Article

Landslide Susceptibility Mapping Based on Particle Swarm Optimization of Multiple Kernel Relevance Vector Machines: Case of a Low Hill Area in Sichuan Province, China

Yongliang Lin [1,2], Kewen Xia [1,*], Xiaoqing Jiang [1,3], Jianchuan Bai [1,4] and Panpan Wu [1]

[1]   School of Electronics and Information Engineering, Hebei University of Technology, No. 5340 Xiping RD, Beichen District, Tianjin 300401, China; lyl@tcu.edu.cn (Y.L.); ujnjxq@126.com (X.J.); chuanjianbai888@163.com (J.B.); pwu0320@163.com (P.W.)

[2]   Information Center, Tianjin Chengjian University, No. 26 Jinjing RD, Xiqing District, Tianjin 300384, China

[3]   School of Information Science and Engineering, University of Jinan, No. 336 West Road of Nan Xinzhuang, Jinan 250022, China

[4]   Centre of Intelligent and Networked System, Central Queensland University, Bruce Highway, North Rockhampton, Queensland 4701, Australia

*   Correspondence: kwxia@hebut.edu.cn; Tel.: +86-22-6043-5739

Abstract: In this paper, we propose a multiple kernel relevance vector machine (RVM) method based on the adaptive cloud particle swarm optimization (PSO) algorithm to map landslide susceptibility in the low hill area of Sichuan Province, China. In the multi-kernel structure, the kernel selection problem can be solved by adjusting the kernel weight, which determines the single kernel contribution of the final kernel mapping. The weights and parameters of the multi-kernel function were optimized using the PSO algorithm. In addition, the convergence speed of the PSO algorithm was increased using cloud theory. To ensure the stability of the prediction model, the result of a five-fold cross-validation method was used as the fitness of the PSO algorithm. To verify the results, receiver operating characteristic curves (ROC) and landslide dot density (LDD) were used. The results show that the model that used a heterogeneous kernel (a combination of two different kernel functions) had a larger area under the ROC curve (0.7616) and a lower prediction error ratio (0.28%) than did the other types of kernel models employed in this study. In addition, both the sum of two high susceptibility zone LDDs ($6.71/100$ km^2) and the sum of two low susceptibility zone LDDs ($0.82/100$ km^2) demonstrated that the landslide susceptibility map based on the heterogeneous kernel model was closest to the historical landslide distribution. In conclusion, the results obtained in this study can provide very useful information for disaster prevention and land-use planning in the study area.

Keywords: particle swarm optimization; multiple kernel learning; relevance vector machine; landslide susceptibility; geographical information system (GIS)

1. Introduction

Landslide susceptibility assessment always requires the consideration of many non-linear relation environmental factors, such as geomorphological, geological, hydrological and land cover data [1,2]. Determining how to map regional landslide susceptibility from these factors has become a focus of geological research. In recent years, several approaches to mapping landslide susceptibility have been proposed, which can be grouped into two broad groups: qualitative and quantitative [3]. Qualitative methods are somewhat subjective and depend on the judgment of

experts, whereas quantitative methods are relatively objective and based on numerical expressions of the relationships between environmental factors and landslides [4]. The quantitative methods are generally considered to be more scientific and accurate than qualitative methods. With the development of computing and geographic information system (GIS) technologies, quantitative methods have developed rapidly in recent decades; these methods include fuzzy logic, decision trees, logistic regression, artificial neural networks (ANNs) and support vector machines (SVMs) [5–9]. Studies have shown that fuzzy logic can address missing values, but it depends on greater experience. Moreover, the fuzzy logic single factor evaluation matrix is difficult to establish, thus making the model prediction unstable. Decision trees can also handle missing values and have better local analysis capabilities than logistic regression does, which inevitably leads to over-fitting problems. Logistic regression models are capable of good global analyses, but are sensitive to extremes. ANN and SVM have been proven to be more effective than the above methods for landslide susceptibility assessments. However, ANN can easily fall into local minima and has slow convergence speeds. SVM performs better than ANN at generalization, but its kernel needs to satisfy the Mercer conditions. Additionally, it cannot directly estimate the prediction uncertainty. Therefore, better techniques are urgently needed.

The relevance vector machine (RVM) proposed by Tipping [10] is a new Bayes probability model based on SVM, and its kernel does not need to satisfy the Mercer conditions [11]. Moreover, the RVM results include classification results and the probability distribution, which are very suitable for landslide susceptibility assessment [12]. However, RVM has been seldom applied to the study of landslide fields, especially in regional landslide susceptibility analyses [13,14]. Similar to the SVM model, the effect of the RVM model depends on the kernel function and kernel parameters. At present, the methods for choosing an effective kernel function and reasonable kernel parameters are still imperfect [15,16].

The multiple kernel learning method provides a proper solution for the kernel selection problem. It always combines some single kernels under certain rules to offer a suitable kernel mapping according to specific samples [16–19]. In that way, the kernel selection problem is transformed into an optimization problem of kernel parameters and kernel weights in a multiple kernel structure. In recent years, the particle swarm optimization (PSO) algorithm [20–23] has been widely used to optimize the parameters of intelligent algorithms in many fields [24–29]. It is a classic group optimization technique that Kennedy and Eberhart [20,21] designed on the basis of the actions of birds feeding. The optimization target can be searched by tracking individual and group particles of the population. At present, studies of the PSO algorithm can be divided into two groups: improvement research and applied research. Most of the improvement research has concentrated on setting the inertia weight and learning factors of the PSO algorithm [22–25]. The adaptive cloud PSO (CPSO) algorithm is one of the most efficient methods in improvement research. It can both increase the convergence speed and ensure the diversity of a population [24,25,28].

In the present study, RVM models using five kernel types based on the CPSO algorithm were applied to landslide susceptibility mapping of the low hill area of China's Sichuan Province. The five-fold cross-validation method was used in the training processes of the five models to avoid the over-fitting problem [30,31], and cloud theory was used to improve the PSO convergence speed [24,25]. In addition, the prediction performances of the five models were verified using a receiver operating characteristic (ROC) curve [32–34] and landslide dot density (LDD) [35,36].

2. Study Area

The study area (27°9′N–32°4′N and 102°9′E–108°1′E) is located in the eastern part of Sichuan Province, China (Figure 1), and covers approximately 112,000 km^2 (23% of Sichuan Province). It includes Chengdu, Zigong, Luzhou, Deyang, Mianyang, Guangyuan, Suining, Neijiang, Leshan, Nanchong, Meishan, Yibin, Guang'an, Dachuan, Ya'an, Bazhong, Ziyang and Liangshan, for a total of 18 cities or districts (Figure 1). It has a high population density and rapid economic development.

According to statistics from 2009 [37–40], the population density was 437 persons per square kilometre, which was much higher than the provincial average (168 persons per square kilometre). The gross domestic product (GDP) was approximately 13,738.54 billion Yuan (96% of the province). However, its landslide disaster density, which is as large as one per 100 km^2, is the highest in China [41]. Therefore, research on the formation mechanisms of landslides and mapping the regional landslide susceptibility in this area are very useful for disaster prevention and land planning [38–40].

Landslides are induced by many factors that can be divided into two broad groups: geographical environment and human activities. First, the specific geographical environment of the study area provides favourable conditions for landslides: (1) low hills mainly cover the study area (98%), the altitudes of which primarily range from 0 m to 800 m (94%); (2) the topographic relief is large, and the altitude differs over 2000 meters from west to east; (3) the study area is located in the subtropical moist monsoon climate zone, and the annual rainfall is usually over 1000 mm from June to September (from records, 2664 landslides induced by rainfall had occurred in the low hilly area of Sichuan Province [42]); (4) many fractures, rivers (1.6 km/km^2) and sedimentary rocks (89%) are concentrated in this area. Second, human activities are usually the major factor for inducing landslides. There is a well-developed transportation system in the study area, and the total length of the road network (including national and provincial roads) exceeded 211,888 km by 2009 [42].
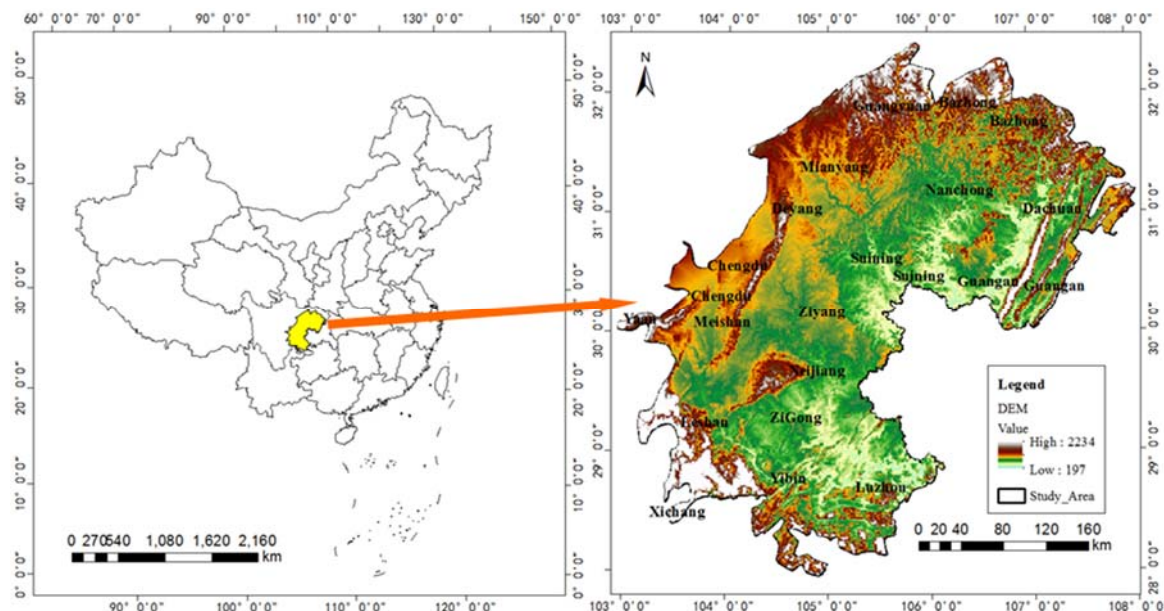

Figure 1. Map of the distribution of hills in Sichuan.

3. Methods

3.1. Relevance Vector Machine

Given a set of nonlinear training samples $\{x_i, t_i\}$ ($i = 1, 2, \cdots, M$), where M is the number of samples, $x_i \in R^n$ are the training influencing factors of landslides and $t_i \in \{0, 1\}$ represents the state of the landslides (the landslide state is 1, and the non-landslide state is 0), assume that the objective function is independent and has additional noise; the output of RVM model can be expressed as Equations (1) and (2):

$$t_i = y(x_i; \omega) + \varepsilon_i \tag{1}$$

$$y(x; \omega) = \sum_{i=1}^{M} \omega_i K(x, x_i) + \omega_0 \tag{2}$$

where $\varepsilon_i = N\left(0, \sigma^2\right)$ is the additive Gaussian noise, x is the input vector that needs to be predicted, $\omega = [\omega_0, \omega_1, \omega_2, \ldots, \omega_M]^T$, ω_i is the model "weight", ω_0 is the deviation and $K(x, x_i)$ is the selected kernel function.

A Bayesian probability model is used to explain the effects of the noise (ε_i) on the predicted results. The model not only can improve the problem of setting the error parameter in SVM, but can also output the probability of the results, which is used to describe the landslide susceptibility. The Bayesian probability is:

$$p\left(t_*|t\right) = \int p\left(t_*\Big|\omega, \sigma^2\right) \; p\left(\omega, \sigma^2\Big|t\right) d\omega d\sigma^2 \tag{3}$$

where t_* is the prediction target of the new input vector x.

To prevent the over-fitting problem when evaluating the maximum likelihood estimation of ω and σ^2, the automatic relevance determination (ARD) prior probability distribution can be defined as

$$p(\omega|\alpha) = \prod_{i=0}^{M} N\left(\omega_i\Big|0, \alpha_i^{-1}\right) \tag{4}$$

The prior distribution of the ω_i is a Gaussian distribution with a mean of 0 and variance α_i^{-1}. Following the Bayesian rule, the conditional probability can be written as:

$$\begin{aligned} p\left(t_*|t\right) &= \int p\left(t_*|\omega, \sigma^2\right) \; p\left(\omega|t, \alpha, \sigma^2\right) p\left(\alpha, \sigma^2|t\right) d\omega d\alpha d\sigma^2 \\ &\approx \int p\left(t_*|\omega, \partial_{MP}, \sigma_{MP}^2\right) \; p\left(\omega|t, \partial_{MP}, \sigma_{MP}^2\right) d\omega \end{aligned} \tag{5}$$

Using the Laplace method to make $p\left(\omega, \alpha, \sigma^2|t_i\right)$ approximate a Gaussian distribution:

$$\partial_i^{new} = \frac{\gamma_i}{\mu_i^2}, \left(\sigma^2\right)^{new} = \frac{\parallel t - \Phi\mu \parallel^2}{M - \sum_{i=0}^{M} \gamma_i} \tag{6}$$

where μ_i is the i-th element of the a posteriori mean vector (μ) of weights, $\sum_{i,i}$ is the i-th diagonal element of the covariance matrix ($\sum$) and $\gamma_i = 1 - \partial_i \sum_{i,i}$. In the training process, Equation (6) continues to iterate until ∂_i^{new} and $\left(\sigma^2\right)^{new}$ are approximately ∂_{MP} and σ_{MP}^2. In that case, most ∂_i tend to infinity, and the corresponding ω_i tend to 0. Finally, a few ω tend to finite values, and their corresponding x_i are the relevant vectors of the RVM model.

The logistic model of the regression method was applied to solve the problem of classification, which is written as:

$$p\left(t_i = 1\Big|\omega^T x_i\right) = 1/\left(1 + exp\left(-\omega^T x_i\right)\right) \tag{7}$$

The probability of the prediction results can then be described as:

$$p\left(t|\omega\right) = \prod_{i=1}^{N} \sigma\left\{y\left(x_i; \omega\right)\right\}^{t_i} \left[1 - \sigma\left\{y\left(x_i; \omega\right)\right\}\right]^{1-t_i} \tag{8}$$

where $t = \{t_n\}_{n=1}^{M}$, and the landslide discrimination criterion is:

$$t_i = \begin{cases} 0\,(non-landslide)\,,\; if\; p_i = 1/\left(1 + e^{-y}\right) < 0.5 \\ 1\,(landslide)\,,\; if\; p_i = 1/\left(1 + e^{-y}\right) \geq 0.5 \end{cases} \tag{9}$$

3.2. Multiple Kernel RVM

The adaptive multi-kernel function of the linear combination is the most classic method [23] and is formed of a combination of an overall kernel function (polynomial) and a local kernel function (Gaussian). It can be written as:

$$K_{mix}(x, x_i) = \beta K_{gauss}(x, x_i) + (1 - \beta) K_{poly}(x, x_i) = \sum_{i=1}^{2} \beta_i K(x, x_i) \tag{10}$$

$$K_{gauss}(x, x_i) = exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right), \sigma > 0 \text{ and} \tag{11}$$

$$K_{poly}(x, x_i) = ((x, x_i) + c)^2, c \geq 0 \tag{12}$$

where $K_{gauss}(x, x_i)$ is the Gaussian kernel function and $K_{poly}(x, x_i)$ is the polynomial kernel function. The kernel weight $\beta\,(\beta \in (0,1))$ is a regulatory factor that is used to adjust the contribution of each kernel. When β is 1, the multi-kernel function is the same as a Gaussian kernel function, and when β is 0, the multi-kernel function equates to a polynomial kernel function. The greater the kernel weight, the greater the contribution of the corresponding kernel to $K_{mix}(x, x_i)$.

After substituting Equation (10) into the original RVM model (Equation (2)), the output of the MKRVM model can be written as

$$y(x; \omega) = \sum_{i=1}^{M} \omega_i K_{mix}(x, x_i) + \omega_0 \tag{13}$$

In this way, the multiple kernel RVM (MKRVM) model transforms the expression problem of the samples in the feature space into the setting problem of the kernel parameters and kernel weights [21]. Three types of multi-kernel combinations are encountered in this paper: the combination of a Gaussian kernel and a polynomial kernel (Gauss and Poly), the combination of two polynomial kernels (Poly and Poly), and the combination of two Gaussian kernels (Gauss and Gauss). A combination of two same kernels is called a homogeneous kernel, and a combination of different kernels is called a heterogeneous kernel.

3.3. Particle Swarm Optimization

The iterative update strategy of original PSO algorithm is:

$$\begin{cases} v_{i+1} = \omega \times v_i + c_1 \times r_1 \times (P_{best}(i) - x_i) + c_2 \times r_2 \times (G_{best}(i) - x_i) \\ x_{i+1} = x_i + v_{i+1} \end{cases} \tag{14}$$

where v_i is the particle velocity, x_i is an individual particle, ω is the inertia weight, $P_{best}(i)$ and $G_{best}(i)$ are the individual and global optimizations in the i-th iteration, respectively, r_1 and r_2 are random numbers (0–1) and c_1 and c_2 are the learning factors.

In this paper, the original PSO was improved using cloud theory (CPSO). In the CPSO algorithm, the particle swarm is divided into three groups to calculate the inertia weight (ω) using different generation strategies. Suppose that the fitness (the prediction error of the RVM models) of the i-th particle in the k-th iteration is f_i^k; the average of all of the fitnesses is $f_{avg}^k = \frac{1}{N}\sum_{i=1}^{N} f_i^k$ (N is the population size); the average of the fitness that is lower than f_{avg}^k is set to f_{avg}'; the average of the fitness that is higher than f_{avg}^k is set to f_{avg}''; and the best fitness is set to f_{best}^k. The specific generation strategy is described as follows.

(1)   When $f_i^k \leq f_{avg}'$, the fitness of these particles is closer to the optimal solution (the lowest error rate). Therefore, set a low inertia weight value ($\omega = 0.2$) to speed up local convergence.

(2)   When $f_i^k < f_{avg}''$ and $f_i^k > f_{avg}'$, these particles are relatively far from the best fitness, which can be improved by the cloud model.
The expectation of the cloud model is $\text{Ex} = f_{best}^k$.
The entropy can be calculated using the distance of the expectation and f_{avg}':
$\text{En} = \left(f_{avg}' - f_{best}^k\right)/c_1$.

In addition, the hyper entropy was set using He $= En/c_2$.

The value of the inertia weight can be described as:

$$\omega = 0.9 - 0.5e^{\frac{-(f_i^k - \mathrm{Ex})^2}{2(\mathrm{En}')^2}}, \ \mathrm{En}' = \mathrm{normrnd}\,(\mathrm{En}, \mathrm{He}) \tag{15}$$

According to "3En" rules, the control parameters c_1 and c_2 were set to 3 and 10 [24]. "normrnd" generates normally distributed data.

(3)　When $f_i^k \geq f_{avg}''$, these particles need a higher inertia weight ($\omega = 0.9$) to improve the global search capability.

3.4. PSO-MKRVM

Many of the PSO-MKRVM model parameters require initialization. For the study area, the RVM training samples were comprised of eight condition attributes (the landslides' predisposing factors) and one decision attribute (the states of the landslides). In the multi-kernel structure, one kernel weight β and two kernel parameters needed to be calculated. The optimization target of the PSO algorithm is to minimize the result of the five-fold cross-validation [33]. The PSO population size was set to 20, and the number of iterations was set to 50. The iterative curves of the classic linear kernel function (Figure 2) show that the optimal solution could be found within 20 iterations. In addition, $c_1 = c_2 = 2$ was chosen to make the search region centre on $P_{best}(i)$ and $G_{best}(i)$. The particle of the PSO algorithm was set as $x = (\beta, width1, \, width2)$.

The training process for the PSO-MKRVM models included the following steps.

Step 1: Initialization. Set some of the parameters of the PSO-MKRVM models, including the population size, iterations and learning factors.

Step 2: Optimization. Based on the training samples, the 5-fold cross-validation method and PSO algorithm are used to optimize the parameters and weights of the multi-kernel function.

Step 3: Building. Depending on the results of step 2, the MKRVM prediction models can be built.

Step 4: Prediction. The models built in Step 3 can be used to make distribution maps of the landslide susceptibility index (DMLSI).
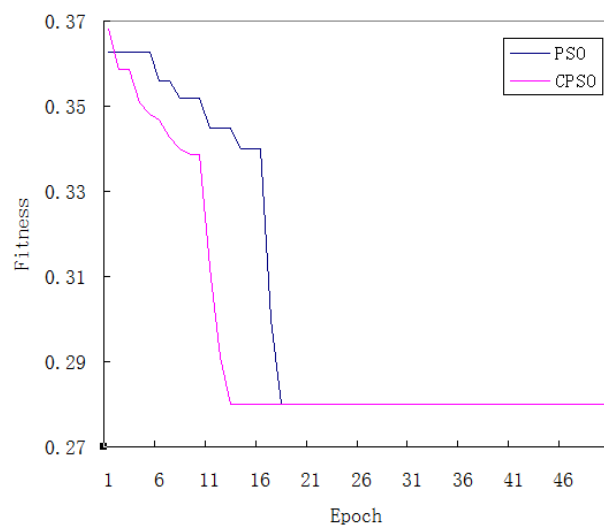

Figure 2. Iterative curves of the classic linear kernel function. CPSO, cloud PSO.

4. Data

4.1. Influencing Factors of Landslides

Overall, 382 landslide samples from the China Institute of Geo-Environmental Monitoring were used for this study; the landslides had occurred in the low hilly area of Sichuan in 2007. Researchers have performed some studies on the susceptibility of landslide predisposing factors based on variable dimension fractal theory and a certainty factor probability model in this area [39,40]. Eight landslide-predisposing factors have been demonstrated to be the major factors responsible for inducing landslides in this area. Based on the landslide inventories and thematic data available in the study area, we used those factors to build landslide prediction models. The factors include landforms (altitude, slope and relief amplitude), geological lithology (lithology), geological structure (faults), cutting slope (river and road) and vegetation (normalized difference vegetation index (NDVI)). The values of the factors are listed in Table 1.

Table 1. Evaluation factor system for landslides in the Sichuan Province hills.

Groups	Factors	Subclasses	Area (%)	Factors	Subclasses	Area (%)
Landform	Slope	0–3°	20.57	Altitude	0–400 m	37.2
		3–5°	17.09		400–600 m	45.84
		5–15°	43.01		600–800 m	11.31
		15–25°	14.42		800–1000 m	3.41
		25–30°	2.55		1000–1200 m	1.18
		30–45°	2.13		>1200 m	1.06
		>45°	0.24			
Geological structure	Faults (buffer distance)	0–2 km	5.83	Faults (buffer distance)	10–12 km	8.84
		2–4 km	14.36		12–14 km	7.54
		4–6 km	13.68		14–16 km	6.26
		6–8 km	12.09		>16 km	5.04
		8–10 km	10.23			
Cutting slope	River network (buffer distance)	0–2 km	10.14	Road network (buffer distance)	0–2 km	13.6
		2–4 km	33.49		2–4 km	11.65
		4–6 km	22.16		4–6 km	10.43
		6–8 km	16.01		6–8 km	9.16
		8–10 km	11.17		8–10 km	8.38
		>10 km	7.02		10–12 km	7.48
	Relief amplitude	0–200 m	71.81		12–14 km	6.36
		200–600 m	25.96		>14 km	32.93
		>600 m	2.23			
Geological lithology	Lithology	Unconsolidated deposits	9.96	Lithology	Conglomerates	37.7
		Mudstone	26.78		Dolomite Carbonate rocks	
		Limestone	3.35		Granite	0.54
		Basalt	0.84			
		Sandstone	20.83		Shale	
Vegetation	NDVI	−1–1				

Based on the above evaluation factor system (Table 1), eight maps of landslide predisposing factors (Figure 3) can be displayed using GIS technologies. Slope, altitude and relief amplitude were derived from the DEM. The slope (Figure 3A) cannot accurately reflect the landforms over equal intervals in the study area. Based on surface features, the slope was divided into 7 intervals. Most areas are in the range of the top four levels (0°–25°, 95.09%). However, many landslides are concentrated in a few areas (25°–45°, 4.68%). In addition, the landslides that occurred above 45° are typically collapsed [5]. The altitude (Figure 3B) was divided into six parts. Area (>800 m) accounts for only 5.65% of the study area, but it gathers a large number of landslides. The relief amplitude (Figure 3C) was derived from the GIS neighbourhood statistics module when the best statistics window radius was 1.1 km. The study area is mainly composed of hills (0–200 m, 71.81%) and low-mountain areas (200–600 m, 25.96%), where landslides have often occurred. Vegetation coverage was extracted

from MODIS products and described by NDVI (Figure 3D). The NDVI values ranged from −1–1; the value increased with an increasing vegetation coverage rate and vice versa. Rivers and roads directly influence the regional geological environments of landslides [2]. The buffer distance maps of rivers (Figure 3E) and roads (Figure 3F) show that landslides primarily occur in the lower buffer distance area. In addition, geological lithology and geological structure are important landslide factors and were extracted from geological maps. The lithology map (Figure 3G) was produced using the ArcGIS vector fusion (Dissolve) method. Limestone, basalt and carbonate rocks represent a small part of the lithology (4.45%), but a number of landslides (2.81, 3.2 and 3.9) occurred in them. Mudstone, sandstone and conglomerates make up 89% of the lithology, but only 1.7, 1.2 and 0.8 landslides per 100 square kilometres occurred in them. In this paper, the impact of faults on landslides was quantified by the buffer distance from the faults (Figure 3H). The fractured structure of the study area presents a spatial "chessboard form", where the lengths of the fracture zones vary from one kilometre to tens of kilometres.
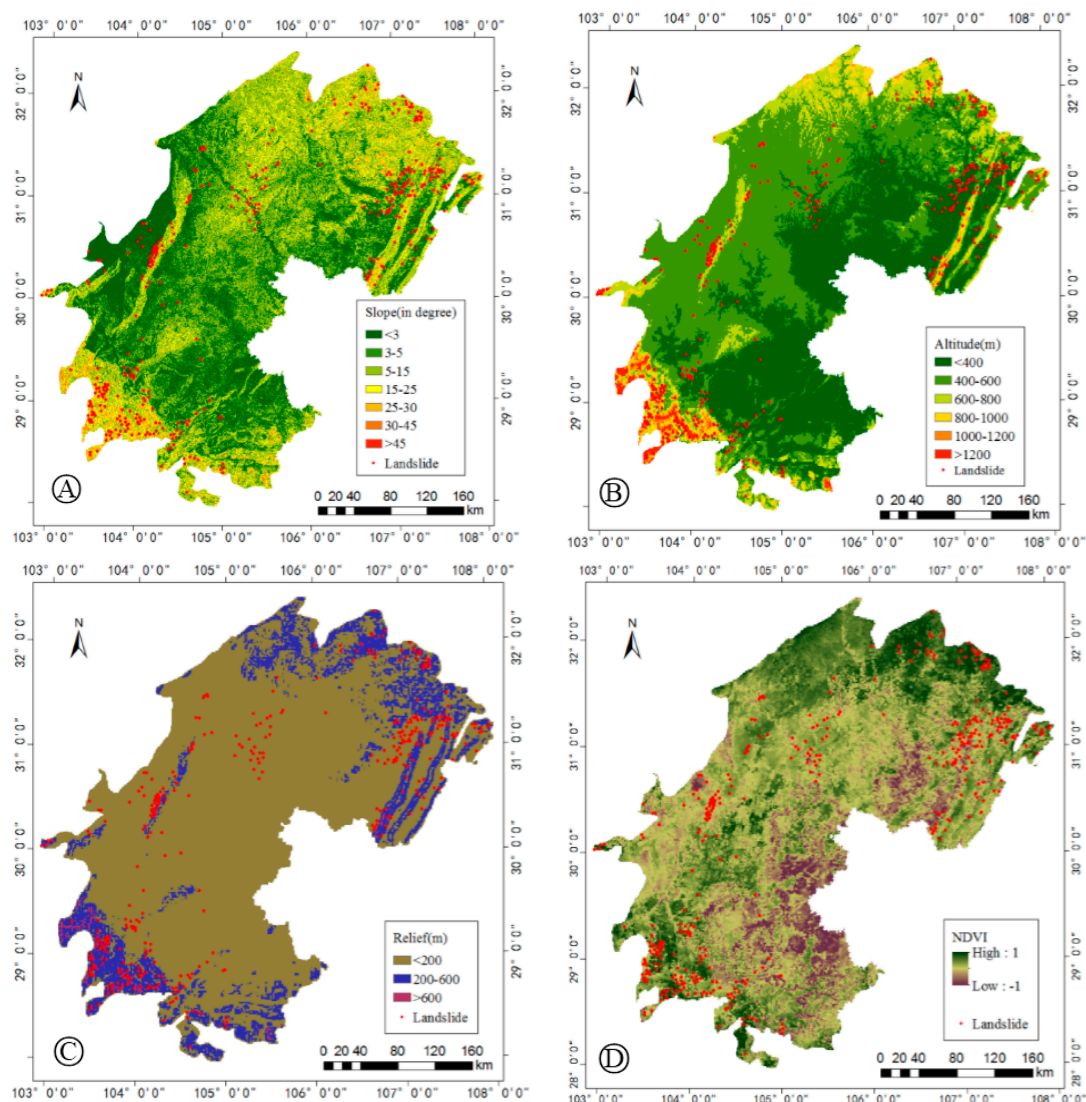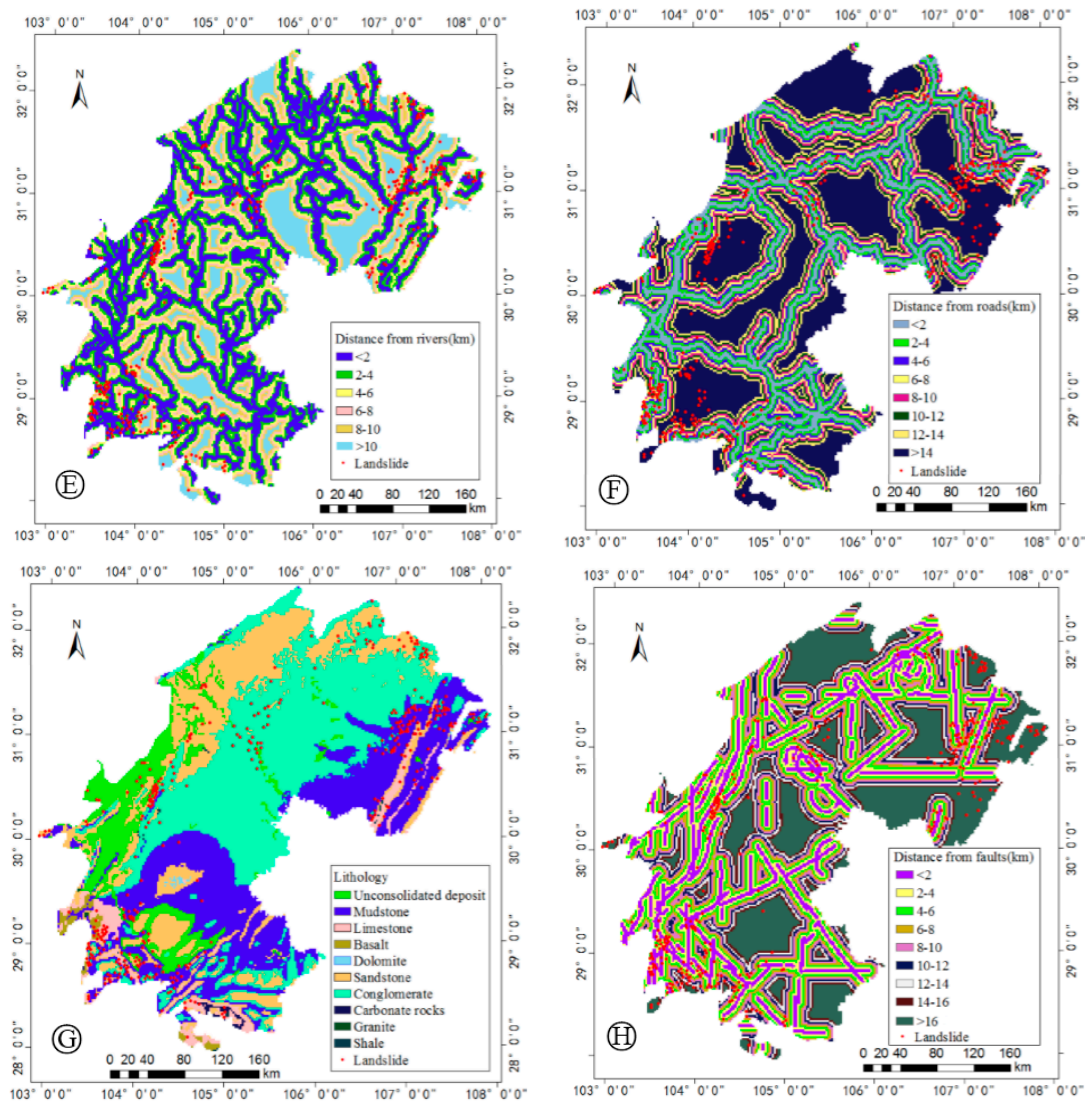

Figure 3. *Cont.*

Figure 3. Map of eight landslide predisposing factors. (**A**) Slope; (**B**) altitude; (**C**) relief; (**D**) NDVI; (**E**) rivers; (**F**) roads; (**G**) lithology; (**H**) faults.

4.2. Normalization Processing

Based on historical landslides and the distribution maps of the eight landslide predisposing factors, 750 training samples (382 landslide samples and 368 non-landslide samples) from the prediction model were extracted using the Arc Toolbox Extraction tool (Figure 4). Similarly, the prediction model inputs of the study area were extracted based on the 27,998 raster units in the study area (the widths and heights of the raster units were 2000 m).

These training samples and regional inputs had to be normalized into unified non-dimensional data using Equation (16):

$$x' = (x - x_{min}) / (x_{max} - x_{min}) \tag{16}$$

where x ($x \in [x_{min}, x_{max}]$) represents the actual value of one factor, x_{min} and x_{max} are, respectively, the minimum and maximum of that factor and $x' \in (0 - 1)$ is the normalization result. The landslides were divided into two states: 1 (landslide) and 0 (non-landslide). To ensure the reliability and stability of the prediction models, a 5-fold cross-validation method was used to divide the 750 training samples into five equal parts. In the training process, four parts were used as training samples, and the other part was used as testing samples. Each part of the samples was used in the training role and

testing role five times. The result of the 5-fold cross-validation method was used as the fitness of the PSO algorithm.
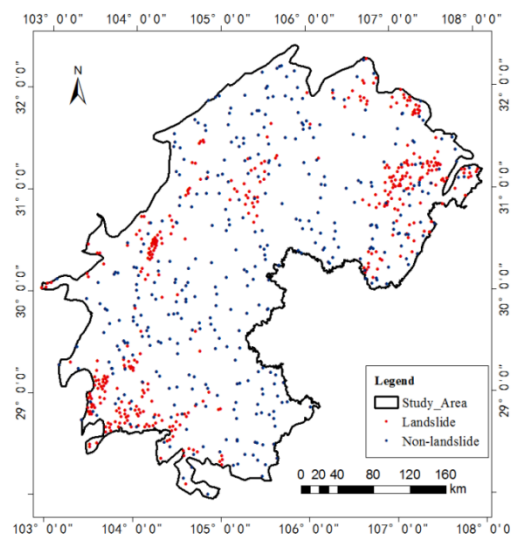

Figure 4. Distribution map of landslides and non-landslides.

5. Results and Discussion

5.1. Model Training

Based on the 750 training samples and the 5-fold cross-validation method, the RVM model parameters were optimized using the PSO algorithm and are shown in Table 2.

Table 2. Kernel parameters and weight coefficient optimized using the PSO algorithm. Poly, polynomial.

Model (Kernel Type)	Weight	Width 1	Width 2	Error Rate (ER)
Gauss		1		0.313
Poly		0.8443		0.333
Gauss and Poly	0.6831	0.6138	2	0.28
Gauss and Gauss	0.2366	0.3396	1	0.333
Poly and Poly	0.114	0.6	0.8844	0.333

The "weight" in the multi-kernel structure represents the weight of the first kernel function. "Width 1" and "Width 2" are the values of the first kernel and second kernel parameters, respectively. The search ranges of the two kernel parameters were [0.3, 1.5] and [0.6, 2.5]. The "error rate" (ER) is the result of the 5-fold cross-validation. Table 2 shows that the "Gauss and Poly" kernel demonstrated the best prediction performance (ER = 0.28) when the three parameters were 0.6821, 0.6138 and 2.

These models were used to make maps of the distribution of the landslide susceptibility index (DMLSI) in the study area. Based on the DMLSI, a cluster analysis of the natural discontinuous points method was used to divide the study area into five landslide susceptibility zones: a very low susceptibility zone (VLS-zone), low susceptibility zone (LS-zone), moderate susceptibility zone (MS-zone), high susceptibility zone (HS-zone) and very high susceptibility zone (VHS-zone). The landslide susceptibility maps (LSM) of the study area developed using this method are presented in Figure 5.
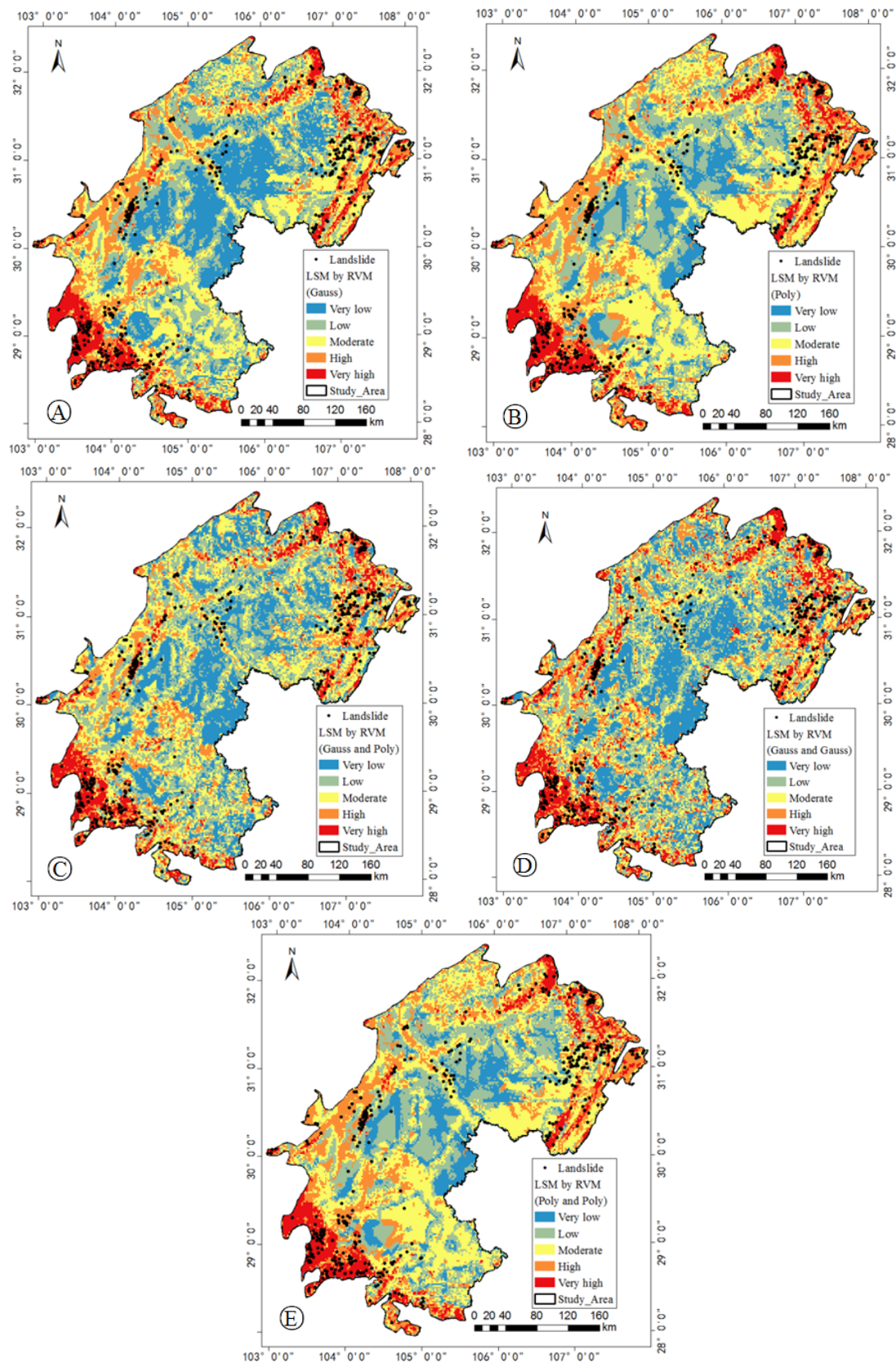
Figure 5. Landslide susceptibility maps of the hilly area in Sichuan. (**A**) Landslide susceptibility map (LSM) by Gaussian; (**B**) LSM by polynomial; (**C**) LSM by Gaussian and polynomial; (**D**) LSM by Gaussian and Gaussian; (**E**) LSM by polynomial and polynomial. RVM, relevance vector machine.

It is obvious that the LSM of the five models conforms well to the distribution of historical landslides. In addition, the models whose kernel function contained a Gaussian function performed better than those employing a polynomial function. This result means that the Gaussian function is more suitable for landslide susceptibility assessment in the study area. However, it is difficult to accurately evaluate the model prediction performances by using LSM.

5.2. Receiver Operating Characteristic Curve

ROC curves are a representative method for testing the accuracies of analysis methods and can reflect the relationship between specificity and sensitivity [32–34]. They were initially applied in the assessment of the receptivity of radar signals and have recently been used to evaluate the prediction performance of medical diagnostic tests [32]. The RIC index that is used to discriminate the accuracy of a "diagnosis" is the area under the ROC curve (AUC) [33]. When the AUC value ranges from 0.5–1, the current method has a good "diagnosis" ability [9]. The diagnostic accuracy can be divided into three levels: low (0.5–0.7), medium (0.7–0.9) and high (>0.9) [33,34].

In this paper, the landslide states were used as the state variable of the ROC curve, and the prediction results of the RVM training models were used as the test variable of the ROC curve. The ROC curves for the different kernel type analyses are shown in Figure 6. The curves indicate that the heterogeneous kernels achieved higher prediction performances (AUC = 0.7616) than did the other kernel types. In addition, the three multi-kernel types that contained a Gaussian function had stronger diagnostic abilities (AUC > 0.7) than did the other types (AUC < 0.7). The strong ability of Gaussian kernel functions for assessing landslide susceptibility in the study area has again been demonstrated.
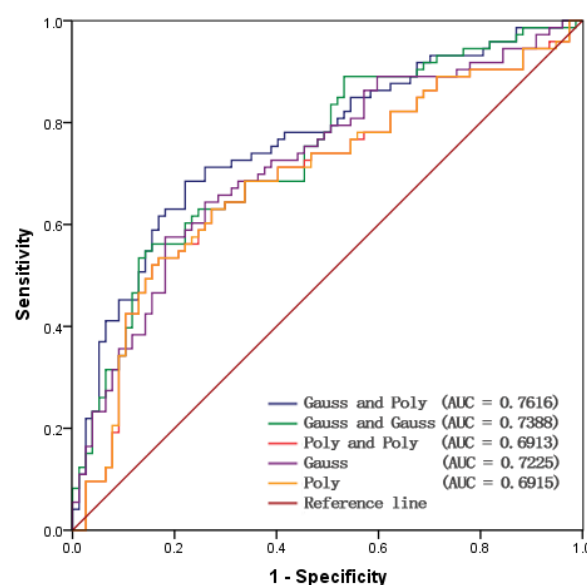

Figure 6. ROC curves for the different kernel type analyses.

5.3. Landslide Dot Density

The frequency ratios of the landslides in each zone of the five LSMs (Figure 5) were used to draw a landslide frequency ratio plot (Figure 7)). The plot results shows that the five frequency ratio lines largely conformed to the actual landslide distribution, in that the frequency ratios continued to grow from the VLS-zone to the VHS-zone. The figure also shows that the "Gauss and Gauss" kernel performed better between the HS-zone and VHS-zone than the other kernel types, but did not perform best in the two low susceptibility zones. In addition, the "Poly and Poly" and "Poly" kernels were almost the same in the five susceptibility zones.
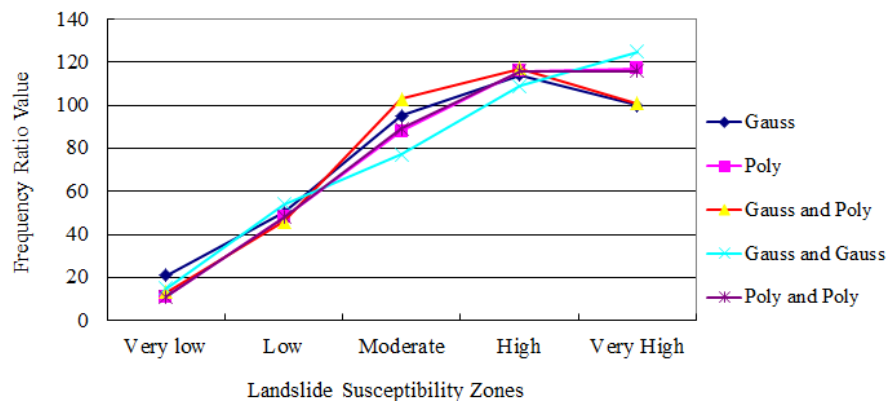
Figure 7. Frequency ratio plots for the five landslide susceptibility zones.

Frequency plots can directly reflect landslides' distributions, but they cannot describe the superiority of a model. Landslide dot density (LDD) is a useful method for making detailed and reasonable analyses of landslide prediction results [35,36]. The LDD results (Table 3) show that the LDD of the polynomial kernel was highest in the VHS-zone (4.47), but not the lowest in the VLS-zone (0.33). The "Gauss and Poly" and "Gauss and Gauss" kernels had identical LDD values (0.24), which were the lowest in the VLS-zone, but did not have the highest LDD values in the VHS-zone. In particular, the "Gauss and Gauss" kernel had the lowest value (3.94) in the VHS-zone. It is difficult to evaluate the effect of each model in a single landslide susceptibility zone. Therefore, we used the summed value of the LDD in the two high susceptibility zones or the two low susceptibility zones to verify the prediction performance. The results show that the "Gauss and Poly" kernel performed best, in that the LDD value was highest in the two high susceptibility zones (6.71) and lowest in the two low susceptibility zones (0.82).

Table 3. The landslide dot densities of the five landslide susceptibility zones. LDD, landslide dot density.

Landslide Susceptibility Zone	LDD (/100 km^2)				
	Gauss	Poly	Gauss and Poly	Poly and Poly	Gauss and Gauss
Very low	0.36	0.33	0.24	0.33	0.24
Low	0.71	0.65	0.58	0.65	0.71
Moderate	1.24	0.98	1.36	0.99	1.22
High	2.3	2.03	2.52	2.04	2.35
Very high	4.1	4.47	4.19	4.42	3.94
Sum of low and very low	1.7	0.98	0.82	0.98	0.95
Sum of high and very high	6.4	6.5	6.71	6.46	6.29

6. Conclusions

RVM is a new Bayesian probability model that performs better than SVM when estimating prediction uncertainties. However, it is seldom used in landslide susceptibility assessments. The effectiveness of RVM depends primarily on the selection of the kernel and kernel parameters. This paper proposed a multi-kernel learning method based on the linear combination of a Gaussian function and a polynomial function to solve the kernel selection problem. In the multi-kernel structure, three parameters were calculated using the adaptive cloud PSO algorithm. To verify the prediction performance of the multi-kernel method, we applied the RVM models in the low hill area of China's Sichuan Province.

Five RVM models used for comparison tests were presented in this paper. The results show that the heterogeneous kernel performed better (ER = 0.28; AUC = 0.7616; the sum of the two HS-zone LDDs was 6.71, and the sum of the two LS-zone LDDs was 0.82) than the other kernel types. This result

occurred mainly because the multi-kernel structure can adaptively adjust its kernel parameters to fit the specific study area. The classic linear kernel combination can not only maintain the interpolation ability of a local kernel function (Gaussian function), but also integrate the generalization ability of a global kernel function (polynomial function). As a final conclusion, the proposed method has a superior prediction skill and higher reliability for regional landslide susceptibility mapping.

Author Contributions: Kewen Xia, Yongliang Lin and Xiaoqing Jiang conceived of and designed the study. Yongliang Lin, Jianchuan Bai and Panpan Wu collected and processed the data. Jianchuan Bai and Panpan Wu analysed the data. Jianchuan Bai, Panpan Wu and Xiaoqing Jiang performed a draft manuscript review, and Kewen Xia polished the paper's language. Yongliang Lin wrote the present manuscript.

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

References

1. Powell, G. Landslide risk management concepts and guidelines. *Aust. Geomech.* **2000**, *35*, 49–92.
2. Chen, X.L.; Ran, H.L.; Qi, S.W. Triggering factors susceptibility of earthquake-induced landslides in 1976 Longling earthquake. *Acta Sci. Nat. Univ. Pekinensis* **2009**, *45*, 104–110. (In Chinese)
3. Pradhan, B. Manifestation of an advanced fuzzy logic model coupled with Geo-information techniques to landslide susceptibility mapping and their comparison with logistic regression modeling. *Environ. Ecol. Stat.* **2011**, *18*, 471–493. [CrossRef]
4. Wu, X.L.; Niu, R.Q.; Ren, F.; Peng, L. Landslide susceptibility mapping using rough sets and back-propagation neural networks in the Three Gorges, China. *Environ. Earth Sci.* **2013**, *70*, 1307–1318. [CrossRef]
5. Ling, P.; Niu, R.Q.; Huang, B.; Wu, X.L.; Zhao, Y.N.; Ye, R.Q. Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. *Geomorphology* **2014**, *204*, 287–301.
6. Melchiorre, C.; Matteucci, M.; Azzoni, A.; Zanchi, A. Artificial neural networks and cluster analysis in landslide susceptibility zonation. *Geomorphology* **2008**, *94*, 379–400. [CrossRef]
7. Wu, X.L.; Ren, F.; Niu, R.Q. Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. *Environ. Earth Sci.* **2014**, *71*, 4725–4738. [CrossRef]
8. Isik, Y. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: Conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ. Earth Sci.* **2010**, *61*, 832–836.
9. Pourghasemi, H.R.; Jirandeh, A.G.; Pradhan, B.; Chong, X.U.; Gokceoglu, C. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. *J. Earth Syst. Sci.* **2013**, *122*, 349–369. [CrossRef]
10. Tipping, M.E. Sparse bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
11. Vapnik, V.N. *The Nature of Statistical Learning Theory*, 2nd ed.; Springer: New York, NY, USA, 2000.
12. Faul, A.C.; Tipping, M.E. Analysis of sparse bayesian learning. In Proceedings of the Advances in Neural Information Processing Systems 14, Vancouver, BC, Canada, 3–8 December 2001.
13. Liu, Z.B.; Shao, J.F.; Xu, W.Y. Comparison on landslide nonlinear displacement analysis and prediction with computational intelligence approaches. *Landslides* **2014**, *11*, 889–896. [CrossRef]
14. Lin, Y.L.; Wang, Z.H.; Xia, K.W.; Li, Z.G. Regional landslide susceptibility assessment based on relevance vector machine. *J. Inf. Comput. Sci.* **2015**, *12*, 6893–6903. [CrossRef]
15. Close, R.; Wilson, J.; Gader, P. A bayesian approach to localized multi-kernel learning using the relevance vector machine. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Vancouver, BC, Canada, 24–29 July 2011.
16. Gönen, M.; Ethem, A. Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **2011**, *12*, 2211–2268.
17. Mehmet, G.; Ethem, A. Localized algorithms for multiple kernel learning. *Pattern Recognit.* **2013**, *46*, 798–807.

18. Wang, H.Q.; Sun, F.C.; Cai, Y.N. On multiple kernel learning methods. *Acta Autom. Sin.* **2010**, *36*, 1037–1050. [CrossRef]

19. Li, D.X.; Wang, J.; Zhao, X.Q.; Liu, Y.; Wang, D.W. Multiple kernel-based multi-instance learning algorithms for image classification. *J. Vis. Commun. Image Represent.* **2014**, *25*, 1112–1117. [CrossRef]

20. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Washington, DC, USA, 27 November–1 December 1995.

21. Shi, Y.; Eberhart, R.C. A modified particle swarm optimizer. In Proceedings of the IEEE International Conference on Evolutionary Computation, Anchorage, AK, USA, 4–9 May 1998.

22. Liang, X.; Li, W.; Zhang, Y.; Zhou, M. An adaptive particle swarm optimization method based on clustering. *Soft Comput.* **2015**, *19*, 431–448. [CrossRef]

23. Li, J.; Zhang, J.; Jiang, C.; Zhou, M.C. Composite particle swarm optimizer with historical memory for function optimization. *IEEE Trans. Cybern.* **2015**, *45*, 2350–2363. [CrossRef] [PubMed]

24. Li, G.D.; Hu, J.P.; Xia, K.W. Intrusion detection using relevance vector machine based on cloud particle swarm optimization. *Control Decis.* **2015**, *30*, 698–702.

25. Liu, Q.F.; Bo, H.L.; Qin, B.K. Optimization of direct action solenoid valve based on Cloud PSO. *Ann. Nucl. Energy* **2013**, *53*, 299–308. [CrossRef]

26. Duan, Q.; Zhao, J.G.; Ma, Y. Relevance vector machine based on particle swarm optimization of compounding kernels in electricity load forecasting. *Electr. Mach. Control* **2010**, *14*, 33–38. (In Chinese)

27. Fei, S.W.; He, Y.; Ma, X.J.; Miao, Y.B. A hybrid model of RVM and PSO for dissolved gases content forecasting in transformer oil. *Recent Pat. Electr. Electr. Eng.* **2013**, *6*, 183–189. [CrossRef]

28. Fei, S.W.; He, Y. A multiple-kernel relevance vector machine with nonlinear decreasing inertia weight PSO for state prediction of bearing. *Shock Vib.* **2015**, *2015*, 1–6. [CrossRef]

29. Zhang, C.L.; He, Y.G.; Yuan, L.F.; Deng, F.M. A novel approach for analog circuit fault prognostics based on improved RVM. *J. Electr. Test. Theory Appl.* **2014**, *30*, 343–356. [CrossRef]

30. Antoine, S.; Isabel, M.; Van Wesemael, B. Soil organic carbon predictions by airborne imaging spectroscopy: Comparing cross-validation and validation. *Soil Sci. Soc. Am. J.* **2012**, *76*, 2174–2183.

31. Fushik, T. Estimation of prediction error by using K-fold cross-validation. *Stat. Comput.* **2011**, *21*, 137–146. [CrossRef]

32. Milos, M. Comparing the performance of different landslide susceptibility models in ROC space. In Proceedings of the Landslide Science and Practice: Landslide Inventory and Susceptibility and Hazard Zoning, Rome, Italy, 3–9 October 2011.

33. Bradley, A.P. Use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [CrossRef]

34. John, M.; Jha, V.K.; Rawat, G.S. Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. *Landslides* **2009**, *6*, 17–26.

35. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [CrossRef]

36. Fratinni, P.; Crosta, G.; Carrara, A. Techniques for evaluating the performance of landslide susceptibility models. *Eng. Geol.* **2010**, *111*, 62–72. [CrossRef]

37. Yu, S.H. Analyses on spatial-temporal characteristics of mud-rock flow and landslide in Sichuan basin and its meteorological cause. *Plateau Meteorol.* **2003**, *22*, 83–89. (In Chinese)

38. Tang, X.C.; Xie, S.Y. The exploration on the causes of geotectonic to form the regularity of distribution of the mountain calamity landforms surrounding Sichuan Basin. *J. Soil Water Conserv.* **1994**, *8*, 76–84. (In Chinese)

39. Wang, Z.H.; Hu, Z.W.; Liu, C.Q. Susceptibility analysis of disaster-pregnant environmental factors of landslide in the hilly area in Sichuan based on variable dimension fractal theory. *Earth Environ.* **2013**, *41*, 680–687. (In Chinese)

40. Wang, Z.H.; Hu, Z.W.; Zhao, W.J.; Gong, H.L.; Deng, J.X. Susceptibility analysis of precipitation-induced landslide disaster-pregnant environmental factors based on the certainty factor probability model—Taking the hilly area in Sichuan as example. *J. Catastrophol.* **2014**, *29*, 109–115. (In Chinese)

41. Zhang, S.Y. *Basis of Geological Disaster Weather Forecast*; China Meteorological Press: Beijing, China, 2009. (In Chinese)
42. Huang, Q. On prevention of landslide in low mountains and hills country in Jiangxi province. *J. Jiangxi Norm. Univ.* **1992**, *2*, 161–166. (In Chinese)