Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268 e-ISSN: 2395-3454	Vol. 18	No. 2	pp. 359-368	2019

Original Research Paper

Open Access

Association Analysis Between the Site Index Model and the Site Factors of *Cunninghamia lanceolata* Timber Forest in Western Zhejiang Province

Chen Dong*(**)† and Luming Fang*(**)
*College of Information Engineering of Zhejiang, A&F University, Hangzhou 311300, China
**Key Laboratory of Forestry Intelligent Monitoring and Information Technology in Zhejiang Provincial, Hangzhou 311300, China
†Corresponding author: Chen Dong

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 09-08-2018
Accepted: 09-12-2018

Key Words:
FP-growth algorithm
Association rules
Site factors
Stand growth factors
Site index model

ABSTRACT

Forest site conditions affect the tree growth. In this study, a site index model based on Richard's theoretical equation was established using the dynamic monitoring data of forest resources in Lin'an District in Zhejiang Province, Chinese fir [*Cunninghamia lanceolata* (Lamb.) Hook.] as the study object, and a difference equation to determine the internal relationship between site factors and forest stand growth. The FP-growth algorithm was used to explore the association rules between site and forest stand growth factors. A total of 175 factor association rules were obtained with the confidence coefficient and support degree of more than 80% and 10%, respectively. The main factors that influence the site quality of Chinese fir timber forest in Lin'an District are altitude, slope position, slope direction, slope gradient, soil type, soil texture, soil layer thickness, humus layer thickness, undergrowth vegetation species, and undergrowth vegetation coverage. The rules revealed the hidden associations between site factors and site quality and between site and forest stand growth factors in the Chinese fir forest. The results can provide a theoretical foundation for follow-up site evaluation work and are of great practical significance to scientific afforestation and forest management.

INTRODUCTION

In forest management, site quality, which is an important index for measuring forest growing environment and vegetation production potential on one site, is of great importance to the growth and harvest of forest stands. Forest stands with different site conditions have different tree growth states, which affect their final harvests. Site conditions and site factors, such as slope, slope direction and altitude, influence the tree growth. Selecting correct site factors and finding their influencing laws on forest stand growth are of great practical significance to introduction and cultivation of trees, matching of species with the site, selection of planting site, and scientific evaluation of the site qualities of forest stands.

Many studies have reported the association between site factors and forest tree growth, and the research methods used are mainly divided into two types. The first method comprises grouping and classification of forest stand investigation data according to different site factors and independent analysis of differences in forest tree growth under different classifications of each site factor. For instance, Sun & Jiang (2004) grouped *Abies holophylla* heights according to the actual distribution intervals of single site factors and compared growth conditions through variance analysis. Chen (2010) conducted an afforestation test of different site factors by using an orthogonal design and analysed the influence degrees of slope position, slope gradient, and slope direction on diameter at breast height (DBH), height, volume of timber, and biomass of *Phoebe zhennan*. Li et al. (2014) conducted a comparative analysis of the unit area stand volumes of large-scale provincial Chinese fir under four different site factors: landform, slope direction, slope gradient and slope position; results indicated that landform and slope position significantly influence Chinese fir stand volume, whereas slope direction and slope gradient have insignificant influence. The second method explores the association between site and forest stand growth factors by establishing a quantitative model. In this aspect, Xun et al. (1995) established a multiple regression model of five site factors with DBH and tree height increment of *Larix kaempferi*; they found that altitude, coating thickness, slope direction, and slope gradient significantly influence forest stand growth. Wu et al. (2014) utilized principal component analysis to determine the influences of five site conditions, namely, slope direction, slope position, slope gradient, soil bulk density, and soil organic content, on the growth of *Camellia oleifera* Abel. However, research on the association between site and forest tree growth factors only analysed statistical associations between site and growth factors but did not quantify their importance degrees or elucidate the relationship among site factors. Meanwhile, site factors that influence forest tree growth are usually selected by experience, while some potential association factors are ignored. Thus, research results are quite subjective. Few works have conducted profound analysis of multisource mass data on forest resources through a simple statistical method and experience. Therefore, scholars must develop an effective method for rapid and automatic in-depth analysis of the association between site and forest stand growth and reveal their hidden relationships to conduct later-stage evaluation of forest resource data and provide guidance for forestry development.

Data mining, as an emerging interdisciplinary application field, is a process of extracting hidden and potentially useful information and knowledge that is not known before from the incomplete, noisy, fuzzy, and random mass data (Witten & Frank 1999). Since 1993, domestic research and development institutions and institutions of higher learning have carried out basic theoretical and application studies on data mining in succession (Han & Kamber 2012). In the field of forest resource management and decision making, data mining technology has been applied to forest stand harvesting (Ashraf et al. 2013), forestry remote sensing (Yan et al. 2015), biomass prediction (Vatsavai 2015), and forest fire simulation (Cheng & Wang 2008, Pourtaghi et al. 2016). An important task in data mining is association analysis. Association rule mining refers to finding certain associations or relationships among itemsets from mass data (Wang et al. 2006). Many types of association rule mining algorithms exist, and the most classic and extensively applied are Apriori and FP-growth algorithms. The Apriori algorithm is easy to implement, but the database should be frequently scanned and a large quantity of candidate itemsets is generated during practical application, resulting in high time complexity and low algorithm efficiency; thus, this algorithm is unsuitable for rule mining of big data. The FP-growth algorithm can be used for real-time construction of an undirected itemset graph without needing to frequently scan the database, leading to high mining efficiency.

Chinese fir plays a significant role in timber forests in Zhejiang Province (Zhou 2001). Within a certain range of climatic region, the growth of Chinese fir forest stands is influenced by site environment and their own physiological conditions. In this study, the Chinese fir plantation in Western Zhejiang was taken as the study object, and the FP-growth algorithm was used for association analysis between the site factors of the Chinese fir plantation and its growth indices. Results can be used to mine the association rules between site quality and site factors and between site and forest stand growth factors. This study can provide a scientific basis for local site quality evaluation and site model establishment to promote afforestation, greening, and matching of species with the site. Moreover, sufficiently determining soil fertility and tree production potential and facilitating high-yield, high-quality, and high-efficiency development of the forestry are of great practical significance.

MATERIALS AND METHODS

Profile of the data sources: Lin’an District is in the west of Hangzhou City in Zhejiang Province, China from east longitude 118°51’ to 119°52’ and from north latitude 29°56’ to 30°23’. Data were derived from the dynamic monitoring data of forest resources in Lin’an District from 2008 to 2012, which was the annual dynamic monitoring system of counter-level forest resources established based on the planning and design investigation of forest resources in Zhejiang Province. The investigation factors, taking a small class as a unit, included basic information like stand plot, site factors, forest stand factors, ownership information, forest management measures, plant diseases and insect pests, and fire disaster information. Site factors included landform, altitude, slope direction, slope position, slope gradient, soil name, soil texture, soil layer thickness, humus layer thickness, undergrowth vegetation species, undergrowth vegetation height, and coverage, while forest stand factors included forest category, origin, tree species, age, stand DBH, average tree height, dominant tree height, canopy density, degree of closeness, unit number of plants, and unit stand volume.

Among the tree species, small-class retesting data of artificial Chinese fir forest each year between 2008 and 2012 were selected, from which two-stage retesting data were selected to establish the site index model, and 2012 data were selected to study the association rules between forest stand growth and site factors. As the seedlings were in the recovery and rooting phase, they truly entered the fast-growing phase five years later. Meanwhile, the related literature indicates that only forest stands with canopy densities exceeding 0.2 could sufficiently embody the forest tree growth status (Li et al. 2008). Thus, small-class data with ages and canopy densities smaller than five years and 0.2, respectively, were excluded in this study. In addition, data integrity and consistency were checked, and abnormal data were excluded by taking three times the standard deviation as the criterion. Through data processing, 521 small-class data were obtained and respectively distributed in 54 villages, including Linglong Village and Jinxi Village in Lin’an District. According to the survey data, small-class site and forest stand growth conditions were obtained as depicted in Tables 1 and 2.

This study aims to mine potential association rules of site factors with site quality and forest stand growth. Besides the existing site factors, an index that can comprehensively measure site quality is required. In the evaluation process of the site quality of one forest stand, site class, site-class index, and site index are the three commonly used evaluation criteria. Compared with site class and site-class index, site index has been widely applied because of its specific mathematical expression and it is not affected by many artificial interference factors during the modelling process (Raulier et al. 2003, Dong et al. 2016). Therefore, site index was selected as the index for determining the site conditions in this study.

According to retesting data about dominant tree height and age in the Chinese fir plantation, the algebraic difference approach (ADA) was used to establish a site index model of Chinese fir, the site indexes of each small class were calculated according to the model, the site indexes were taken as site factors to be combined with 12 other site factors, and the FP-growth algorithm was used to analyse the association rules according to DBH, tree height, and stand volume of the forest stand growth factors.

Establishment of the site index model: ADA is one of the common methods used to establish forest stand site indexes, and its principle is to select a theoretical equation as the basic equation and select one parameter in the equation as the element elimination parameter for the difference element elimination of the equation to obtain a difference equation, including two groups of dependent and independent variables (Guo et al. 2013).

In establishing the site index model, one group represents the dominant tree height and age of the forest stand at the time, while the other group represents the benchmark age and site index, and the concrete expression of difference equation parameters is obtained through equation fitting. The Richards equation is taken as an example, and the application method of the difference equation is explained. The dominant tree height of the forest stand in year t_1 is HT_1, which is substituted into the Richards equation, as shown in Formula 1.

$$HT_1 = a(1-\exp(-ct_1))^b \quad ...(1)$$

As the site index equation is fixed, the parameters of the equation in which the two data are substituted are unchanged. The dominant tree height of the forest stand in year t_2 is selected as HT_2 and substituted into the Richards equation as shown in Formula 2.

$$HT_2 = a(1-\exp(-ct_2))^b \quad ...(2)$$

Parameter a in Formulas 1 and 2 is placed at the left side of the equation. Then, the logarithm of the two sides is taken to obtain the following:

$$\ln\left(HT_1 / a\right) = b\ln(1-\exp(-ct_1)) \quad ...(3)$$

$$\ln\left(HT_2 / a\right) = b\ln(1-\exp(-ct_2)) \quad ...(4)$$

Formula 4 is divided by Formula 3 to eliminate parameter b, and the difference equation is obtained through Formula 5.

$$HT_2 = a(HT_1 / a)^{\frac{\ln(1-\exp(-ct_2))}{\ln(1-\exp(-ct_1))}} \quad ...(5)$$

Model fitting and test: To measure the fitting accuracy of the model scientifically, R-square (R^2) and residual sum of squares (RSS) were used to judge the model fitting accuracy. Model fitting data were independent of the test data, and root-mean-square error (RMSE) and mean absolute error (MAE) were used to test the model applicability. The model fitting and test indexes are respectively expressed in Formulas 6 to 9.

(1) Coefficient of Correlation, R^2:

$$R^2 = 1-\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 / \sum_{i=1}^{n}(y_i-\bar{y})^2 \quad ...(6)$$

(2) Residual Sum of Squares, RSS:

$$RSS = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \quad ...(7)$$

(3) Root Mean Square Error, RMSE:

$$RMSE = \sqrt{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 / (n-p)} \quad ...(8)$$

(4) Mean Absolute Error, MAE:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \quad ...(9)$$

In the formulas, y_i is the actual value of the *i* (th) dependent variable, $\hat{y}_i$ is the predicted value of the *i* (th) dependent variable, $\bar{y}$ is the mean value of the actual values of dependent variables, *n* is the number of samples, and *p* is number of independent variables in the model.

Association rule and FP-growth algorithm: Association rule is a description of association or correlation between things or between relational data concentration items. Support degree and confidence level are the two measurements for the association rule to mine the degree of interest.

Definition 1**:** X and Y are different transactions in the item, where the support degree of rule R_1: X $\cup$ Y is as follows:

$$\text{Support } (R_1) = \frac{count\,(X \cup Y)}{|D|} \quad ...(10)$$

Where, count ($X \cup Y$) is the quantity of items in the X and Y union and |D| is the quantity of all transactions in D. The support degree of rule R_1 is the ratio of all transaction sets containing X and Y simultaneously in the transaction set.

***Definition 2*:** X and Y are different transactions in the item, and the confidence level of rule R_2: X⇒Y is as follows:

$$\text{Confidence } (R_2) = \frac{Support(X \cup Y)}{Support(X)} \quad \text{...(11)}$$

Where, Support($X \cup Y$) is the support degree of the X and Y union and Support(X) is the support degree of transaction X. The confidence level of R_2 is the probability of the appearance of transaction Y on the condition that transaction X is contained.

When the itemset in the transactions satisfies the minimum support degree, it is called the frequent itemset. Association rule mining includes two processes overall, namely, finding all frequent itemsets and finding their generated strong association rules, where the support degree of itemsets satisfies minimum support counts and the confidence level of association rules satisfies the minimum confidence level.

The FP-growth algorithm, a highly efficient algorithm improved based on the classical association rule algorithm (Apriori algorithm), was proposed by a famous scholar, Professor Jiawei Han (Han et al. 2004). This algorithm uses a special prefix tree, called the FP-tree (frequent pattern tree), as the data structure for storing candidate sets that are compressed by a large margin.

During the establishment process of the FP-tree, transactions should be sorted in a descending order of single support degrees. They are sorted in a lexicographic order when their support degrees are the same, and each transaction data is input in succession. Considering the branches added to the input transactions, whether the transaction is located in the common prefix branch of the current FP-tree is checked. If it is in the common prefix branch, then 1 is added to the support count of this node on the branch; otherwise, a new node should be established to be added to the child node of the last node in the prefix branch, 1 is assigned to its support degree, and then it is recursively added to the residual transactions.

The algorithm description is as follows:
FP-growth (Tree, α)
{
(1) If the tree contains a single path P, then
(2) For each combination (recorded as β) of nodes in path P;
(3) Generate the pattern, $\beta \cup \alpha$ minimum support degree of nodes in its support degree support = β;
(4) Else, for ach αi, at tree head {
(5) Generate a pattern β = αi, its support degree is support = αi.support
(6) Construct the conditional pattern base of β and the conditional FP tree of β, namely, Tree β
(7) If Tree $\beta \neq \emptyset$, then;
(8) Call P-growth (Tree β, β)}
}

Data preprocessing: Before data mining work is formally implemented, data cleaning and data conversion of the original data should be carried out. In this study, data cleaning refers to eliminating abnormal data and invalid data, and the precondition for data mining is the normalization processing of data of different dimensions after cleaning. The data in this paper not only have qualitative factors, but also quantitative factors, where the former include multiple property dimensions and the latter are continuous data. For the convenience of association rule data mining work, qualitative factors are classified while quantitatively continuous data are discretized, and the classification results are expressed by English letters and digital subscripts. The concrete data conversion process is as follows:

1. Landforms (low mountain and hill) are identified as A_1 and A_2.
2. The altitude is divided into three groups (10-200 m, 201-400 m, 401-600 m), which are identified as B_1, B_2 and B_3, respectively.
3. The slope positions (upper, middle, lower, valley, whole) are identified as C_1-C_5.
4. The slope directions (east, south, west, north, northeast, southeast, northwest, southwest) are identified as D_1–D_8.
5. The slope gradients (flat, gentle, inclined, steep, abrupt, dangerous) are identified as E_1–E_6.
6. The soil names (red soil, yellow soil) are identified as F_1 and F_2, respectively.
7. The soil textures (sandy soil, loamy soil, clay) are identified as G_1, G_2, and G_3, respectively.
8. The soil layer thickness values (thick, medium, thin) are identified as H_1, H_2, and H_3, respectively.
9. The humus layer thickness values (thick, medium, thin) are identified as I_1, I_2, and I_3, respectively.
10. The undergrowth vegetation species (grass cluster, shrub, bush wood, no vegetation) are identified as J_1-J_4, respectively.

The undergrowth vegetation height is divided into two groups (0-50 cm, 50-85 m), which are identified as K_1-K_2, respectively.

The undergrowth vegetation coverage rate is divided into three groups (0%-30%, 31%-60%, 61%-90%), which are identified as L_1, L_2 and L_3, respectively.

Table 1: Basic site conditions of Chinese fir plantation in Western Zhejiang Province.

Site factors	Related values
Landform	Low mountains and hills
Altitude (m)	10-570 m
Slope position	Upper, middle, lower, valley, whole
Slope direction	East, south, west, north, northeast, southeast, northwest, southwest
Slope gradient	Flat, gentle, inclined, steep, abrupt, dangerous
Soil name	Red soil and yellow soil
Soil texture	Sandy soil, loamy soil, clay
soil layer thickness (cm)	Thick, medium, thin
Humus layer thickness	Thick, medium, thin
Undergrowth vegetation species	Grass cluster, shrub, bush wood, no vegetation
Undergrowth vegetation height	0-85 cm
Undergrowth vegetation coverage	0%-90%

Table 2: Basic forest stand growth conditions of Chinese fir plantation in Western Zhejiang Province.

Stand growth factors	Numbers	Minimum value	Maximum value	Average value	Standard deviation
Age (t)/a	721	5	38	21.4	9.38
DBH/cm	721	5.9	18.8	15.6	4.87
Tree height/m	721	3.6	17.7	9.8	4.09
Unit stand volume/m^3	721	22	164	88.0	49.44

Table 3: Preprocessed forest stand data.

Stand plot number	Age	DBH	Tree height	Landform	Altitude	Slope position
004	P_3	M_2	N_2	A_2	B_1	C_5
011	P_2	M_2	N_1	A_1	B_1	C_1
018	P_3	M_3	N_3	A_1	B_2	C_1
022	P_1	M_1	N_1	A_1	B_1	C_1

For the forest stand growth factors, the most representative DBH, tree height and stand volume are selected as association rule transactions, and these continuous data go through discretization processing. According to the range of collected data, DBH is divided into three groups (5.0-10.0 cm), (10.1-15.0 cm) and (15.1-20.0 cm), which are expressed by M_1, M_2 and M_3, respectively. The tree height is divided into five groups (3.0-6.0 m), (6.1-9.0 m), (9.1-12.0 m), (12.1-15.0 m) and (15.1-18.0 m), which are expressed by N_1-N_5, respectively. The unit stand volume is divided into five groups (20.0-50.0 m^3), (50.1-80.0 m^3), (80.1-110.0 m^3), (110.1-140.0 m^3) and (140.1-170.0m^3), which are expressed by O_1-O_5, respectively.

Forest stand age is also considered. According to the planning and design survey regulations of forest resources in Zhejiang Province, the age group of the Chinese fir plantation in Western Zhejiang is divided and expressed by P1-P5, namely, sapling forest (≤ 10 years), half-mature forest (11-20 years), nearly mature forest (21-25 years), mature forest (25-32 years), and over mature forest (>32 years).

Table 3 shows some preprocessed data according to the classification, and the data in each row can be understood as one transaction in the FP-growth algorithm.

In this study, the FP-growth algorithm is first used to calculate the frequent itemsets and the support degrees. Then, the association rules between itemsets are obtained according to the confidence levels.

RESULTS AND ANALYSIS

Site index model: Two theoretical equations (Richards and Mitscherlich), which are commonly used to fit the dominant high guide curves of forest stands, and three empirical equations (Schumacher, Hyperbola, and Logarithmic hyperbolic) are taken as basic models to construct a difference equation. The difference equations of the same basic equation will vary from free parameters. In terms of a theoretical equation, when progressive parameter *c* is selected as the free parameter, the difference equation is a multi-shape single-asymptote equation. When potential growth parameter *a* is selected as the free parameter, the difference equation is single-shape variable-asymptote equation (Pyo 2017). For two theoretical growth equations, parameter *a* is the maximum potential growth value of forest trees and parameter *c* is the forest tree growth rate, both of which should be reserved, and thus *b* is taken as the free parameter. An empirical equation takes each parameter as a free parameter for analysis. According to the above conditions, algebraic difference conversion is conducted for the two theoretical equations and three empirical equations to respectively obtain eight difference equations with different forms, and their concrete expressions are given in Table 4.

In data fitting, small-class retesting data are selected, former-stage data in the same small class represent *t* and *HT*, and next-stage data represent *T* and *SI*, where *T* is the benchmark age of Chinese fir, which is taken as 20 years, and *SI* is its site index. According to the 7:3 principle, 365 small-class retesting data are selected for the fitting of site index equations, with R^2 and RSS being the fitting indexes. A total of 156 groups of data, which are not participating in the modelling process, are selected for testing, and RMSE and MAE are taken as the testing indexes. The expression, fitting parameters, fitting indexes and testing indexes of the model are given in Table 5.

Table 5 shows the fitting and testing results of the eight difference equations. The fitting effects of the theoretical equations are superior to those of the empirical equations, and R^2 is 0.716 in the Richards equation, which is higher than those in other equations, and its RSS is the minimum among the eight equations, indicating that the deviation between the predicted and measured values is small. In the model test, the RMSE and MAE in the Richards equation are smaller than those in the other models, indicating that the fitted equation is applicable to the tested data. Therefore, the Richards difference equation is selected in this paper to calculate the site indexes of stand plots.

According to the established site index model, the site indexes of Chinese fir are within 8-18 as transactions for association analysis. The site conditions of indexes 16 and 18 are defined as good and expressed by Q_1, the site conditions of indexes 12 and 14 are moderate and expressed by Q_2, and the site conditions of indexes 8 and 10 are poor and expressed by Q_3.

Association rules and analysis: The FP-growth algorithm is used for the association rule analysis of preprocessed data. According to the related literature (Agrawal et al. 1993), the minimum support degree and the minimum confidence level are set as 10% and 80%, respectively. The Matlab2011a software was used for the association analysis between the site and forest stand growth factors, and 175 association rules that conform to the conditions are finally obtained. As this study aims to mine the association rules between factors that influence site quality and forest growth, the association rules with the high support and confidence degrees, which take the site indexes and the forest stand factors as consequent items, are screened and extracted as given in Table 6.

According to the association rules mining results of several high confidence levels and support degrees in Table 6, the rules can be grouped into five classes:

***Rule class 1*:** 96.2% of Chinese fir forest stands that are growing under the site environment with an altitude below 400 m (B_1, B_2), yellow soil (F_2), a large soil layer thickness (H_1), and an average tree height of above 15 m (N_5) have good site quality; 95.7% of forest stands under the site environment with yellow soil (F_2), undergrowth vegetation being shrub (J_2), a low slope position (C_3), and a gentle slope (E_2) have good site quality; 87.8% of forest stands in the site environment with low slope position (C_3), undergrowth vegetation shrub (J_2), large soil layer thickness (H_1), northeast (D_5) and northwest (D_7) slope directions, and a loamy soil texture (G_2) have a good site quality; 86.4% of forest stands (L_2, L_3) under the site environment with the undergrowth vegetation being shrub and grass cluster (J_1, J_2), a vegetation coverage rate of 31%-90%, and a large humus layer thickness (I_1) have good site quality.

***Rule class 2*:** 82.1% of Chinese fir forest stands under an upper slope (C_1) and gentle slope (E_2) and a grass cluster (J_1) or shrub (J_2) as the undergrowth vegetation have moderate site conditions; 85.8% of forest stands under the environment with red soil (F_2), medium soil layer thickness (H_2), medium humus layer thickness (I_2), undergrowth vegetation of shrub, and undergrowth vegetation coverage rate of 31%-60% (L_2) have moderate site quality.

***Rule class 3*:** 97.8% of forest stands under the site environment with 400-600 m altitude (B_3), upper slope (C_1), hilly land (A_2), and small soil layer thickness (H_3) have poor site quality; 98.2% of forest stands under the site environment with a small soil layer thickness (H_3), upper slope (C_1), south slope (D_2), and inclined slope (E_3) have poor site quality; 94.5% of forest stands with no undergrowth vegetation (J_4), a vegetation coverage rate smaller than 30% (L_1), a sandy soil texture (G_1), and a small humus layer thickness (I_3) have poor site quality.

***Rule class 4*:** 89.6% of forest stands growing on lower (C) gentle (E_2) slope with tree heights within (12.1-15.0 m) (N_4) have their DBHs within (15.1-20.0 cm) (M_3); 87.6% of those

Table 4: Basic equations and difference equations after conversion.

Basic equations	Formulas	Free parameter	Difference equations	Difference equations Number
Richards	$HT = a(1-\exp(-ct))^b$	b	$HT_2 = a(HT_1/a)^{\frac{\ln(1-\exp(-ct_2))}{\ln(1-\exp(-ct_1))}}$	1
Mitscherlich	$HT = a(1-b\exp(-ct))$	b	$HT_2 = \exp(ct_1 - ct_2)(HT_1 - a) + a$	2
Schumacher	$HT = a\exp(-b/t)$	a	$HT_2 = HT_1 \exp(b/t_1 - b/t_2)$	3
		b	$HT_2 = a(HT_1/a)^{t_1/t_2}$	4
Hyperbola	$HT = a + b/t$	a	$HT_2 = b(1/t_2 - 1/t_1) + HT_1$	5
		b	$HT_2 = t_1(HT_1 - a)/t_2 + a$	6
Logarithmic	$\lg(HT) = a + b/t$	a	$HT_2 = 10^{b(1/t_2 - 1/t_1) + \lg(HT_1)}$	7
hyperbolic		b	$HT_2 = 10^{t_1(\lg(HT_1) - a)/t_2 + a}$	8

Table 5: Statistics of parameters of site index equations and fitting and testing indexes.

Difference equations number	Parameters values			Fitting indexes		Testing indexes	
	a	b	c	R^2	RSS	RMSE	MAE
1	12.941	-	0.141	0.716	459.646	1.429	1.217
2	13.055	-	0.125	0.713	453.217	1.488	1.307
3	-	6.416	-	0.553	1272.419	1.874	1.497
4	15.888	-	-	0.681	667.168	1.577	1.333
5	-	-61.364	-	0.599	1037.782	1.763	1.433
6	14.899	-	-	0.699	603.352	1.546	1.289
7	-	-2.786	-	0.553	1272.419	1.877	1.499
8	1.201	-	-	0.681	667.168	1.543	1.311

growing on upper (C_1) inclined (E_3) slope have their DBHs within (5.0-10.0 cm) (M_1); 92.3% of forest stands with a thick humus layer (I_1), gentle slope (E_1), and good site quality have their tree heights at supreme grade, within 15.1-18.0 m (N_5); 86.6% of half-mature Chinese fir forest trees located on low mountain (A_1) with a middle slope (C_2) have their DBHs within 10.1-15 cm (M_2), and 83.1% of those located on hill (A_2) with middle (C_2) inclined (E_3) slope have their DBHs within 5-10 cm (M_1); 92.8% of half-mature (P_2) forest stands with good site quality with red (F_2) loamy (G_2) soil have their unit stand volumes within (50-80 m^3) (O_2). Under the same good site conditions, 90.2% of the half-mature forest stands on red loamy soil have their unit stand volumes within (80.1-110 m^3) (O_3).

***Rule class 5*:** 94.9% of the forest stands with a thin humus layer (I_3), medium soil layer thickness (H_2), and DBH within 5-10 cm (M_1) are sapling forests; 89.6% of the forest stands with a medium humus layer thickness (I_2), medium soil layer thickness (H_2), and vegetation coverage rate within 31%-60% are half-mature forests; 92.1% of those with thick humus layer (I_1), a shrub undergrowth vegetation (J_2), a stand volume within 80-140 m^3(O_4,O_5) and a vegetation coverage rate within 61%-90% (L_3) are nearly mature forests; 99.8% of those with a grass cluster undergrowth vegetation (J_1) or no vegetation (J_4) and a vegetation coverage rate below 30% belong to mature forests.

According to rule classes (1)-(3), the site factors that influence the site quality of Chinese fir timber forest in Western Zhejiang are altitude, slope position, slope direction, slope gradient, soil type, soil texture, soil layer thickness, humus layer thickness, undergrowth vegetation species, and undergrowth vegetation coverage. The influences of landform and undergrowth vegetation height on site quality are not significantly embodied. The analysis of the three rule classes indicates the following: i) Site quality declines with the altitude because altitude affects the temperature and humidity of forest stand growth (Guo et al. 2013), and the general law is that as the altitude increases, the temperature decreases progressively, and humidity increases. The differences in altitude and temperature in Lin’an region are large, and the region with a high altitude has low temperature, which is not conducive for the growth of the Chinese fir forest. ii) Slope position, direction, and gradient have a certain bearing on forest stand site quality, and the rule results indicate that the higher the slope position, the steeper the slope and the poorer the forest stand site quality be-

Table 6: Partial association rules by the FP-growth algorithm.

No.	Association rules	Support degrees %	Confidence degrees %
1	$B_1,B_2,F_2,H_1,N_5 \rightarrow Q_1$	25.3	96.2
2	$B_3,C_1,A_2,H_3 \rightarrow Q_3$	20.0	97.8
3	$C_1,E_2,J_1,J_2 \rightarrow Q_2$	15.7	82.1
4	$H_3,C_1,D_2,E_3 \rightarrow Q_3$	22.3	98.2
5	$F_2,J_2,C_3, E_2 \rightarrow Q_1$	21.2	95.7
6	$F_2,H_2,I_2,J_2,L_2 \rightarrow Q_2$	10.0	85.8
7	$C_3,J_{2,} H_1, D_5,D_7,G_2 \rightarrow Q_1$	22.4	87.8
8	$E_2,C_3,N_4 \rightarrow M_3$	31.4	89.6
9	$I_1,E_1,Q_1 \rightarrow N_5$	28.7	92.3
10	$I_1, J_2, L_3,O_4,O_5 \rightarrow P_3$	27.1	92.1
11	$J_4,J_1,L_1 \rightarrow P_5$	18.0	99.8
12	$A_1,C_2,E_1,P_2 \rightarrow M_2$	17.4	86.6
13	$A_2,C_2,E_3,P_2 \rightarrow M_1$	21.1	83.1
14	$Q_1,G_2,F_1,P_3, \rightarrow O_2$	17.7	92.8
15	$Q_1,G_2,F_2,P_3, \rightarrow O_3$	33.1	90.2
16	$I_{3,} H_2,M_1 \rightarrow P_1$	17.7	94.9
17	$I_2, H_2,L_2 \rightarrow P_2$	13.8	89.6
18	$J_1,J_2,L_2,L_3,I_1 \rightarrow Q_1$	17.6	86.4
19	$J_4,L_1,G_1,I_3 \rightarrow Q_3$	30.1	94.5
20	$E_3,C_1 \rightarrow M_1$	22.2	87.6

cause slope gradient and slope position have influences on microclimate in forest stands. A high slope is usually windward, and the steeper the slope, the thinner the soil layer, and according to physiological features of Chinese fir, windward plantation is not good for its growth (Song 2008). The rules also indicate that the Chinese fir trees growing on the northeast and northwest slopes have better quality than those growing on south slope, thus verifying the benefit of growing them on shady and half-shady slopes. iii) Site quality is directly proportional to soil layer thickness and humus layer thickness. The study indicates that the fast-growing period of Chinese fir forest stands with loose and wet soil and large thickness usually has a long duration (Chi 1996). The amount of nutrients, such as nitrogen, phosphorus, and potassium, in soil and humidity will increase with increasing soil layer thickness; that is, the thicker the soil layer is, the lower the soil erosion degree will be (Monserud & Marshall 2001). Humus layer can improve soil structure and fertility, that is, the thicker the humus layer is, the higher the soil fertility and the more sufficient its ability to supply nutrients to aboveground plants will be. Meanwhile, loamy soil structure is more suitable for the growth of the Chinese fir forest than the sandy soil structure, and this law is also embodied in the association rules. iv) Forest stand site quality also has a strong association with undergrowth vegetation. The main undergrowth vegetation in the Chinese fir plantation in Western Zhejiang is grass cluster and shrubs among the undergrowth vegetation factors, and the higher the vegetation coverage, the higher the site quality because the increase of undergrowth vegetation contributes to water and soil conservation and improves soil permeability and nutrient storage ability, thus improving the ability of maintaining soil fertility, and this law has been proved in the research results of Akpo (1997), Caccia & Ballaré (1998), He & Fu (2002).

Rule class (4) is the embodiment of the influence of site factors on forest stand growth. Study results indicate that forest stands with a thick humus layer have high trees with a favourable growth status; low mountain, middle slope and lower slope are more suitable for the DBH growth of Chinese fir than hill, inclined slope and upper slope, respectively, thus indirectly reflecting the influences of humus layer, landform and slope position on forest stand growth environment. Under the same site conditions, more stand volume can be obtained by planting Chinese fir under yellow soil than red soil, and this study result is consistent with the study result of Song (2008) on the growth environment of Chinese fir. Rule (4) is another expression of the first three rules and an embodiment of site quality in forest stand growth.

Rule class (5) indicates that some forest stands also present a series of change laws with stand age and is specifically manifested by two factors like vegetation coverage and humus layer thickness. In the sapling forest phase, the undergrowth vegetation coverage is low with a relatively scarce undergrowth biomass and thin humus layer. The vegetation coverage increases with the stand age. The main vegetation in the half-mature forest phase is shrub, accompanied by increasing withered products and increasing humus layer thickness. Undergrowth vegetation coverage and humus layer thickness reach their maximum values in the nearly mature forest phase. When forest stands gradually become mature, leaf canopies are closed, shrub and herbs gradually start disappearing, and undergrowth vegetation coverage declines, while the productivity of vegetation declines, and this result is consistent with study result of Alaback (1984) on the dynamic evolution laws of undergrowth vegetation.

CONCLUSIONS

Mining potential rules and patterns from mass data are the basic problem in data mining at present. Chinese fir timber forest was taken as the study object, and dynamic monitoring of the data of forest resources in Lin'an of Zhejiang Province was performed. The difference equation method was used to establish a site index model of Chinese fir plantation, on which the FP-Growth algorithm was used to conduct an analysis of association rules between the site and growth factors of Chinese fir forest stands, which reveal the causal relationships between different factors. The study results indicate the following:

1. Among the eight difference equations, the site index model (SI=12.941×(Ht /12.171)ln(1-exp(-0.141×20))/ ln(1-exp(-0.141t))) of the difference equation based on Richards theoretical growth model has the highest performance in predicting fitting indexes and testing indexes, and this is closely related to the fact that Richards equation has a favourable biological significance (Guo et al. 2013).
2. Within a certain range of climatic region, the site conditions of Chinese fir are mainly influenced by altitude, slope position, slope direction, slope gradient, soil type, soil texture, soil layer thickness, humus layer thickness, undergrowth vegetation species and undergrowth vegetation coverage. The rules and analysis indicate that the zone with low altitude, shady slope or half-shady slope and gentle slope should be selected for Chinese fir afforestation. When the altitude is high, the temperate zone should be selected for plantation; and to maintain a superior forest stand site environment, a certain biological diversity and vegetation coverage should be kept for the undergrowth vegetation of forest stands, in addition to artificial fertilization, weeding and other measures for the forest land. Landform is the main factor that influences site quality under general circumstances. However, landform influence is only embodied on low mountains and hills due to the limited study data in this paper; as such, the influences of the middle and high mountains on Chinese fir forest growth cannot be compared.
3. The rules indicate that Chinese fir height growth is greatly influenced by humus layer thickness, while DBH growth is greatly influenced by landform and slope position. Thus, Chinese fir should be planted in the zone with low hill and middle-lower slope positions to cultivate forest trees with medium and high diameters. As for soil selection, yellow soil is a better choice than red soil, and trees should be planted in thick and fertile soil to increase the harvest yield of forest stands.
4. The main undergrowth vegetation is shrub. The vegetation coverage first presents an increasing and then decreasing change law with the age of the forest stands. In China's man-made forest system, simple forest stand structure, high density and undeveloped undergrowth vegetation are common. The association rules indicate that undergrowth vegetation diversity can increase humus layer thickness and improve forest stand site quality to promote forest tree growth. Therefore, the richness and diversity of the undergrowth habitat can be properly improved according to the changed laws of undergrowth vegetation. The development of undergrowth vegetation can be facilitated through proper forest culturing and management measures.

Moreover, the study results reveal a certain auxiliary effect on the site quality model with various site factors as independent variables. The site factors in the past site quality model were usually decided by subjective factors with a small application range. The mutually dependent relations between site factors were analysed and evaluated through the data mining technology to objectively extract factors related to site quality. The model established using the extracted site factors has great scientific application and practicability.

The FP-growth algorithm was used in forest stand site selection and forest stand growth to find and extract objective laws hidden behind data, and the matching of species with the site was realized according to the rules. Association rules can not only be applied to the mining of forest stand growth data but can also be promoted and applied to the studies on multifunctional forest management and forest resources-related decision making, thus providing a reference for the accurate management of forest resources to facilitate the stable and diversified development of the entire man-made forest ecosystem.

ACKNOWLEDGEMENTS

This study was supported by the Research Development Fund Project of Zhejiang Agriculture & Forestry University (No. 2016FR036), Zhejiang Education Department General Project (No. Y201636009) and Zhejiang key research and development program funded projects (No. 2018C02013).

REFERENCES

Agrawal, R., Imielinski, T. and Swami, A.N. 1993. Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216.

Akpo, L.E. 1997. Phenological Interactions between tree and understory herbaceous vegetation of a Sahelian semi-arid Savanna. Plant Ecology, 131(2): 241-248.

Alaback, P.B. 1984. Plant succession following logging in the Sitka Spruce- western hemlock forests of southeast Alaska: Implications for management. USDA Forest Service General Technical Report PNW-United States.

Ashraf, M.I., Zhao, Z., Bourque, P.A., MacLean,D. A. and Meng, F. R. 2013. Integrating biophysical controls in forest growth and yield predictions with artificial intelligence technology. Canadian Journal of Forest Research, 43(12): 1162-1171.

Caccia, F.D. and Ballare, C.L. 1998. Effects of tree cover, understory vegetation, and litter on regenerati. Canadian Journal of Forest Research, 28(5): 683-692.

Chen, Shurong 2010. Effect of different site conditions on the growth of Phoebeboumei. Journal of Forest and Environment, 30(2): 157-160.

Cheng, T. and Wang, J. 2008. Integrated spatiotemporal data mining for forest fire prediction. Transactions in GIS, 12(5): 591-611.

Chi, Jian 1996. Technology for Rapid Growth and High Yield of Chinese Fir. Jin Dun Press, Beijing.

Dong, C., Wu, B., Wang, L., Qi, Y. and Han, Y. 2016. Algorithms of thinning simulation for timber plantation-a case study of Chinese fir. Nature Environment and Pollution Technology, 15(2): 429-436.

Guo, Y., Han, Y., Wu, B. and Liu, Y. 2013. Study on modelling of site quality evaluation and its dynamic update technology for plantation forests. Nature Environment and Pollution Technology, 12(4): 591.

Han, J. and Kamber, M. 2012. Data mining: Concepts and techniques. Data Mining Concepts Models Methods & Algorithms Second Edition, 5(4): 1-18.

Han, J., Pei, J., Yin, Y. and Mao, R.Y. 2004. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data Mining & Knowledge Discovery, 8(1): 53-87.

He, Yiling and Fu, Maoyi 2002. Review of studies on understorey of plantations. Forest Research, 15(6): 727-733.

Li, Mingjun, Du, Mingfeng and Yulifei 2014. Study on the relationship between the stand volume and site actors in Chinese fir forest of middle age in Guizhou Province. Forest Resources Management, 5: 58-63.

Yongning, L., Binlan, Z. and Shuying, Q. 2008. Review of research and application of forest canopy closure and its measuring methods. World Forestry Research, 21(1): 40-46.

Monserud, R.A. and Marshall, J.D. 2001. Time-series analysis of 13C from tree rings. I. Time trends and autocorrelation. Tree Physiology, 21(15): 1087-1102.

Pourtaghi, Z.S., Pourghasemi, H.R., Aretano, R. and Semeraro, T. 2016. Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. Ecological Indicators, 64: 72-84.

Pyo, J. 2017. Developing the site index equation using a generalized algebraic difference approach for *Pinus densiflora* in central region, Korea. Forest Science & Technology, 13(2): 87-91.

Raulier, F., Lambert, M.C., Pothier, D. and Ung, C.H. 2003. Impact of dominant tree dynamics on site index curves. Forest Ecology & Management, 184(1): 65-78.

Song, Jing 2008. Discussion of the relationship between Chinese fir growth and environmental conditions. Western China Technology, 7(11): 54-55.

Sun, Yueying and Jiang, Wenjun 2004. The relationships between the growth of *Abies holophylla* Maxim. plantation and the site factors. Journal of Northeast Forestry University, 32(4): 86-87.

Vatsavai, R.R. 2015. Multitemporal data mining: From biomass monitoring to nuclear proliferation detection. Analysis of Multitemporal Remote Sensing Images. IEEE, pp. 1-4.

Wang, Tian, Peng, Shikui and She, Guanghui 2006. Application of Apriori algorithm on data analyses of forest inventory. Journal of Nanjing Forestry University (Natural Sciences Edition), 30(3): 63-66.

Witten, Frank I. H. 1999. Data mining. Practical Machine Learning Tools & Techniques with Java Implementations, 13(4): 1-1.

Wu, Y., Lang, N., Li, T., Liu, Z. and Zhao, L. 2014. Growth indexes of a new variety of Camellia oleifera in Yunnan and relativity and dominant analysis of its site factors. Journal of West China Forestry Science, 43(2): 79-84.

Xun, Shouhua, Cao, Hanyu and Wu, Dejun 1995. Correlation analysis between the increment and site factors of Japan Larch. Journal of Shandong Normal University (Natural Science), (2): 35-37.

Yan, F., Gong, Y. and Feng, Z. 2015. Combination of artificial neural network with multispectral remote sensing data as applied in site quality evaluation in Inner Mongolia. Croatian Journal of Forest Engineering, 36(2): 307-319.

Zhou, G.M., Guo, R.J., Wei, X.L. and Wang, X.J. 2001. Growth model and cutting age of Chinese fir planted forest in Zhejiang Province. Journal of Zhejiang Agricultural and Forestry University, 18(3): 219-222.