14.9 °C for Bio8 (Mean temperature of the wettest quarter), 0.95% for TOC (Topsoil organic carbon), 6.8 for TPH (Topsoil pH), and 330 m for ASL (Elevation above sea level).

Table 6. The percentage of areas of habitat suitability distribution of *G. pentaphyllum* in different provinces.

Province	Unsuitable Habitat (%)	Marginally Suitable Habitat (%)	Suitable Habitat (%)
Shaanxi	63.80	23.39	12.81
Anhui	38.79	38.23	22.98
Zhejiang	12.46	56.51	31.03
Hubei	7.08	56.99	35.93
Chongqing	13.67	48.06	38.27
Fujin	47.95	49.48	2.57
Jiangxi	10.02	74.94	15.04
Hunan	1.04	43.69	55. 27
Guizhou	12.42	75.25	12.33
Yunnan	36.50	56.19	7.31
Taiwan	89.14	10.64	0.22

4. Discussion

4.1. The Rationality of the Model

Highly accurate and reliable SDMs model results depends on authentic species data, the relevance completeness of the environmental variables, and the rationality of the mathematical algorithm [14,31]. In this study, we chose field surveys data as species distribution data to ensure the authenticity and accuracy of the species distribution records. Only under the circumstances of appropriate temperatures and sufficient water can plants exhibit a normal growth and accumulate particular organics [4]. Therefore, reliable and sufficient environmental variables for the environmental characteristics of a high-quality medicinal plants distribution area are the decisive factor for model building. Three types of environmental variables were selected to present the feature of suitable habitat characteristics, and we used principal component analysis (PCA), correlation analysis, expert knowledge, and biological characteristics to select the environmental variables. This ensures that the environmental variables involved in modelling have no multi-collinearity and can effectively and reliably describe the environmental characteristics.

FME is derived from fuzzy theory, and it represents a very useful tool for the simulation of high-quality medic plant distributions [3,4,6]. During the modelling procedure, determining the membership function is a key step [32–34], and usually, researchers determine the type and parameters of fuzzy membership functions based on expert knowledge [35,36]. However, in this study, in order to avoid the influence of subjective factors, we used the information on the relationship between the value of 19 environmental variables and the gypenosides content of *G. pentaphyllum* samples to choose the membership function type and calculate the parameters. Hence, the FME model can standardize variables with no prior knowledge, as well as retain information about the original data.

In practice, the FME model is a universal model for the simulation of the quality of wild herbs in different environmental conditions, and this model only depends on the information provided by the statistical data, determining the mathematical relationship between the value of each impact factor and the content of effective composition in medicinal plants. It represents a tractable modelling method when knowledge about the habitat requirements of some species is insufficient [3]. In the modelling process, the FME model needs the data of the effective components content for every sample site. This required all of the sample data in these models to be derived from field work, and the content of the medicinal composition of the plant must then be tested. A large amount of labor, material, and financial resources are consumed during this process, and then, we usually only get limited data. Thus, only limited sample data are provided for the FME model, which will increase the model uncertainty to some extent [4]. Future research should collect more sample data, and try to combine the FME model with other machine learning models, in order to make the modelling more accurate and exact.

4.2. The Dominant Variables' Response to Suitability

The quality and efficacy of medicinal plants are closely related to their geographical environment [37,38]. In this research, the estimates of the relative contributions of the environmental variables indicate that, among the climatic variables, Bio6 (Minimum temperature of the coldest month), AAT10 (Accumulated temperature above 10 °C), Bio1 (Annual mean air temperature), Bio11 (Mean temperature of the coldest quarter), Bio16 (Precipitation of the wettest quarter), and Bio8 (Mean temperature of the wettest quarter) are the key variables, with contribution rates 10.10%, 8.07%, 7.56%, 7.33%, 7.25%, and 6.65%, respectively. For soil variables, the key variables determining the distribution of high-quality *G. pentaphyllum* are the TOC (Topsoil organic carbon) and TPH (Topsoil pH(H_2O)), with contribution rates of 6.67% and 5.94%, respectively. In addition, the ASL (Elevation above sea level) and ASPE (Aspect) are the key variables among the topographical variables, with contribution rates of 5.40% and 4.81%, respectively.

According to the fuzzy membership functions, we calculated the optimal value (membership degree = 1) and the suitable ranges of environmental variables for the distribution area of *G. pentaphyllum* (membership degree > 0.5) [4]. The results showed that the optimal values are 5300 d·°C for AAT10, 17 °C for Bio1, 3.25 °C for Bio6, 7.6 °C for Bio11, 14.9 °C for Bio8, and 530 mm for Bio16. These data summarized the environmental characteristics of high-quality *G. pentaphyllum*, namely, warm humid climates. The ranges are 16.7–27.2 °C for bio8 and 350–910 mm for Bio16, and these data indicate that during the summer growing season, low temperatures and drought would be disadvantageous for the growth of *G. pentaphyllum* and the synthesis and accumulation of gypenosides.

Li and Zhou [39] produced a preliminary summary of the climatic environmental characteristics in *G. pentaphyllum*-producing areas. They reported that in such areas, the accumulated temperature above 10 °C was greater than 3660 d·°C, the annual average temperature was 14–25 °C, the annual precipitation was above 850 mm, and the precipitation in the summer growing season accounts for more than 80% of the annual precipitation. This record was consistent with the results of our research. Because of the shallow fleshy stem roots, it is easy to cause damage to *G. pentaphyllum* under the condition of an extremely low temperature and continuous low temperature in winter. The results show that the suitable ranges are 0.14–4.2 °C for Bio6 and 0–14.3 °C for Bio11, and these data describe the climatic characteristic of a milder winter, consistent with the biological characteristics of *G. pentaphyllum*.

This study indicated that the silt loam was the most suitable for the accumulation of gypenosides content of *G. pentaphyllum*, followed loam and sandy clay loam. At the same time, the suitable range of TPH is 5.5–7.6 and the optimal value is 6.8, and the TOC optimal value is 0.95%, which means a high soil carbon content. These data describe the soli characteristics of suitable habitats for *G. pentaphyllum*, namely, subacidity and a high carbon content in soil. Pang [40] studies have shown that sandy loam soil and humus soil with a high carbon content are the most appropriate soil types for the growth

of *G. pentaphyllum*, and the most appropriate soil pH for this plant was 6.5–7.0. This record is also consistent with the results of our study.

According to the result of membership function, the suitable range for altitude is 60–1670 m, and Zhang [41] has shown that the most appropriate altitude for the growth of *G. pentaphyllum* is 200–2000 m; our results show that the optimal aspect value is 290 degrees, which means that the northwest is suitable for the growth of *G. pentaphyllum*. It is consistent with the biological characteristics of *G. pentaphyllum*, which love shadowy places.

5. Conclusions

We have successfully predicted the suitable habitats of high-quality *G. pentaphyllum* in China. Moreover, the FME model has provided satisfactory results based on the data of the gypenosides content of *G. pentaphyllum* derived from field work and three types of environmental factors, which are selected from the result of PCA, correlation analysis, expert knowledge, and biological characteristics. According to the result of membership function, we can obtain suitable range of each factor for *G. pentaphyllum*, and this information is significant for the management and protection of *G. pentaphyllum* resources. Based on the results of this research, combined with the environmental characteristics of key variables which have greatly influenced the distribution of *G. pentaphyllum*, we can make a reasonable plan for the introduction and cultivation of these species.

The ability of drought resistance of *G. pentaphyllum* is poor, and only in the condition of high humidity and a high soil moisture content can it grow well. Hence, when planting in the field, we should first select the suitable habitat area for *G. pentaphyllum*, and then conserve soil and water, improve soil moisture, and during the winter cold season, insulation measures should be taken. Finally, we need to provide a good developmental situation for *G. pentaphyllum* by providing shading conditions or intercropping with other tall crops and fruit trees.

Author Contributions: Z.Z., H.W., Y.G., and W.G. collected data and processed the input data. Z.Z. and Y.G. performed the modelling and analyzed the data. Z.Z., H.W., and W.G. organized the paperwork and all of the authors contributed to the writing of each part.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Beeran, A.A.; Maliyakkal, N.; Rao, C.M.; Udupa, N. The enriched fraction of *Elephantopus scaber* Triggers apoptosis and inhibits multi-drug resistance transporters in human epithelial cancer cells. *Pharmacogn. Mag.* **2015**, *11*, 257–268. [PubMed]
2. Williamson, E.M.; Lorenc, A.; Booker, A.; Robinso, N. The rise of traditional Chinese medicine and its materia medica: A comparison of the frequency and safety of materials and species used in Europe and China. *J. Ethnopharmacol.* **2013**, *149*, 453–462. [CrossRef] [PubMed]
3. Lu, C.Y.; Gu, W.; Dai, A.H.; Wei, H.Y. Assessing habitat suitability based on geographic infomation system (GIS) and fuzzy: A case study of *Schisandra sphenanthera* Rehd. et Wils. in Qinling Mountains, China. *Ecol. Model.* **2012**, *242*, 105–115. [CrossRef]
4. Guo, Y.L.; Wei, H.Y.; Lu, C.Y.; Gao, B.; Gu, W. Predictions of potential geographical distribution and quality of *Schisandra sphenanthera* under climate change. *Peerj* **2016**, *4*. [CrossRef] [PubMed]
5. Lentini, P.E.; Wintle, B.A. Spatial conservation priorities are highly sensitive to choice of biodiversity surrogates and species distribution model type. *Ecography* **2015**, *38*, 1101–1111. [CrossRef]
6. Guo, Y.L.; Wei, H.Y.; Gu, W.; Zhang, H.L. Potential distributions of *Sinopodophyllum hexandrum* based on fuzzy matter element model. *Acta. Ecol. Sin.* **2015**, *35*, 770–778. (In Chinese)
7. Committee of Flora of China. Angiospermae. In *Flora of China*; Lu, A.M., Chen, S.K., Eds.; Science Press: Beijing, China, 1986; Volume 73, pp. 265–277. (In Chinese)

8.　Chang, C.K.; Chang, K.S.; Lin, Y.C.; Liu, S.Y.; Chen, C.Y. Hairy root cultures of *Gynostemma pentaphyllum* (Thunb.) Makino: A promising approach for the production of Gypenosides as an alternative of ginseng saponins. *Biotechnol. Lett.* **2005**, *27*, 1165–1169. [CrossRef] [PubMed]

9.　Chew, Y.L.; Wong, H.C. Gypenosides, the cancer buster from *Gynostemma pentaphyllum*, (Thunb.) Makino and the apoptotic pathways: A review. *OPEM* **2016**, *16*, 1–12. [CrossRef]

10.　Cui, J.F.; Eneroth, P.; Bruhn, J.G. *Gynostemma pentaphyllum*: Identification of major sapogenins and differentiation from Panax, species. *Eur. J. Pharm. Sci.* **1999**, *8*, 187–191. [CrossRef]

11.　Mishra, R.N.; Joshi, D. Jiao Gu Lan (*Gynostemma pentaphyllum*): The Chinese rasayan-current research scenario. *Int. J. Res. Pharm. Biomed. Sci.* **2011**, *2*, 1483–1502.

12.　Lv, Y.; Yang, X.B.; Zhao, Y.; Zhao, Y.; Ruan, Y.; Yang, Y.; Wang, Z.Z. Separation and quantification of component monosaccharides of the tea polysaccharides from *Gynostemma pentaphyllum* by HPLC with indirect UV detection. *Food. Chem.* **2009**, *112*, 742–746. [CrossRef]

13.　Le, Y. Study on Saponin Content in *Gynostemma pentaphyllum*. Master's Thesis, Shaanxi Normal University, Xi'an, China, 2010. (In Chinese)

14.　Elith, J.; Leathwick, J.R. Species distribution models: Ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. S.* **2009**, *40*, 677–697. [CrossRef]

15.　Phillips, S.J.; Anderson, R.P.; Schapire, R.E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **2006**, *190*, 231–259. [CrossRef]

16.　Guo, Y.L.; Li, X.; Zhao, Z.Z.; Wei, H.; Gao, B.; Gu, W. Prediction of the potential geographic distribution of the ectomycorrhizal mushroom *Tricholoma matsutake* under multiple climate change scenarios. *Sci. Rep.* **2017**, *7*, 46221. [CrossRef] [PubMed]

17.　Tantipisanuh, N.; Gale, G.A.; Pollino, C. Bayesian networks for habitat suitability modeling: A potential tool for conservation planning with scarce resources. *Ecol. Appl.* **2016**, *24*, 1705–1718. [CrossRef]

18.　Pérez-Miñana, E. Improving ecosystem services modelling: Insights from a Bayesian network tools review. *Environ. Model. Softw* **2016**, *85*, 184–201. [CrossRef]

19.　Ashraf, U.; Ali, H.; Chaudry, M.N.; Ashraf, I.; Batool, A.; Saqib, Z. Predicting the potential distribution of *Olea ferruginea* in Pakistan incorporating climate change by using maxent model. *Sustainability* **2016**, *8*, 722. [CrossRef]

20.　Márquez, A.L.; Real, R.; Olivero, J.; Eatrada, A. Combining climate with other influential factors for modelling the impact of climate change on species distribution. *Clim. Chang.* **2011**, *108*, 135–157. [CrossRef]

21.　Yang, X.Q.; Kushwaha, S.P.S.; Saran, S.; Xu, J.C.; Roy, P.S. Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills. *Ecol. Eng.* **2013**, *51*, 83–87. [CrossRef]

22.　Rödder, D.; Lötters, S. Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (*Hemidactylus turcicus*). *Glob. Ecol. Biogeogr.* **2009**, *18*, 674–687.

23.　Ranc, N.; Santini, L.; Rondinini, C.; Boitani, L.; Poitevin, F.; Angerbjörn, A.; Maiorano, L. Performance tradeoffs in target–group bias correction for species distribution models. *Ecography* **2016**, *39*, 1–12. [CrossRef]

24.　Food and Agriculture Organization of the United Nations (FOA): Harmonized World Soil Database (version1.1). Available online: http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en (accessed on 18 September 2014).

25.　Giannini, T.C.; Takahasi, A.; Medeiros, M.C.M.P.; Saraiva, A.M.; Alves-dos-Santos, I. Ecological niche modeling and principal component analysis of *Krameria, Loefl.* (Krameriaceae). *J. Arid Environ.* **2011**, *75*, 870–872. [CrossRef]

26.　Cai, W. The extension set and incompatibility problem. *J. Sci. Explor.* **1983**, *1*, 81–93. (In Chinese)

27.　Ding, L.; Shao, Z.F.; Zhang, H.C.; Xu, C.; Wu, D.W. A comprehensive evaluation of urban sustainable development in China based on the TOPSIS-Entropy method. *Sustainability* **2016**, *8*, 746. [CrossRef]

28.　Wang, Q.S.; Yuan, X.L.; Zhang, J.; Gao, Y.; Hong, J.L.; Zuo, J.; Liu, W. Assessment of the sustainable development capacity with the entropy weight coefficient method. *Sustainability* **2015**, *7*, 13542–13563. [CrossRef]

29.　Jamieson, P.D.; Porter, J.R.; Wilson, D.R. A test of the computer simulation model arcwheat1 on wheat crops grown in New Zealand. *Field Crop Res.* **1991**, *27*, 337–350. [CrossRef]

30. Wang, S.; Wang, Q.B.; Adhikari, K.; Jia, S.H.; Jin, X.X.; Liu, H.B. Spatial-temporal changes of soil organic carbon content in Wafangdian, China. *Sustainability* **2016**, *8*, 1154. [CrossRef]

31. Elith, J.; Graham, C.H. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* **2009**, *32*, 66–77. [CrossRef]

32. Rocchini, D. Fuzzy species distribution models: a way to represent plant communities spatially. *J. Veg. Sci.* **2014**, *78*, 317–318. [CrossRef]

33. Mouton, A.M.; Alcaraz-Hernández, J.D.; De-Baets, B.; Goethals, P.L.M.; Martínez-Capel, F. Data-driven fuzzy habitat suitability models for brown trout in Spanish Mediterranean rivers. *Environ. Model. Softw.* **2011**, *26*, 615–622. [CrossRef]

34. Lee, Y.C.; Chou, C. Technology evaluation and selection of 3DIC integration using a Three-Stage Fuzzy MCDM. *Sustainability* **2016**, *8*, 114. [CrossRef]

35. Zhu, A.X.; Yang, L.; Li, B.L.; Qin, C.Z.; Pei, T.; Liu, B.Y. Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma* **2010**, *155*, 164–174. [CrossRef]

36. Wang, C.H.; Hong, T.P.; Tseng, S.S. Integrating membership functions and fuzzy rule sets from multiple knowledge sources. *Fuzzy Sets Syst.* **2000**, *112*, 141–154. [CrossRef]

37. Deng, S.; Chen, S.N.; Yang, J. Chemistry of medicinal plants, foods, and natural products 2015. *J. Anal. Methods Chem.* **2015**. [CrossRef] [PubMed]

38. Street, R.A.; Stirk, W.A.; Van Staden, J. South African traditional medicinal plant trade—Challenges in regulating quality, safety and efficacy. *J. Ethnopharmacol.* **2008**, *119*, 705–710. [CrossRef] [PubMed]

39. Li, X.Y.; Zhou, G.F. Climate adaptability analysis of *Gynostemma pentaphyllum*. *Chin. Agric. Meteorol.* **1994**, *15*, 18–19. (In Chinese)

40. Pang, M. Study on Germplasm Resources of Medicinal Plant *Gynostemma pentaphyllum*. Master's Thesis, Shaanxi Normal University, Xi'an, China, 2006. (In Chinese)

41. Zhang, X. Study on Population Genetics of the Genus *Gynostemma pentaphyllum*. Master's Thesis, Shaanxi Normal University, Xi'an, China, 2015. (In Chinese)