Article

Mapping the Soil Texture in the Heihe River Basin Based on Fuzzy Logic and Data Fusion

Ling Lu [1,2], Chao Liu [1,2,3], Xin Li [1,2,4] and Youhua Ran [1,2,*] (iD)

[1] Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China; luling@lzb.ac.cn (L.L.); liuchao4@csrzic.com (C.L.); lixin@lzb.ac.cn (X.L.)
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Zhuzhou Electric Locomotive Research Institute Co., Ltd., China Railway Rolling Stock Corporation (CRRC), Zhuzhou 412001, China
[4] CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing 100101, China
[*] Correspondence: ranyh@lzb.ac.cn; Tel.: +86-931-4967964

Abstract: Mapping soil texture in a river basin is critically important for eco-hydrological studies and water resource management at the watershed scale. However, due to the scarcity of in situ observation of soil texture, it is very difficult to map the soil texture in high resolution using traditional methods. Here, we used an integrated method based on fuzzy logic theory and data fusion to map the soil texture in the Heihe River basin in an arid region of Northwest China, by combining in situ soil texture measurement data, environmental factors, a previous soil texture map, and other thematic maps. Considering the different landscape characteristics over the whole Heihe River basin, different mapping schemes have been used to extract the soil texture in the upstream, middle, and downstream areas of the Heihe River basin, respectively. The validation results indicate that the soil texture map achieved an accuracy of 69% for test data from the midstream area of the Heihe River basin, which represents a much higher accuracy than that of another existing soil map in the Heihe River basin. In addition, compared with the time-consuming and expensive traditional soil mapping method, this new method could ensure greater efficiency and a better representation of the explicitly spatial distribution of soil texture and can, therefore, satisfy the requirements of regional modeling.

Keywords: soil texture; soil land inference model; data fusion; arid region

1. Introduction

Soil texture is one of the most basic physical properties of soils. Different soil textures that vary in their composition and contents of specific particles have very different soil hydraulic characteristics, such as water retention and hydraulic conductivity, and soil thermal parameters, such as thermal conductivity and heat capacity [1]. All of these differences significantly affect soil hydrological processes, including soil moisture content, evaporation, transpiration, and infiltration. Therefore, the soil texture type, which has an important influence on the simulation of water cycle and surface fluxes, is a critical parameter in land surface process models, hydrological models, and land surface process-coupled atmospheric models. Currently, these models always require soil properties as inputs according to the soil texture classification scheme of the United States Department of Agriculture (USDA). However, most available soil maps in China are based on the Chinese occurrence taxonomy or diagnosis taxonomy classification criteria, which cannot be directly used in land surface and distributed hydrological models. Therefore, the exploration of digital soil mapping methods that can directly generate accurate Chinese soil texture maps using existing soil environmental data and soil survey data would be of great value and necessity.

The digital soil mapping technique began with the proposal of a "clorpt equation" by an American pedologist, Hans Jenny, in 1941. This equation describes the soil as a function of the climate, living beings, topography, parent material, and time [2]. Since then, many digital soil mapping methods based on the soil-landscape relationship have been developed [3], including linear regression models [4], discriminant analysis [5], decision tree methods [6], geo-statistical models [7], artificial neural network models [8], and fuzzy clustering and fuzzy logic methods [9,10]. Among these statistical models, the linear regression model has a simple structure and is widely accepted, with relatively good outcomes from its application. However, a linear model cannot explain the complicated interactions between processes in the soil [11]. Discriminant models have a similar structure to linear regression models and can also be used to derive relatively good results when used to predict soil types. However, the discriminant model has problems similar to those of the linear model. A decision tree model can represent non-linear relationships and provides relatively good support for different data types. However, in situations with relatively few nodes, it often predicts unrealistically abrupt changes in soil type [11]. Geo-statistical models have the best prediction accuracy among the statistical models; their primary disadvantage is their complexity and high computation cost. The variance function needs to be manually selected based on experience, and the applicability of such models is limited to the region set in the model [12]. The fuzzy clustering model has certain advantages in terms of spatial extrapolation and can be combined with expert knowledge [13]. Its disadvantage is that there are still many problems with knowledge acquisition, and the validity and accuracy of the rules for such a model cannot be guaranteed.

The aforementioned digital soil mapping methods have mostly been developed for regional scales and serve for soil texture mapping at multiple scales. However, efficient digital soil mapping has not been conducted in the Heihe River basin (HRB). Many soil geological surveys have been conducted in the past, amassing a considerable collection of profile survey data on soil texture. There are also some existing soil comparisons at the catchment scale. For example, Gao et al. collected 53 sub-type soil profiles within the basin and established the distribution of soil texture types in the HRB at a 30-s resolution based on a sub-type soil distribution map (1:10,000 scale) of the HRB by comparing the map with a soil triangular diagram relating the mapped sub-types to the corresponding international universal soil texture type classification criteria of State Soil Geographic (STASGO) [14]. However, integrated eco-hydrological research in the HRB still uses global and regional soil texture maps, such as the global soil texture map (1:1,000,000 scale) published by the United Nations Food and Agriculture Organization (FAO) in 2009 [15], the China soil texture map developed by Institute of Soil Science, Chinese Academy of Sciences [16,17] and the regional soil texture map of China developed by Beijing Normal University [17]. The HRB is an experimental basin used for the integrated study of biological hydrology in China, and a set of high-quality soil texture maps is urgently needed for an integrated eco-hydrological model of Heihe. With the accumulation of ground survey data and the development of mapping methods recently, the new soil texture map in HRB have been produced or are developing [18–20].

This study develops a digital mapping scheme for soil texture at the basin scale. Based on fuzzy logic inference theory, this scheme uses a Geographic Information System (GIS) to combine multiple environmental factors that affect soil texture with available soil type maps and soil texture survey data to extract the relationship between soil and environment through artificial intelligence, machine learning, and expert knowledge. This produced a high-quality and high-resolution soil texture map for the eco-hydrological model of the HRB. This paper is organized as follows: This section introduces the need for soil texture mapping in Heihe River Basin. Section 2 primarily describes the characteristics of the research area of the HRB. Section 3 describes the data and methods. Section 4 analyzes and verifies the results and compares them with those of other soil maps, and Section 5 provides a summary of this paper.

2. Heihe River Basin

The HRB, the second largest inland river basin of China, has been used as an experimental river basin to pursue the integrated study of watershed science and integrated river basin management

for a long time [21–23]. It is located at 97.1° E–102.0° E and 37.7° N–42.7° N, and the basin's area is approximately 143,000 km^2. Its upstream area includes the Qilian Mountains at the northern margin of the Tibetan Plateau, where the elevation is approximately 2000–5500 m and the annual precipitation exceeds 350 mm. This is the runoff-generating area and the water catchment area of the HRB. The landscapes includes glaciers, cold desert, alpine meadow, shrub meadow, forest, grassland, and desert grassland from high to low elevation. In the midstream area, irrigated crops and shelter forest are distributed in the lower part of the piedmont alluvial fan and the river alluvial plain, exhibiting an oasis landscape dominated by artificial vegetation. The annual average air temperature, precipitation and potential evaporation are approximately 7 °C, 117 mm, and 2390 mm, respectively. The elevation ranges between 1300 and 1700 m. The downstream area is dominated by the Gobi Desert. In the lake basin depression along the delta and the alluvial fan on both sides of the Heihe River, there are areas of desert riparian forest, shrub and meadow vegetation. On the basis of natural landscape differentiation along the altitudinal gradient and block integrity, the HRB can be divided into six landscape zones [24]: Forest-steppe zone in the Qilian Mountains (region 1), Oasis zone in the Hexi Corridor (region 2), Desert-gobi zone in the Northwestern Hexi Corridor (region 3), Ruoshui River Delta zone (region 4), Desertified zone in the Gurinai Lake Basin (region 5), and Desert-gobi zone in the Northern Alxa High-plain (region 6), as shown in Figure 1.
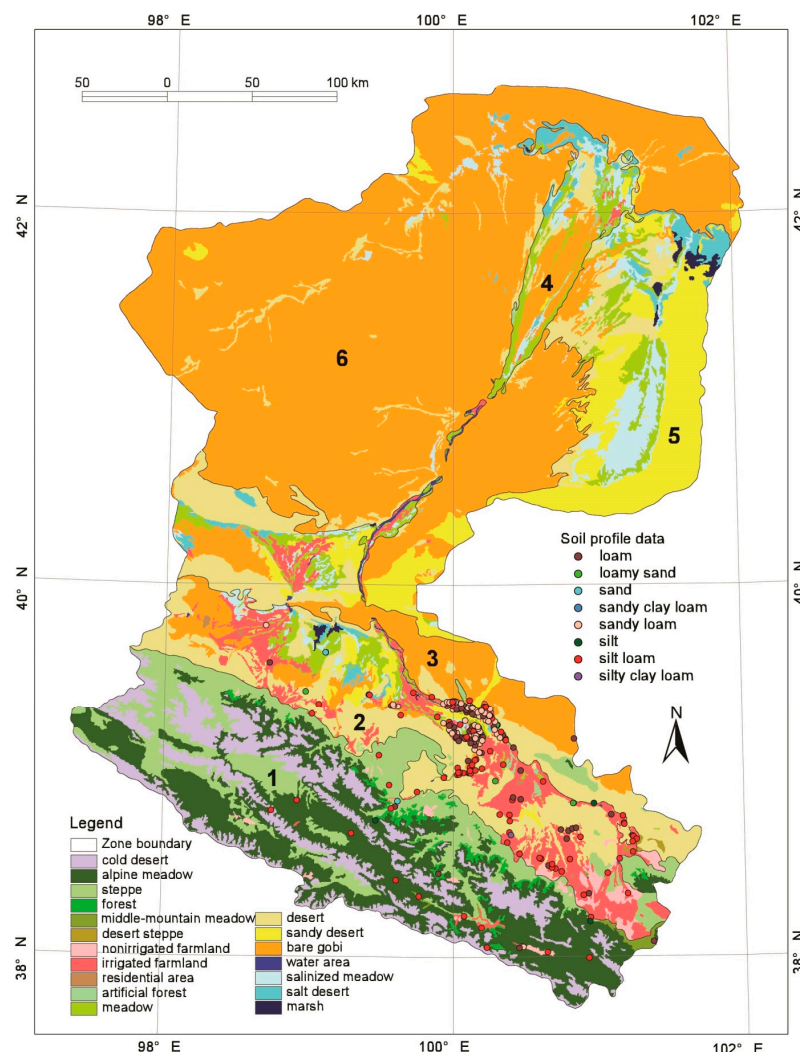

Figure 1. The landscape and the soil profile distribution of the Heihe River basin (the meanings of landscape zone numbers can be found in the text) (Modified from [24]).

3. Data and Methods

3.1. Data

3.1.1. Soil Profile Data

Measurement data were collected regarding the soil texture from a total of 220 profiles in the HRB. Due to the variety of data sources, the vertical sample soil layer and grading of the grain sizes was not uniform. First, we unified the grain size according to the percentage by weight of individual sizes in the soil, including sands (0.05–2 mm), clay (<0.002 mm), and silt (0.002 mm–0.05 mm), and we then applied the soil texture classification criteria of the USDA to derive the soil texture types present in the HRB. Under the soil texture classification criteria of the USDA, soil texture is divided into 12 types, whereas the soil profile of the HRB only covers eight types, the three dominant types of soil texture being silt loam, loam, and sandy loam. Specifically, the compositional proportion of silt loam is 42%, the proportion of loam is 25%, and that of sandy loam is 26%. The combined proportion of the other five types of soil texture present in the basin (i.e., loamy sand, sand, sandy clay loam, silty clay sand, and silty sand) is less than 8%. The soil profiles are mainly distributed in the midstream and upstream area of the HRB (Figure 1), and the soil profiles in the midstream area were mostly concentrated in Linze County, Ganzhou District, and Shandan County of Zhangye City, with fewer profiles in the west region; in the upstream area of the basin, the soil profiles were primarily distributed in the valley area.

3.1.2. Environmental Factors

The formation of soil is influenced by many environmental factors, such as topography, climate, biology, parent material, and time. According to the characteristics of the HRB and the data availability, this study selected 13 environmental factors as impact factors that affect the spatial characteristics of soil texture distribution in the HRB, including topographic factors (i.e., slope, curvature along the cross section, and curvature along the plane) derived from digital elevation models (DEM), biological factors (vegetation type and vegetation index), climatic factors (precipitation, air temperature and radiation), parent material factors (topographic types and geological types), land use types, and distance from rivers.

The DEM data were derived from the SRTM (Shuttle Radar Topography Mission) elevation product with a 90-m resolution. The vegetation type map is sourced from the vegetation map of China (1:1,000,000 scale) [25]. The vegetation index used a long-term series of vegetation index data based on SPOT (Système Pour l'Observation de la Terre) imagery in China [26]. The climatic factors were estimated by using the Global Land Data Assimilation System (GLDAS) and a digital elevation model of the HRB with a spatial resolution of 1 km to obtain atmospheric driving data for the HRB from 2001–2010 with a spatial resolution of 1 km (multi-year average precipitation, air temperature and radiation were used here). The topographic data were sourced from a digital topographic dataset (1:1,000,000 scale) of Western China [27]. The soil parent material data were derived from a national geological data gallery (1:500,000 scale) [28,29], and the land use factor was based on a land use map (1:100,000 scale) for China in 2000 [30]. These data were all obtained from the Environmental and Ecological Science Data Center for West China [31].We pre-processed nine continuously-distributed environmental factors, including treatment of outliers and standardization. The nine environmental factors include elevation, slope, curvature along the cross section, curvature along the plan, vegetation index, precipitation, air temperature, radiation, and distance from rivers. The treatment of outliers was performed to eliminate extreme values in the data. The basis for the classification of values as outliers was the distribution histogram for statistics of environmental factors; singular values at the end of the distribution histograms were excluded. This is the simplest technique and is commonly used for detecting outliers. In addition, the order of magnitude differed significantly between different environmental factors, and this could seriously affect the classification results. Therefore, before the clustering, we performed a standardization of the data by uniformly stretching the environmental

factors with values greater than 0 into the range of 0–100 and stretching the environmental factors with negative values into the range of −50 to +50. The 0 values, such as curvature, remain unchanged. Moreover, the order of magnitude of the original values for the two curvature factors was 10^{-3} or even lower, and there were too many significant digits. Therefore, before the calculation, we performed a uniform multiplication of the values. Although the absolute values of the data became larger, this did not affect the differences or spatial distribution of the values or the performance of the soil texture predictions.

3.2. Method

The soil-landscape model is now widely used for regional soil prediction. This method is based on the classical soil genesis theory, known as the Jenny equation, which defines the soil as a function of climate, biology, topography, parent material, and time. This function was later simplified to the following form:

$$S = f(E) \tag{1}$$

where S is soil, E represents a vector of the formative environments, and f represents the relationships between the soil and the environmental conditions.

Using geographic information analysis techniques, such as GIS and remote sensing (RS), we can quantitatively develop relationships between soil distributions and their environmental factors, including climate, terrain, vegetation, and so on. Combining these relationships with adequate soil profile samples, it becomes possible to derive regional soil prediction maps based on the soil-landscape model. In this study the Soil Land Inference Model (SOLIM) was used to map the soil texture, which was jointly developed by the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, the Department of Geography of the University of Wisconsin, and the Department of Geography of Dartmouth College [10,32]. SOLIM uses a suite of GIS and RS techniques to characterize environmental conditions and knowledge acquisition techniques to extract soil-landscape relationships. Environmental conditions are integrated with the extracted soil-landscape relationships to infer the spatial distribution of soil types under fuzzy logic. Fuzzy clustering is used to obtain different environment combinations to better represent transition zones in the soil-landscape. A more detailed description of SOLIM can be found in Zhu et al. and Zhu [10,32]. SOLIM assumes the soil at any location can be expressed by a soil similarity vector. The element of the vector is a membership value that used to measures the similarity between the soil and environmental factors at a certain location. Soil texture measurements from 220 profile points and an old soil texture map were used to train the soil-landscape model (see below for the details). The key procedure in using the SOLIM model for soil mapping is constructing the membership function curve of soil-environmental factors. Based on the varying landscape features of the six landscape zones in the HRB (Figure 1), we applied different methods to construct the membership function curves of the soil-environmental factors.

3.2.1. Soil Texture Mapping for the Forest-Steppe Zone in the Qilian Mountains

A small number of soil profile points were distributed in the forest-steppe zone in the Qilian Mountains. These profiles have a high degree of topographic variation. The climate and vegetation have clear characteristics of vertical differentiation. It is appropriate to use a method that combines targeted sampling, a non-probability sampling, and fuzzy logic mapping. The objective of targeted sampling used here is to capture the spatial pattern of soil variation effectively. Except the existing small number of soil profile points, an old soil texture map is used to support the targeted sampling (see below). The fuzzy C-means (FCM) method is used to identify the environmental configuration in this study. For this zone, we implemented FCM clustering for the topographic factors (i.e., elevation, slope, curvature along the cross-section, curvature along the plan), climatic factors (precipitation, air temperature, radiation), and distance from rivers, and derived the quantitative membership relationship between the combination of environmental factors and soil type according to a small

number of soil profile points. Then, we used the values of environmental factors in regions with high membership as the typical environmental value for the soil distribution and then used the SOLIM model to generate a soil texture map and fuzzy membership graph. This process includes two steps as follows:

(1) Extraction of the Soil-Landscape Relationship

We first implemented the FCM clustering of nine treated environmental factors (see Section 3.1.2) for the landscape in this zone and selected the optimal numbers of categories 'c' and fuzzy degree 'm'. First, we calculated the fuzzy matrix derived through fuzzy clustering with different 'm' and calculated the partition coefficient and entropy. Then, according to the selected criteria of the optimal number of categories, we determine the optimal classification. As shown in Figure 2, for m = 2.5, there is no optimal number of categories; for m = 1.75, the optimal number of categories is 13 or 14; for m = 1.5, the optimal number of categories is 6, 8, 12, or 15; for m = 1.25, the optimal number of categories is 7, 8, 11, 14, 16, or 17. The most frequent optimal numbers of categories were 8 and 14. According to the mapping scale, we determined the optimal number of categories to be 8. Then, according to the determination principle of optimal 'm', we determined the fuzzy objective function and fuzzy constraint function for different 'm' values given that the number of categories is eight. The objective function and fuzzy function curves intersect at m = 1.55 and, thus, we selected m =1.55 as the optimal fuzzy degree.
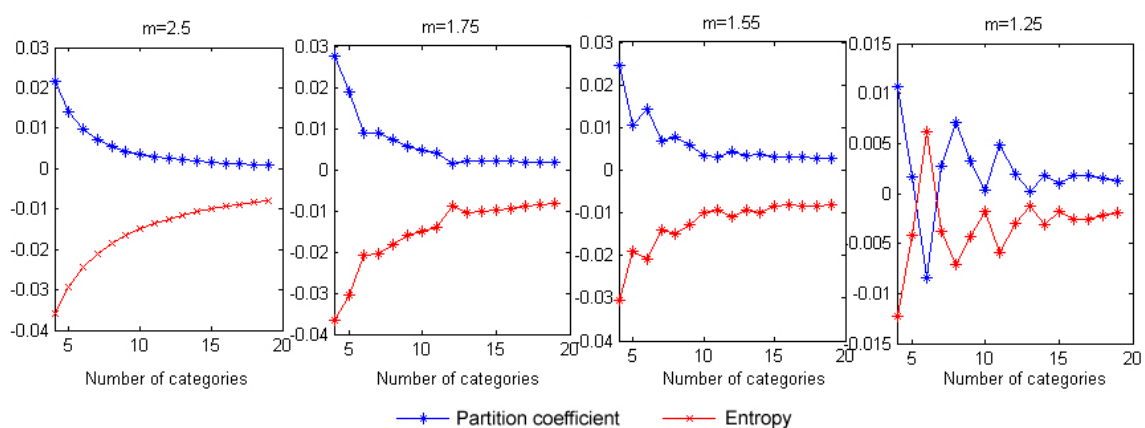

Figure 2. Variation of partition coefficient and partition entropy with the number of categories for different fuzziness values 'm'.

After the number of categories 'c' and the fuzziness value 'm' were determined, we implemented the FCM clustering of environmental factors to generate membership graphs for individual environmental combinations. We obtained a total of eight different environmental combinations, which correspond to different spatial distributions. Then, we calculated the type of environmental combination for the maximum membership of individual environmental combinations as the new discrimination type. For example, categories 3, 4, and 5 were adjacent and their environmental combination results were similar; thus, categories 3, 4, and 5 were combined. In the spatial analysis module of ArcGIS (Redlands, California, USA), we calculated the region where the membership of individual environmental combinations is greater than 0.7 and used the range of environmental factors in that region as the typical values of environmental conditions for the soil type distribution. The range of values with the existence of six major categories was used as the type-II knowledge rule for the existence of soil type (see Table 1). The range of values for the first row of environmental factors represents the range of environmental values with membership greater than 0.7 and is used as the type-I knowledge for the mapping of the SOLIM model. The second row represents the range of values when the environmental condition deviates from typical environmental conditions, where 's' indicates an s-type curve and 'z' a z-type curve.

Table 1. Typical environment and deviation environment for the existence of soil types.

	Knowledge Type	Type 1	Type 2	Types 3/4/5	Type 6	Type 7	Type 8
Elevation	Type-I knowledge	3208–4377	3088–4168	3656–4820	1837–3479	2566–3803	2840–4231
	Type-II knowledge	2918–4653	2558–4702	2569–s	z–5025	2334–4326	1937–4890
Vegetation Index	Type-I knowledge	54.2–90.8	66.4–102.9	26.6–64.5	42.8–82.3	84.3–152.5	44–85
	Type-II knowledge	31–105	35–128	z–112	24–98	66–s	24–126
Plane curvature	Type-I knowledge	−17.3–20.8	−20.3–16.1	−25.7–12.1	−27.3–25.7	−16.4–22.4	−11–15
	Type-II knowledge	−287–196	−299–169	z–192.6	−173–s	−179–349	−63–164
Cross section curvature	Type-I knowledge	−11.4–13.8	−11.6–15.3	−28.4–13.2	−13.8–19.4	−8.9–20.1	0.5–27.9
	Type-II knowledge	−53–66	−73–66	−73.2–50.7	z–85	−64–83	−86–s
Radiation	Type-I knowledge	4724.8–5152	4647–5228	4194–4997	4782–5658.3	4795.3–5431	4681–5317
	Type-II knowledge	4376–5535	4156–5544	z–5414	4431–s	4309–5669	3617–5719
Precipitation	Type-I knowledge	255.1–500.8	611–971.4	298.2–725	99.6–387.3	320.6–673	243–596
	Type-II knowledge	212–720	404–s	172–1247	z–676	212–833	129–1248
Slope	Type-I knowledge	0.2–7.1	4.9–13.5	6.9–17.2	0.5–10.7	2.1–11.6	11.8–19.9
	Type-II knowledge	z–11.6	0.75–28.47	0.91–31.9	z–17.67	z–19.3	5.58–s
Air temperature	Type-I knowledge	−8.7 to −1	−10.4 to –1	−12.6 to −2.2	−3.3–7.3	−5.5–2.2	−9.43 to −0.24
	Type-II knowledge	−14.6–1.11	−14.1–0.1	z–2.5	−17.3–s	−14.3–4.8	−19.8–6.5

(2) Mapping with the SOLIM Model

The upstream area of the HRB contained 24 soil measurement profiles including four soil types. Silt loam accounted for 18 of the profiles, representing the vast majority. There was also one sample point of sandy loam, two sample points of silty sand, and three sample points of loam. The old soil texture map [18] was used to connect the category combination generated by fuzzy c-means clustering and soil types, and the membership curve corresponding to soil types and environmental factors was edited in the SOLIM model. According to the membership curve, we can infer the membership graph of individual soil types. The range of the membership graph is 0–100. A higher membership value indicates greater similarity to this soil type. A membership of 100 indicates full belonging to a certain type of soil, and a membership of 0 indicates a total absence of belonging to a certain type of soil. The membership graph was used to generate the preliminary soil type map after hardening. Except the four soil types, i.e., silt loam, loam, sandy loam, and silty sand included in soil survey data, silty clay loam, and sandy clay loam included in old soil texture map also was integrated to derive a soil texture map of the Qilian Mountains forest meadow landscape.

3.2.2. Soil Texture Mapping of the Oasis Zone in the Hexi Corridor

The oasis zone in the oasis zone in midstream area of the Hexi Corridor contained a large number of soil profile points, where the overall terrain is flat. The dominant environmental factors are land use and topography and, thus, for this region, a knowledge mining method based on a decision tree, characterized by a hierarchical organization of rules, was adopted. We first used a decision tree algorithm to generate the soil type distribution and decision-making rules for the environmental data, determining the typical environmental values of the soil distribution and then using the SOLIM model to generate the soil map and the fuzzy membership graph.

(1) Extraction of the Soil-Landscape Relationship

In this area, there were a total of 190 soil profiles. Among these, 141 soil profiles were used as the training sampling points to develop soil-landscape model. The selection criteria of training sampling points is that distribution can reflects the difference and diversity of the environmental conditions governing soil distribution. The remaining 49 soil profiles were used as verification points. As for the selection of environmental factors, we first used the soil mapping method based on a decision tree to rapidly select environmental factors from 13 factors mentioned in Section 3.1.2. By comparing the mapping effects of different combinations of environmental factors, we selected the optimal combination of environmental factors by comprehensively considering the prediction accuracy of training points and verification points. The terrain in this zone is relatively flat and, thus, there

was no need to include many topographic factors. Additionally, the testing results for the mapping based on the decision tree indicated that the introduction of terrain curvature factor resulted in a soil map that was fragmented and unrealistic and, therefore, the curvature factors were not taken into account. As for type factors, the geological map and vegetation map did not improve mapping accuracy, which was higher after these factors were excluded. By comprehensively considering the results of the analysis above, we finally selected a total of nine environmental factors in this area, including elevation, slope, precipitation, air temperature, radiation, vegetation index, distance from rivers, geomorphology, and land use.

We extracted the environmental condition values of individual profile points and the decision tree and decision-making rules was generated. In this region, we adopted a decision tree with more abundant information as the basis for extracting the soil texture type-landscape relationship. By working downward in sequence from the root node of the decision tree, we recorded the generated paths of the decision tree and converted them to descriptions of soil environments. Since not all combinations of environmental factors have a distribution of representative soil profiles in this area, the decision tree model arbitrarily allocated the leaf nodes of the decision tree in the prediction of unknown regions. This created some rules in the decision tree model that were unreasonable and, thus, some expert knowledge pertaining to the relationship between soil texture types and environmental factors was used for supplementation. For example, in the region where the type of geomorphology is sand, the sand content in the corresponding soil texture type is high. In the generated decision tree, the utilization rate of individual environmental factors was ordered from high to low as follows: air temperature (100%), vegetation index (81%), landscape (74%), precipitation (66%), distance from rivers (48%), slope (42%), land use type (36%), elevation (32%), and radiation (30%). This reflects the different influences of the various environmental factors in this area on the distribution of soil textures.

(2) Mapping with the SOLIM Model

For the oasis landscape zone, the relationship between soil and environment derived from decision tree learning can be used as the typical environmental condition for soil texture existence, i.e., the first kind of knowledge required by SOLIM model mapping. Mapping with the SOLIM model also requires a second kind of knowledge, namely, the width of the membership curve. The second kind of knowledge in this situation is generally established by comprehensively considering the range of environmental conditions and the amplitude of variation. In this area, the second kind of knowledge for individual continuous environmental conditions is as follows: elevation (200), slope (5), distance from rivers (100), vegetation index (3), radiation (100), air temperature (5), and precipitation (100). Therefore, according to the obtained first and second kinds of knowledge, in the SOLIM model, we edited the membership graph between individual soil types and existing environmental factors and then derived a fuzzy membership graph of individual soil types through automatic inference. The membership graph is processed through hardening to generate an initial soil type map, and the land use type map is then integrated to extract the non-soil factors, such as bare rocks and water. Fusion then generates a soil texture map for the oasis landscape zone in the midstream area.

3.2.3. Soil Texture Mapping in the Downstream Area

No soil profile is located in the downstream area of the HRB, or in where the overall terrain is flat. Thus, it is neither suitable for the targeted sampling method nor for the knowledge mining method based on a decision tree. Therefore, in this area, we used an old soil texture map generated by previous studies to perform knowledge mining. Based on this existing soil map and environmental data, we examined the distribution histogram of each soil type and environmental factor, and determined the typical environment values associated with each type of soil. Then the SOLIM was used to generate a soil map and fuzzy membership graph.

(1) Extraction of the Soil-Landscape Relationship

The available soil texture maps for the HRB include the global soil texture map (1:1,000,000) published by the United Stations Food and Agriculture Organization (FAO) in 2009 [15] and the China

soil texture map developed by Beijing Normal University [18]. The validation results indicate that the latter has better accuracy than the soil texture map of FAO. Therefore, it is used for knowledge mining in this study. Based on the similar characteristics of their environmental factors, we combined the Ruoshui River Delta zone (region 4) and the Desertified zone in the Gurinai Lake Basin (region 5) as a new region and mapping soil texture for this new region, region 3 and region 6, respectively. Since the quality of the data on climatic factors in region 4, 5, and 6 is relatively poor, the climatic factors were excluded; the terrain in regions 3, 4, and 5 is flat and, thus, the curvature factor was not taken into account.

In the soil map, one soil type often includes many polygons. If each polygon were treated as an independent entity, we can calculate the soil-environmental factor relationship curve corresponding to each polygon. If all of the polygons for one soil type were treated as one entity, we can also calculate an average soil-environmental factor relationship curve, which is called the overall average relationship curve. For the independent relationship curve of each polygon, we can compute its similarity with the overall average relationship curve, which is expressed as the similarity curve. If the similarity between an independent relationship curve and the overall average relationship curve is too low, we classify this polygon as a singular polygon, which is called an outlier. The soil-environmental factor distribution curve expressed by the outlier does not follow the general rule, and the outlier should be excluded and the average curve recalculated. Similarly, the mode curve expresses the occurrence frequency of environmental factors values. For values with a particularly low occurrence frequency can also be considered as outliers or excluded individually to recalculate the overall average relationship curve. The newly-generated curve. after excluding outliers. can be considered as the curve that correctly reflects the relationship between soil type distribution and landscape factors by eliminating erroneous values.

(2) Mapping with the SOLIM Model

For the downstream area of the basin, we investigated the frequency distribution curve of each soil type for individual environmental factors in each landscape zone. After the outliers were excluded, this curve was directly introduced into the SOLIM model as the membership curve for the distribution of soil and environmental factors, and then the fuzzy membership graphs of the soil types in individual areas were generated through automatic inference. Finally, the soil type map was generated by a hardening process of the soil membership graph yields.

4. Validation and Results

4.1. Results

For each of the six ecological landscape zones in the HRB, we applied the extracted knowledge rules and SOLIM model to create a predictive map of soil texture and then integrated non-soil information (such as a glacier map and water map) to generate a final soil texture map of the HRB, as shown in Figure 3. Figure 3 indicates that the upstream area of the HRB is dominated by loam, sandy loam, and silt loam. Loam and sandy loam are primarily distributed in the mountainous area of the upstream area, and silt loam is primarily distributed in the valley area of the upstream area. The soil texture of the upstream area has distinct altitudinal zonation, and topographic factors play a dominant role in the distribution of soil texture types. The soil of the artificial oasis in the midstream area of the HRB is dominated by loam and sandy loam, and other areas are dominated by silt loam and loamy sand. The distribution of soil textures in the midstream area of the basin is primarily affected by land use, topography, and distance from rivers, and the structure of the distribution is relatively clear and definite. The soil structure of the downstream area is relatively simple and is dominated by loamy sand, silt loam, and large areas of bare rock.
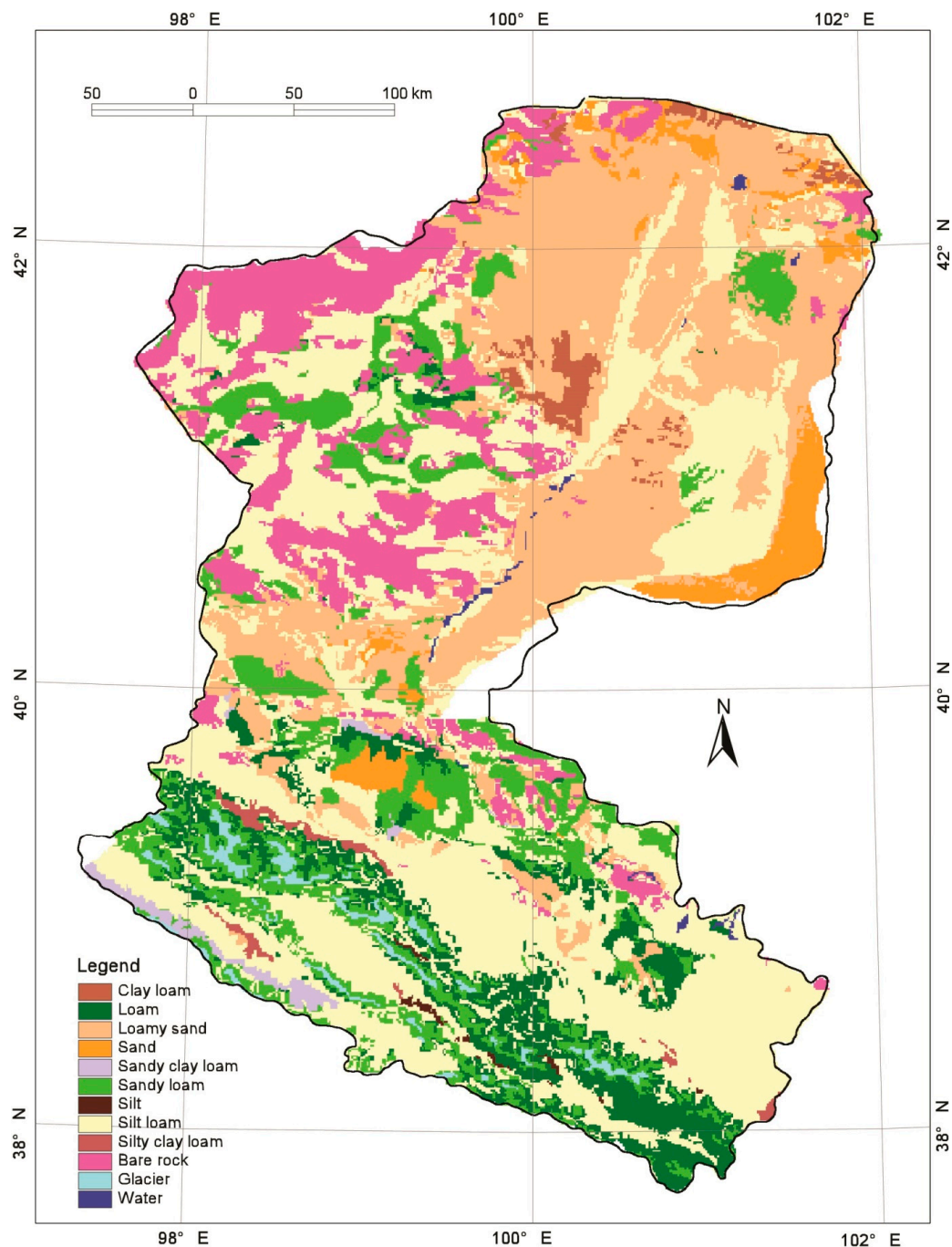
Figure 3. Soil texture map in the Heihe River Basin from this paper.

4.2. Validation

The reliability of the results was validated in two ways. We first used 49 verification points to assess the accuracy of the derived soil texture type map. After selecting a subset of the soil profile data that represents the diversity of the soil texture distribution as training points, the remaining profile points were used as verification points (See 3.2.2). Since the majority of the profile points were distributed in the midstream area due to the limitations of the available data, we used the verification points to assess the accuracy of the soil texture map by focusing on the midstream area of the HRB. The confusion matrix calculated from the comparison between verification points and the soil map derived in this study is shown in Table 2. The rows in Table 2 represent the number of sampling points

for each soil type in the field measurements, and the columns represent the number of sampling points for each soil type in the result map. For example, for silt loam in Table 2, there are a total of 25 points for this type of soil in the soil survey data, and there are only 21 points for this soil type in result map. Among the 25 silt loam sampling points in the field data, 17 sampling points are correctly classified in the map, and eight are classified incorrectly as other types of soil. The producer's accuracy is 17/25 = 0.68 and, thus, the producer's accuracy represents the level of correct classification of the soil type map. The overall accuracy of soil texture classification in the midstream area of HRB was 69.4%.

Table 2. Confusion matrix for the comparison of the soil survey data with the soil map produced by this study.

	Silt Loam	Loam	Sand	Loamy Sand	Sandy Loam	Total	Producer's Accuracy
Silt Loam	17	2	0	2	4	25	0.68
Loam	2	5	0	0	2	9	0.56
Sand	0	0	1	0	0	1	1
Loamy Sand	0	0	0	0	0	0	0
Sandy Loam	2	1	0	0	11	14	0.79
Total	21	8	1	2	17	49	
User's accuracy	0.81	0.63	1	0	0.65		
Overall accuracy = 0.694							

Since the soil texture map for the desert and lake basin landscape zone in the downstream area of the HRB is based on the old soil texture map, the extracted relationship between soil texture and environment is derived from fuzzy logic inference mapping; the soil texture mapping for the Qilian Mountains forest meadow landscape zone in the upstream area of HRB also refers to the old soil texture map and, thus, we also used the 49 verification points to validate the old soil texture map. The results indicate a classification accuracy of 38.7% for the old soil texture map, meaning that the accuracy of the results in this study was considerably improved. Based on a confusion matrix of the two texture maps, the overall consistency is 46.7%, and the kappa coefficient is 30.5%.

In comparison with the old soil texture map, the soil texture map generated in this study based on the fuzzy logic method yields considerably higher accuracy by characterizing the correlations between soil types and typical environmental factors, and the new map clearly reflects the different conditions of soil formation in the upstream, midstream and downstream area of the HRB. The spatial distribution of soil texture is more reasonable in the updated map, with richer details and stronger continuity. For example, in the upstream area of the HRB, the map clearly exhibits the altitudinal variation in soil texture and better retains the complexity and integrity of plaque structures in the midstream area of the HRB.

4.3. Data Availability

The soil texture map produced by this paper are available upon request from the corresponding author or download from Cold and Arid Regions Science Data Center at Lanzhou, a member of World Data System [33].

5. Conclusions

This paper applies the fuzzy logic inference mapping theory to the entire HRB, which has a large area and strong natural differentiation, and covers three distinct landscape features (i.e., mountains, plain oasis, and desert). Using multiple knowledge mining methods, we extracted the soil-environment relationships in each zone of the basin, providing input data for the fuzzy logic mapping. In addition, by combining multi-sources information, several thematic maps, such as a glacier map, a water map, and a desert map, were integrated to improve the soil texture map. The main conclusions are as follows:

(1)　The new soil texture map produced by this study are more reliable than existing maps. This indicates that by combining the large number of soil profile data with environmental factors, we

can obtain relatively good soil texture prediction results by applying the method based on the combination of a decision tree and fuzzy logic.

(2) The targeted sampling method of fuzzy c-means is more applicable to regions with relatively large variations in topographic factors, when we can achieve the goal of increasing efficiency and implement soil prediction with the fewest possible sampling points; in vast areas with flat terrain, the environmental factors are the dominant controls, and the fuzzy c-means method is not applicable. In regions with flat terrain, we still require a number of sampling points to characterize the soil-landscape relationship and achieve fuzzy logic inference mapping.

(3) Since the influence of the training samples is very strong in the shaping of the decision tree, the level of utilization efficiency changes significantly for the environmental factors based on the differences among the training samples and the combination of different environmental factors. Therefore, from the perspective of the level of utilization efficiency for the environmental factors in the decision tree, the influence of each environmental factor on soil formation is relative rather than absolute. We can only determine the dominant environmental factor once the combination of training samples and environmental factors is determined.

(4) In the application of the SOLIM model to mapping soil texture for six landscape zones, there was no occurrence of texture plaque rupture at the boundaries of the individual zones. It is, therefore, reasonable to implement soil mapping by dividing the HRB into six zones based on ecological function. This also reflects the feasibility of using the fuzzy logic method for soil mapping.

(5) Based on the fuzzy logic method, we can overcome the disadvantages of relatively low efficiency and relatively poor accuracy associated with the traditional soil mapping method. The generated soil map is expressed in raster format, which can more accurately characterize the spatial gradient features of soil. In comparison with traditional mapping methods, most processes in the mapping scheme in this study were accomplished with computers, and the mapping cycle is short. For example, we can update the current soil map by introducing new soil profiles or other soil distributions.

Therefore, we conclude that, with limited soil profile data and full consideration of the impact of environmental factors, the method presented here is validated for digital soil mapping based on fuzzy logic to obtain a digital soil texture map and real-time updates at the river basin scale. This approach can generate soil texture data that satisfies the requirements of watershed-scale eco-hydrological models, regional climate atmospheric models, and potentially improve the simulation accuracy of those models.

Author Contributions: All authors worked together on the manuscript. L.L. and X.L. conceived and designed the research; C.L. collected and analyzed the data; Y.R. prepared the drafts of the paper, submitted it online and response to reviewer's comments. All authors wrote and reviewed the paper.

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

References

1. Gao, Y.H.; Chen, F.; Barlage, M.; Liu, W.; Cheng, G.D.; Li, X.; Yu, Y.; Ran, Y.H.; Li, H.Y.; Peng, H.C.; et al. Enhancement of land surface information and its impact on atmospheric modeling in the Heihe River Basin, northwest China. *J. Geophys. Res. Atmos.* **2008**, *113*, D20. [CrossRef]
2. Jenny, H. *Factors of Soil Formation: A System of Quantitative Pedology*; McGraw-Hill: New York, NY, USA, 1941; p. 281.
3. McBratney, A.; Santos, M.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [CrossRef]

4. McKenzie, N.J.; Austin, M.P. A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. *Geoderma* **1993**, *57*, 329–355. [CrossRef]

5. Sinowski, W.; Auerswald, K. Using relief parameters in a discriminant analysis to stratify geological areas with different spatial variability of soil properties. *Geoderma* **1999**, *89*, 113–128. [CrossRef]

6. Henderson, B.L.; Bui, E.N.; Moran, C.J.; Simon, D.A.P. Australia-wide predictions of soil properties using decision trees. *Geoderma* **2005**, *124*, 383–398. [CrossRef]

7. Odeh, I.O.; McBratney, A.B.; Chittleborough, D.J. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. *Geoderma* **1995**, *67*, 215–226. [CrossRef]

8. Chang, D.; Islam, S. Estimation of Soil Physical Properties Using Remote Sensing and Artificial Neural Network. *Remote Sens. Environ.* **2000**, *74*, 534–544. [CrossRef]

9. Goktepe, A.B.; Altun, S.; Sezer, A. Soil clustering by fuzzy c-means algorithm. *Adv. Eng. Softw.* **2005**, *36*, 691–698. [CrossRef]

10. Zhu, A.X.; Hudson, B.; Burt, J.E.; Lubich, K. Soil mapping using GIS, expert knowledge and fuzzy logic. *Soil Sci. Soc. Am. J.* **2001**, *65*, 1463–1472. [CrossRef]

11. Mckenzie, N.J.; Ryan, P.J. Spatial prediction of soil properties using environmental correlation. *Geoderma* **1999**, *89*, 67–94. [CrossRef]

12. Li, X.; Cheng, G.D.; Lu, L. Comparison of spatial interpolation methods. *Adv. Earth Sci.* **2000**, *3*, 260–265.

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

14. Shi, X.Z.; Yu, D.S.; Warner, E.D.; Pan, X.Z.; Petersen, G.W.; Gong, Z.G.; Weindorf, D.C. Soil Database of 1:1,000,000 Digital Soil Survey and Reference System of the Chinese Genetic Soil Classification System. *Soil Surv. Horiz.* **2004**, *45*, 129–136. [CrossRef]

15. Nachtergaele, F.; Velthuized, H.V.; Verelst, L. *Harmonized World Soil Database (version 1.1)*; FAO: Rome, Italy; IIASA: Laxenburg, Austria, 2009.

16. Shi, X.Z.; Yu, D.S.; Warner, E.D.; Pan, X.Z.; Sun, X.; Petersen, G.W.; Gong, Z.G.; Lin, H. Cross Reference System for translating Between Genetic Soil Classification of China and Soil Taxonomy. *Soil Sci. Soc. Am. J.* **2006**, *70*, 78–83. [CrossRef]

17. Shi, X.Z.; Yu, D.S.; Yang, G.X.; Wang, H.J.; Sun, W.X.; Du, G.H.; Gong, Z.T. Cross-reference Benchmarks for Correlating the Genetic Soil Classification of China and Chinese Soil Taxonomy. *Pedosphere* **2006**, *16*, 147–153. [CrossRef]

18. Shangguan, W.; Dai, Y.J.; Liu, B.Y.; Ye, A.Z.; Yuan, H. A soil particle-size distribution dataset for regional land and climate modelling in China. *Geoderma* **2012**, *171–172*, 85–91. [CrossRef]

19. Song, X.D.; Liu, F.; Zhang, G.L.; Li, D.C.; Zhao, Y.G. Estimation of Soil Texture at a Regional Scale Using Local Soil-Landscape Models. *Soil Sci.* **2016**, *181*, 435–445. [CrossRef]

20. Yang, R.M.; Liu, F.; Zhang, G.; Zhao, Y.G.; Li, D.C.; Yang, J.L.; Yang, F.; Yang, F. Mapping soil texture based on field soil moisture observations at a high temporal resolution in an oasis agricultural area. *Pedosphere* **2016**, *26*, 699–708. [CrossRef]

21. Li, X.; Cheng, G.D.; Liu, S.M.; Xiao, Q.; Ma, M.G.; Jin, R.; Che, T.; Liu, Q.H.; Wang, W.Z.; Qi, Y.; et al. Heihe Watershed Allied Telemetry Experimental Research (HiWATER): Scientific objectives and experimental design. *Bull. Am. Meteorol. Soc.* **2013**, *94*, 1145–1160. [CrossRef]

22. Cheng, G.D.; Li, X.; Zhao, W.Z.; Xu, Z.M.; Feng, Q.; Xiao, S.C.; Xiao, H.L. Integrated study of the water-ecosystem-economy in the Heihe River Basin. *Natl. Sci. Rev.* **2014**, *1*, 413–428. [CrossRef]

23. Cheng, G.D.; Li, X. Integrated research methods in watershed science. *Sci. China Earth Sci.* **2015**, *58*, 1159–1168. [CrossRef]

24. Lu, L.; Li, X.; Cheng, G.D. Landscape evolution in the middle Heihe River Basin of northwest China during the last decade. *J. Arid Environ.* **2003**, *53*, 395–408. [CrossRef]

25. Editorial Board of Vegetation Map of China, Chinese Academy of Sciences. *Vegetation Map (1:1,000,000) in China*; Science Press: Beijing, China, 2001. (In Chinese)

26. Ma, M.G.; Frank, V. Reconstructing Pathfinder AVHRR Land NDVI time-series data for the Northwest of China. *Adv. Space Res.* **2006**, *37*, 835–840. [CrossRef]

27. Editorial Board of Geomorphological Atlas of the People's Republic of China. *Geomorphological Atlas of the People's Republic of China (1:1,000,000)*; Science Press: Beijing, China, 2009. (In Chinese)

28. Editorial Group of Digital Geological Map Database of P.R. China. *Digital Geological Map Database of the People's Republic of China (1:500,000)*; China Geological Survey Bureau: Beijing, China, 1999. (In Chinese)

29. Zhang, W.H.; Lian, C.; Chen, H.; Li, S.L. MapInfo-based 1/500,000 digital geological map database of the People's Republic of China. *Seismol. Geol.* **2005**, *27*, 470–478. (In Chinese, English abstract)

30. Liu, J.; Liu, M.; Deng, X.; Zhuang, D.; Zhang, Z.; Luo, D. The land use and land cover change database and its relative studies in China. *J. Geogr. Sci.* **2002**, *12*, 275–282.

31. Li, X.; Nan, Z.T.; Cheng, G.D.; Ding, Y.J.; Wu, L.Z.; Wang, L.X.; Wang, J.; Ran, Y.H.; Li, H.X.; Pan, X.D.; et al. Toward an improved data stewardship and service for environmental and ecological science data in west China. *Int. J. Digit. Earth* **2011**, *4*, 347–359. [CrossRef]

32. Zhu, A.X. *Fine Diginal Soil Survey Model and Method*; Science Press: Beijing, China, 2008; p. 227.

33. Lu, L.; Liu, C.; Li, X.; Ran, Y.H. Soil Texture Dataset of the Heihe River Basin. Cold and Arid Regions Science Data Center at Lanzhou. 2011. Available online: http://westdc.westgis.ac.cn/data/03f0754b-8ce3-40aa-ab69-b3167154082b (accessed on 10 July 2017).