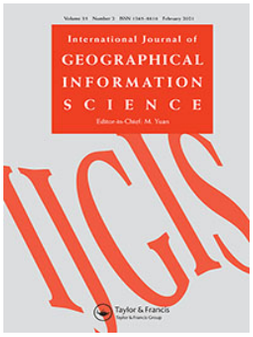

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Extracting knowledge from legacy maps to delineate eco-geographical regions

Lin Yang, Xinming Li, Qinye Yang, Lei Zhang, Shujie Zhang, Shaohong Wu & Chenghu Zhou

To cite this article: Lin Yang, Xinming Li, Qinye Yang, Lei Zhang, Shujie Zhang, Shaohong Wu & Chenghu Zhou (2021) Extracting knowledge from legacy maps to delineate eco-geographical regions, International Journal of Geographical Information Science, 35:2, 250-272, DOI: 10.1080/13658816.2020.1806284

To link to this article: https://doi.org/10.1080/13658816.2020.1806284

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2021, VOL. 35, NO. 2, 250–272
https://doi.org/10.1080/13658816.2020.1806284

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Extracting knowledge from legacy maps to delineate eco-geographical regions

Lin Yang[a,b], Xinming Li[b,c], Qinye Yang[d], Lei Zhang[a], Shujie Zhang[e], Shaohong Wu[c,d] and Chenghu Zhou[a,b,c]

[a]School of Geography and Ocean Science, Nanjing University, Nanjing, China; [b]State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing, China; [c]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China; [d]Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing, China; [e]Urban Planning Academic Infromation Center, China Academy of Urban Planning & Design, Beijing, China

ABSTRACT

Legacy ecoregion maps contain knowledge on relationships between eco-region units and their environmental factors. This study proposes a method to extract knowledge from legacy area-class maps to formulate a set of fuzzy membership functions useful for regionalization. We develop a buffer zone approach to reduce the uncertainty of boundaries between eco-region units on area-class maps. We generate buffer zones with a Euclidean distance perpendicular to the boundaries, then the original eco-region units without buffer zones serve as the basic units to generate the probability density functions (PDF) of environmental variables. Then, we transform the PDFs to fuzzy membership functions for class-zones on the map. We demonstrate the proposed method with a climatic zone map of China. The results showed that the buffer zone approach effectively reduced the uncertainties of boundaries. A buffer distance of 10–15 km was recommended in this study. The climatic zone map generated based on the extracted fuzzy membership functions showed a higher spatial stratification heterogeneity (compared to the original map). Based on the fuzzy membership functions with climate data of 1961–2015, we also prepared an updated climatic zone map. This study demonstrates the prospects of using fuzzy membership functions to delineate area classes for regionalization purpose.

ARTICLE HISTORY
Received 29 January 2020
Accepted 3 August 2020

KEYWORDS
Knowledge extraction; ecological regionalization; legacy map; buffer zone; fuzzy membership function

1. Introduction

Significant differences in environmental conditions, natural resources and human activities can cause the spatial-temporal heterogeneity of land surface (Bailey 1983, Fu *et al*. 2001, Chen *et al*. 2020). In this context, regionalization (zoning) appears to be an effective way to recognize the spatial differentiation rules of land surface elements, and provide management guidance towards sustainability (Huang 1989, Fu *et al*. 2001, Fan and Li 2009, Wu *et al*. 2016). The idea of regionalization is widely utilized in many research fields, such as agriculture (Xu *et al*. 2006, Jiang *et al*. 2018, Gelcer *et al*. 2018, Bonfante *et al*. 2018),

CONTACT Shaohong Wu wush@igsnrr.ac.cn

climate (Walsh *et al*. 2017, Verichev *et al*. 2019, Praene *et al*. 2019), geomorphology (Chai *et al*. 2009, Miliaresis 2013, Jasiewicz *et al*. 2014), landscape (Zhang and Du 2015, Fei *et al*. 2017, Liu *et al*. 2018, 2019, Peng *et al*. 2019, Zhang *et al*. 2020, Du *et al*. 2020), and ecology (Hargrove and Hoffman 2004, Omernik and Griffith 2014, Wang and Pan 2017, Liu *et al*. 2017, Guo *et al*. 2018, Wu *et al*. 2018, Xu *et al*. 2019, Yu *et al*. 2020). Better knowledge of regional differentiation is necessary for tackling global environmental changes and achieving regional sustainable development. To this end, regionalization approach is essential for managing and utilizing natural resources.

In general, ecological regionalization discriminates different spatial patterns of ecosystem functions or ecological processes by separating broad environments into smaller regions (Bailey 1983, McMahon *et al*. 2001, Fu *et al*. 2004). Ecoregion maps are widely adopted as strata for sampling or as covariates for modeling (Kearney *et al*. 2019). Given the growing interactions between natural environment and human activities, ecoregion maps are increasingly needed for environmental monitoring and management (Powers *et al*. 2013, Guo *et al*. 2017, Lin and Li 2019).

Traditionally, expert evaluation involving semi-qualitative analysis has been the primary method to generate ecoregion maps. These include the Canadian Ecological Framework (Ecological Stratification Working Group 1995); Land Resource Regions and the Major Land Resource Areas (MLRAs) produced by the US Department of Agriculture (USDA 1981); and China national eco-geographical map (Zheng *et al*. 2008). Whilst these ecoregion maps might be outdated as climate change and anthropogenic activities have intensively interfered ecological processes and earth system, they are still commonly used. Based on remote sensing or other spatial data, some recent studies employed machine-learning methods to prepare ecoregion maps (Zhou *et al*. 2003, Li and Zheng 2003, Li *et al*. 2008, Guo *et al*. 2018, Wu *et al*. 2018, Xu *et al*. 2019). The commonly applied approaches include classification, clustering, decision tree, etc. Although the machine-learning-based method has the advantage of advanced and automatic/semi-automatic approaches for regionalization, it remains difficult to verify the reliability of the generated ecoregion maps. In most cases, expert knowledge on the spatial differentiation of ecoregions is still required for the verification.

Whilst legacy ecoregion maps might be outdated, they contain the understanding of experts on spatial distribution of ecoregions and relationships between ecoregion units and their environmental factors. The knowledge about 'what environmental conditions each ecoregion units exist' is generalized from experts' extensive field works and remains invaluable. Once the geographic knowledge embedded in ecoregion maps are explicitly extracted and expressed, the knowledge would be useful in two ways. First, the knowledge could facilitate the understanding of ecosystem patterns and the associated ecological processes in a quantitative way (Loveland and Merchant 2004), which would provide references for successive scientists or support the verification of machine-learning-based methods. Second, the extracted knowledge could be used for updating regionalization maps automatically with updated environmental data under global change. To retain the existing knowledge and update regionalization, it is necessary to extract the knowledge embedded in ecoregion maps using robust data mining approaches.

The key objectives of this paper are to extract quantitative knowledge about relationships between ecoregions and their environmental factors from legacy ecoregion maps; and update the regionalization maps using the extracted knowledge. To achieve these

objectives, we utilize an eco-geographical regionalization map of China (Zheng *et al*. 2008) as an example of legacy ecoregion maps for knowledge extraction.

2. The eco-geographical region map of China

Geographical regionalization is considered a classic research topic in integrated physical geography and regional geography (Bailey 1976, Omernik 1987, Zheng and Fu 1998, Zheng 1998). Since the first introduction of the concept of natural regionalization (Herbertson 1905, Dokuchaev 1951), many studies have been conducted to develop different regionalization systems (Köppen 1931, Bailey 1976, Olson *et al*. 2001, Ellis and Ramankutty 2008). In China, the Chinese Academy of Sciences (CAS) organized the 'Natural Regionalization Work Commission' in 1956 to initiate the systematic regionalization research in China (Wu *et al*. 2003b). Notably, Huang (1989) established the integrated natural regionalization which is widely applied in agriculture production management. Zheng *et al*. (2008) developed a hierarchic eco-geographic regional system of China through the division of natural features. This system is one of the most important eco-region systems in China (Wu *et al*. 2003a, Gao *et al*. 2010) which is similar to Bailey (1976).

In Zheng's system (Zheng *et al*. 2008), the highest level of eco-geographical regions focuses on climatic and biological differences, which mainly depicts horizontal zonality. The second-level unit focuses on the differentiation of moisture states, while the third level uses medium-scale topographical factors to identify different units. In this study, we use the highest level of regionalization (i.e. climatic zone; Figure 1) as the target unit for knowledge extraction. There are 11 climatic zones with different sizes (see Figure 1). Most of the climatic zones show clear latitude zonality from north to south, except two climatic zones of the Tibetan Plateau.

According to Zheng *et al*. (2008), several experts applied semi-quantitative analysis on environmental factors to produce the climatic zone map (Figure 1). They performed the analysis based on their understanding about distinguishing different climatic zones in terms of environmental characteristics. Extensive field works and experiences enabled them to plot boundary lines by referring to several maps with environmental variables. Climate variables were the key factors in the boundary delineation process. The specific procedure is as follows.

Firstly, some climate variables (e.g. annual accumulated temperature ≥10°C, average annual temperature of January, and average annual extreme minimum temperature, etc.; period: 1961–1995) were obtained from weather observations of the National Meteorological Administration. Note that the annual accumulated temperature ≥10°C is the sum of daily mean temperatures no less than the minimum biology temperature (10°C). It is an important indicator of thermal conditions in crop ecology. Accordingly, days of temperature ≥10°C is the number of days with daily mean temperatures no less than 10°C in a year. Based on ArcView GIS software, isolines of the climatic variables with a scale of 1:4, 000, 000 were generated. Following experts' knowledge, isoline maps were then used to determine the boundaries of climatic zones. For example, 220 days of temperature ≥10°C and 4500°C of annual accumulated temperature ≥10°C were two main criteria to delineate the boundary line between the temperature zone and the subtropical zone. Besides, national vegetation map of China (1: l, 000, 000) (Hou *et al*. 2001), national soil map of

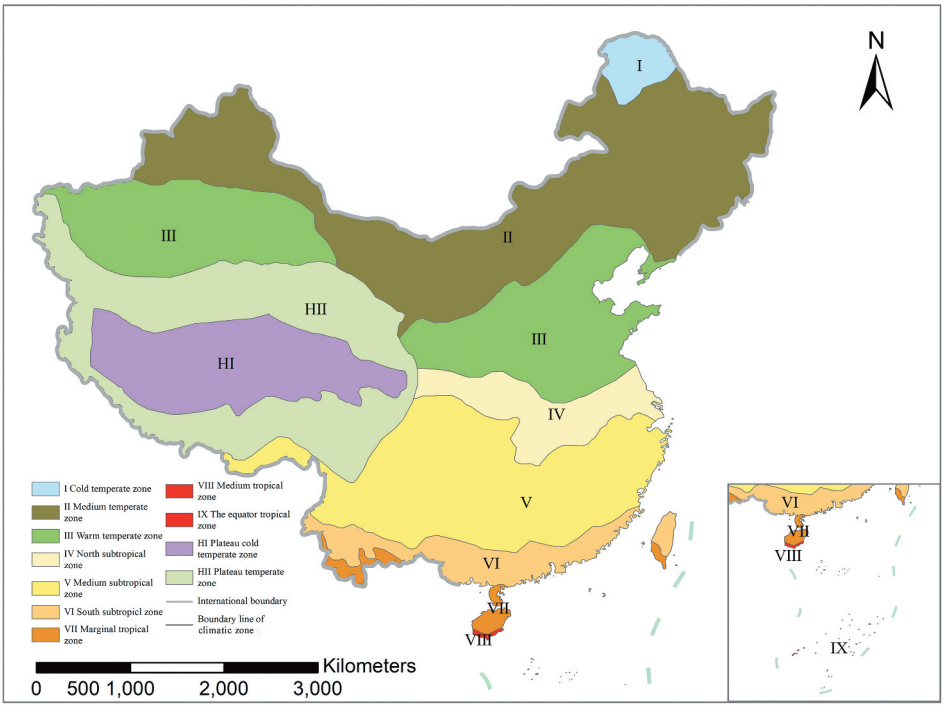

Figure 1. The highest level units (climatic zones) in the eco-geographic regional system of China (Zheng *et al*. 2008). IJGIS remains strictly neutral with respect to jurisdictional claims on disputed territories and the naming conventions used in the maps.

China (l:20, 000, 000) (Editorial Committee of Physical Geography of China et al. 1982) were adopted as supplementary data. Landform was also considered to determine the boundaries of some zones (e.g. the boundary of the plateau temperate zone was determined from the spatial extent of the Tibet plateau). Table 1 shows the applied variables and the criteria for the differentiation of climatic zones generalized by the experts.

In the above process, the boundaries between different zones were determined subjectively in accordance with the integrated knowledge of many experts (Wu *et al*. 2003b). It is difficult to determine boundary lines because the spatial distribution of climate variables used to identify different zones can change gradually (Wu and Zheng 2000, Gao *et al*. 2010). In addition, whilst climate variables were taken as one main criterion, vegetation and soil information were also considered for determining the boundaries. When various variables were integrated to define climatic zones, the judgment of boundaries might be subjective and not reproducible (Wu and Zheng 2000).

3. Methodology

Whilst the determination of boundaries using expert judgment was a subjective process, the uncertainties mainly existed in boundaries between the climatic zones (Wu and Zheng 2000, Gao *et al*. 2010). The majority of areas (pixels) within one unit (climatic zone) could present the characteristics of this specific unit, thus those inner areas (except for the

Table 1. Criteria for the differentiation of climatic zones (Zheng *et al.* 2008).

	Main criteria		Supplementary criteria		
Climatic zones	Days of temperature ≥10°C (day)	Annual accumulated temperature ≥10°C (°C)	Average annual temperature of Jan. (°C)	Average annual temperature of July (°C)	Average annual extremely lowest temperature(°C)
I: Cold temperate zone	<100	<1600	<-30	<16	(<-44)
II: Medium temperate zone	100–170	1600–3200(3400)	−30-(−12)(−6)	16–24	(−44/-25)
III: Warm temperate zone	170–220	3200(3400)-4500(4800)	−12(−6)-0	24–28	−25-(−10)
IV: North subtropical zone	220–240	4500(4800)-5100(5300) 3500–4000 (Yunnan)	0–4 3(5)-6 (Yunnan)	28–30 18-20 (Yunnan)	−14(−10)-(−6)(−4) −6-(−4) (Yunnan)
V: Medium subtropical zone	240–285	5100(5300)-6400(6500) 4000–5000 (Yunnan)	4–10 5(6)-9(10) (Yunnan)	28–30 20-22 (Yunnan)	−5-0 − 4–0 (Yunnan)
VI: South subtropical zone	285–365	6400(6500)-8000 5000–7500 (Yunnan)	10–15 9(10)-13(15) (Yunnan)	28–29 22-24 (Yunnan)	0–5 0-2 (Yunnan)
VII: Marginal tropical zone	365	8000–9000 7500-8000 (Yunnan)	15–18 13-15 (Yunnan)	28–29 > 24 (Yunnan)	5–8 > 2 (Yunnan)
VIII: Medium tropical zone	365	>8000(9000)	18–24	>28	>8
IX: The equator tropical zone	365	>9000	>24	>28	>20
HI: Plateau sub cold temperate zone	<50		−18-(−10)(−12)	6–12	
HII: Plateau temperate zone zone	50–180		−10(−12)-0	12–18	

Note: Yunnan is a province in the southwest of China.

boundary areas) represent the expert knowledge about what environmental conditions the unit exists. Therefore, the basic idea of knowledge extraction in this study is to obtain the relationships between climatic zones and their environmental variables from the climatic zones after reducing the impact of uncertainties of boundaries. The extracted knowledge could be recorded for producing updated ecoregion maps with detailed and updated environmental data.

The method consists of these steps: (i) development of an environmental database, (ii) extraction and quantification of knowledge, (iii) generation of the climatic zone map based on the extracted knowledge, (iv) assessment of the prepared climatic zone map, and (v) updating a climatic zone map for a new time period.

3.1. Development of the environmental database

The first step is to choose and generate environmental variables for extracting knowledge on the ecoregion-environmental variable relationships from the ecoregion map (Figure 1). Based on Table 1, we select five climatic variables including days with temperature ≥10°C, annual accumulated temperature ≥10°C, average annual temperature of January (the coldest month), average annual temperature of July (the warmest month), and average annual extreme minimum temperature (Figure 2). We generate the climatic variables by interpolating the observation data (obtained from weather observation stations of the National Meteorological Administration distributed in every provinces of China; period: 1961–1995). The period (1961–1995) is consistent with the duration of the climatic data used for generating the original eco-region map. We perform the interpolation using the ANUSPLINE software (Hijmans *et al*. 2005) and set the cell size of the interpolated climatic data as 1 km.

To account for different land surfaces and support zone identification, we also use the Digital Elevation Model (DEM) data of China as part of the environmental variables. The original DEM data (spatial resolution: 90 m) were downloaded from the Shuttle Radar Topographic Mission (SRTM) website (http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp) and resampled into a 1 km resolution to be consistent with the climatic variables.

In this study, we do not include the vegetation/soil maps because they are maps of categories which would further complicate the uncertainties of boundary lines.

3.2. Extraction and quantification of knowledge

3.2.1. Reducing the uncertainties of boundary lines

By assuming that the uncertainties mainly exist in transitional areas between two zones, we use a buffer zone approach to reduce the impact of the uncertainties on knowledge extraction. We generate buffer zones for each climatic zone with a specified Euclidean distance perpendicular to the boundary lines, and the original climatic zones without buffer zones are taken as new units for knowledge extraction. We set different distances, including 5 km, 10 km and 15 km. Since the width of climatic zone VIII is no larger than 20 km, a buffer distance of 20 km is not applicable. To evaluate the effect of eliminating the buffer zones from the original climatic zones, we also use the original climatic zones as the units to extract the knowledge for comparison. We apply the tool of Buffer in ArcInfo 10.1 software to generate the buffer zones.

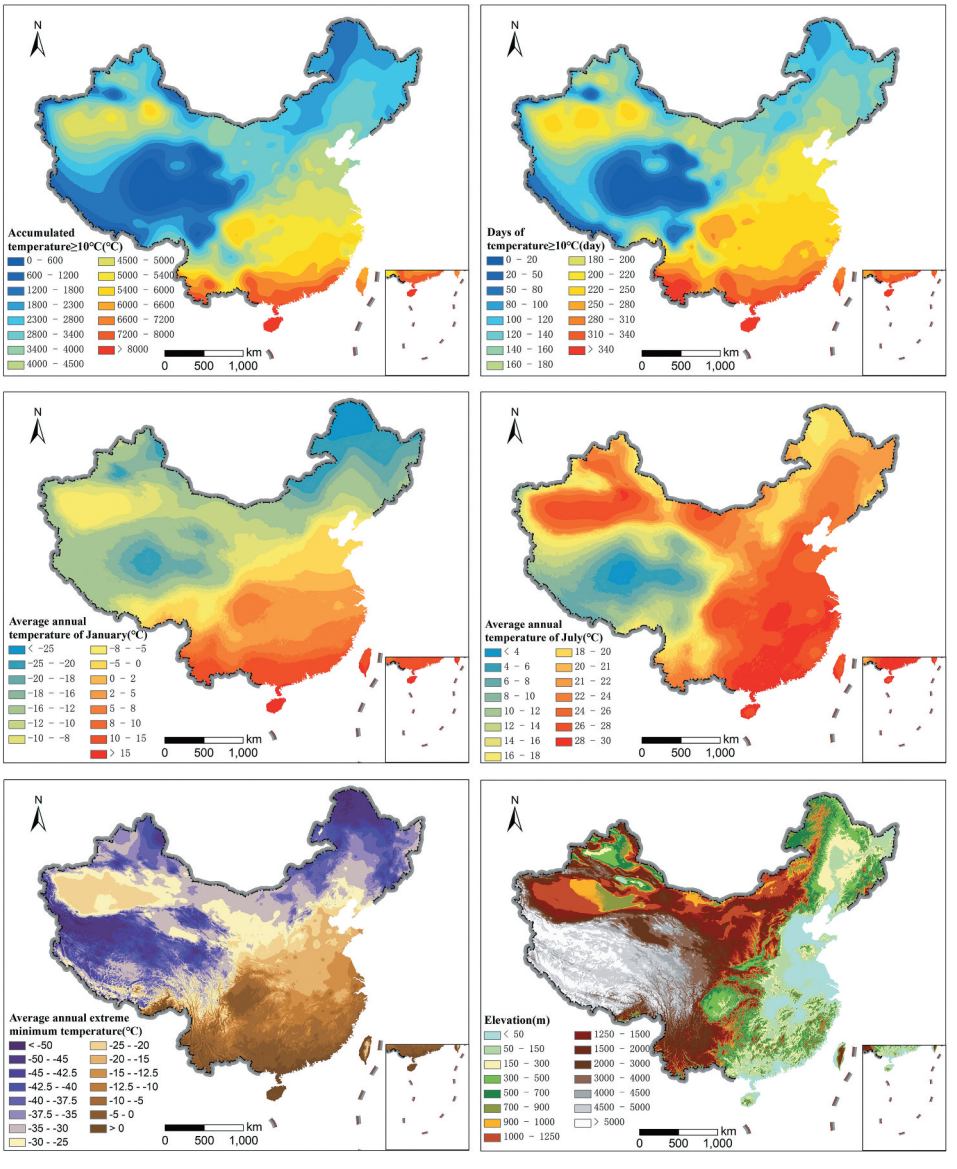

Figure 2. The climatic variables with a period of 1961–1995 year and DEM of China. IJGIS remains strictly neutral with respect to jurisdictional claims on disputed territories and the naming conventions used in the maps included in the figure.

3.2.2. Extraction of knowledge on climatic zone-environment relationships

In this study, we adopt the fuzzy membership function to express knowledge on relationships between climatic zone and environmental variables. A fuzzy membership function can effectively express the knowledge on relationships between continuously distributed and gradually varied geographical elements and their environment. For example, it has been widely used to represent soil–environment relationships (Kaufmann *et al*. 2009,

Beucher *et al*. 2014, Zhu *et al*. 2010, Qi *et al*. 2006, Yang *et al*. 2017, de Gruijter *et al*. 2011), riparian zone–environment relationships (Betz *et al*. 2018), landslide susceptibility–environment relationships (Zhu *et al*. 2018), etc. Furthermore, fuzzy memberships are easy to understand for geographers (Kaufmann *et al*. 2009). A fuzzy membership function describes the grade of membership (similarity) of an individual geographical location to a given class by matching the environmental variable of the individual location with that of the typical case (central concept) in the specific class. For a given class, a fuzzy membership function can be generated using one environmental variable. The fuzzy membership value varies from 0 (which means that the local location is very different from the central concept of a given class) to 1 (which means that the local location is identical to the central concept of a given class). The shape of the membership function curve approximates the manner in which the similarity changes as the environmental value varies.

Several methods have been developed to construct fuzzy membership functions. One approach relies on experts' knowledge (Zhu 1997, Qi *et al*. 2006, Kaufmann *et al*. 2009). Other methods construct fuzzy membership functions based on legacy maps, using an expectation maximization algorithm (Qi *et al*. 2008) or a histogram-based method (Medasani *et al*. 1998). Similar to the histogram-based method, which transforms the frequency information of environmental variables within each soil class to fuzzy memberships, we transform probability density functions (PDFs) of environmental variables to fuzzy membership functions in this study. For each climatic zone, the probability density functions (PDFs) of an environmental variable depict the probability distribution (the likelihood of an outcome) of the environmental variable.

The knowledge extraction procedure is composed of the following three steps. The first step is to generate PDFs of each environmental variable for all climatic zones. If there are L climatic zones with M environmental variables, a total of $L*M$ membership functions would be generated. In our study, we estimate the PDFs using Kernel Density Estimation (KDE), which is a density estimation method capable of estimating continuous PDFs (Silverman 1986, Sheather 2004). The KDE method can be expressed as:

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h_x}\right) \tag{1}$$

where $f(x)$ is the estimated PDF for the environmental variable x, x_i is the value of x at pixel i, n is the total number of pixels, K is a kernel density function, and h_x is the bandwidth. We adopt Gaussian kernel for K, and 'the rule of thumb' algorithm to determine h_x for each variable in the equation (Sheather and Jones 1991):

$$K\left(\frac{x - x_i}{h_x}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h_x{}^2}}, h_x = \sigma_x \left(\frac{4}{3n}\right)^{1/5} \tag{2}$$

where σ_x is the standard deviation of the respective environmental variables for the pixels.

Each of the climatic zones III, VI, and VII includes two separate sub-regions. For each climatic zone, the two separate sub-regions cover a vast area and represent different climatic and topographic characteristics (see Figure 2). Therefore, we name each of the two sub-regions as, III-1 (the west sub-region of III, located in the northwest of China) and III-2 (the east sub-region of III, located in the middle east of China), VI-1 (the sub-region of

VI in mainland of China) and VI-2 (the sub-region of VI in Taiwan), VII-1 (the sub-region of VII in Hainan Island) and VII-2 (the sub-region of VII in Taiwan) for generating PDFs. As such, the original 11 climatic zones turn into 14 zones. In this study, we use all pixels in the original 14 climatic zones and the 14 climatic zones without the buffer zones to generate PDFs of environmental variables for each climatic zone. To do this, we employ the 'stats' package in the R statistical language to generate the PDFs.

The second step involves normalizing the generated PDF curves using Eq. (3). The purpose of normalization is to transform the range of the density values of all PDFs into 0–1, which is consistent with the range of membership values for the constructed fuzzy membership function.

$$S_i^x = \frac{f(x_i)}{\max(f(x))} \tag{3}$$

where x_i refers to the value of environment variable x at location i, max($f(x)$) refers to the maximum value of the PDF, and S^x_i refers to the normalized probability density of environment variable x at location i.

Finally, several characteristic values can be generated from normalized PDF (fuzzy membership) curves to quantitatively represent the characteristics of each climatic zone. Specifically, when a curve contains only one peak, the minimum and maximum of the environmental variable values of the curve, the environmental value of the peak, the environmental values when the fuzzy memberships equaled to 0.5 at both sides were derived. When a curve exhibits two or more peaks, then the above characteristic values are derived for each sub-curve.

3.3. Generation of the climatic zone maps based on the extracted knowledge

We generate new climatic zone maps for 1961–1995 using the above-extracted knowledge from climatic zones without different buffer zones. The procedure includes three steps (Zhu and Band 1994, Zhu 1997): i) determine the membership of a location to each climatic zone based on their fuzzy membership functions of every single environmental variable, ii) determine the final membership to each climatic zone by considering all environmental variables involved, and iii) generate a climatic zone map based on the final membership to each climatic zone.

In the first step, we use the fuzzy membership functions (normalized PDFs) to obtain the membership of a location to each climatic zone based on every single environment variable in accordance with the environmental variable value of the location. For each climatic zone, if there are M environmental variables, then M memberships associated with the M variables are generated for the location.

In the second step, the M memberships generated in the first step for a location are integrated into a final membership for each climatic zone. There are different ways for the integration when considering the influences of environmental variables, e.g. weighted average (McBratney *et al.* 2003), limiting factor method (Zhu and Band 1994), etc. The weighted average strategy assigns each variable with the same or different significances (weights). The limiting factor approach (Zhu and Band 1994), which assumes that the least favorite environmental variable determines the assigned class at a given location, takes the minimum membership value among the membership values of environment variables. In the current

study, the boundaries of climatic zones are controlled by multiple environment variables, but they are mostly determined by the most limiting factor. Thus, we use a minimum operator on memberships for all the environmental variables to generate the final membership of a location for each climatic zone. The outcome of this step is a membership vector M_i at location i, $(M_{i,1}, M_{i,2}, \ldots, M_{i,l})$, in which elements are the environmental memberships of location i for the l climatic zones.

The last step predicts the climatic zone for all locations based on the environmental membership vectors. We conduct a hardening of the environmental memberships generated in this step. In particular, we assign the climatic zone with the maximum membership in vector M_i to location i (Zhu 1997). By assigning the climatic zone for each location, we can prepare a harden climatic zone map.

We add the fuzzy membership functions and environmental variables into the SoLIM solutions software (http://solim.geography.wisc.edu/software/index.htm) to perform the above procedure. Note that programming or other tools that can work with the fuzzy membership functions and environmental data can also realize the prediction of climatic zone maps.

After generating the predicted climatic zone maps using the extracted knowledge, we compare these maps with the original climatic zone map to determine a suitable buffer zone.

3.4. Assessment of the generated climatic zone map based on the extracted knowledge

Typically, it is challenging to quantitatively evaluate the accuracy of regionalization maps (Hargrove and Hoffman 2004). Regionalization reflects a spatial stratification of heterogeneity, which means that locations are more homogeneous within strata than locations among strata (Wang *et al*. 2016). By examining whether (or how much) the environmental variables of pixels in the same climatic zone are similar or different from that of pixels in other climatic zones, we could evaluate the efficiency of a regionalization map for identifying different climatic zones. To evaluate the generated climatic zone map and the original map, we adopt a measure of spatial stratification heterogeneity, the q-statistics. The q index measures the difference among strata versus the similarities within strata by comparing the within-strata variance with between strata-variance (Wang et al. 2015). Earlier studies confirmed its efficiency in evaluating the heterogeneity of spatial patterns of $PM_{2.5}$ concentrations (Chen *et al*. 2019), health status (Li *et al*. 2019), and water-use efficiency (Liu *et al*. 2020), etc.

Since the regionalization for this research is for identifying pixels with similar spatial variations of environmental variables, we calculate the q index for each environmental variable using Eq. (4); and quantify the average q index used to assess the regionalization results using Eq. (5).

$$q_m = 1 - \frac{1}{N\delta_m^2} \sum_{l=1}^{L} N_l \delta_{lm}^2 \tag{4}$$

where q_m is the q value of the environmental variable m, N is the number of all pixels in the study area, L is the number of climatic zones, l is the ID of each climatic zone, N_l is the

number of pixels in climatic zone l, δ_m^2 is the total variance of environmental variable m for all pixels, δ_{lm}^2 is the variance of environmental variable m of pixels in climatic zone l.

$$q_{avg} = \frac{1}{M}\sum_{m=1}^{M} q_m \tag{5}$$

where q_{avg} is the average of q for all the environmental variables, M is the number of environmental variables.

We note that q_m falls into [0, 1], 0 if the environmental variable m in different climatic zones has no difference, 1 if the environmental variable m is the same within each climatic zone and are completely distinct from other climatic zones. A large q_{avg} would indicate an efficient regional differentiation of environmental variables across China.

3.5. Updating the climatic zone map of China for a period of 1961-2015

The boundaries between different climatic zones are not static and can vary with climatic variability. The original climatic zone map was produced using the climatic data over the period 1961–1995. Since our climate has changed over the past few decades, we prepare a new map for the period 1961–2015 using the climatic data with this time period and the extracted knowledge. The new map could provide useful references for climate adaptation. We apply the SoLIM solutions software (http://solim.geography.wisc.edu/software/index.htm) to produce the new map.

4. Results

4.1. The PDFs generated from the climatic zone map

Figure 3 displays the probability density curves of the environmental variables for all the climatic zones. In general, most of the adjacent curves contain overlaps, indicating that there are obvious transitions of environmental variables between adjacent climatic zones. Some of the probability density curves exhibit two or three peaks (e.g. all curves for climatic zone II, and most curves for climatic zone HI), except for DEM and average annual extreme minimum temperature. For most climatic zones, PDFs of average annual temperature of January, average annual extreme minimum temperature, days of temperature ≥10°C, or annual accumulated temperature ≥10°C show large differences. We observe more overlaps for PDFs of average annual temperature of July and DEM for the adjacent climatic zones. Since it is possible to identify different climatic zones from the PDFs of different environmental variables, combining multiple environmental variables could realize the differentiation of most climatic zones. We notice that the distribution characteristics of the probability density curves of days of temperature ≥10°C and annual accumulated temperature ≥10°C for most climatic zones were similar. This is mostly because days of temperature ≥10°C and annual accumulated temperature ≥10°C experience very similar spatial distribution patterns. This reflects that the two variables contain similar indications for identifying climatic zones. In this sense, there is no need to use both variables simultaneously to generate the updated climatic zone maps. In addition, the

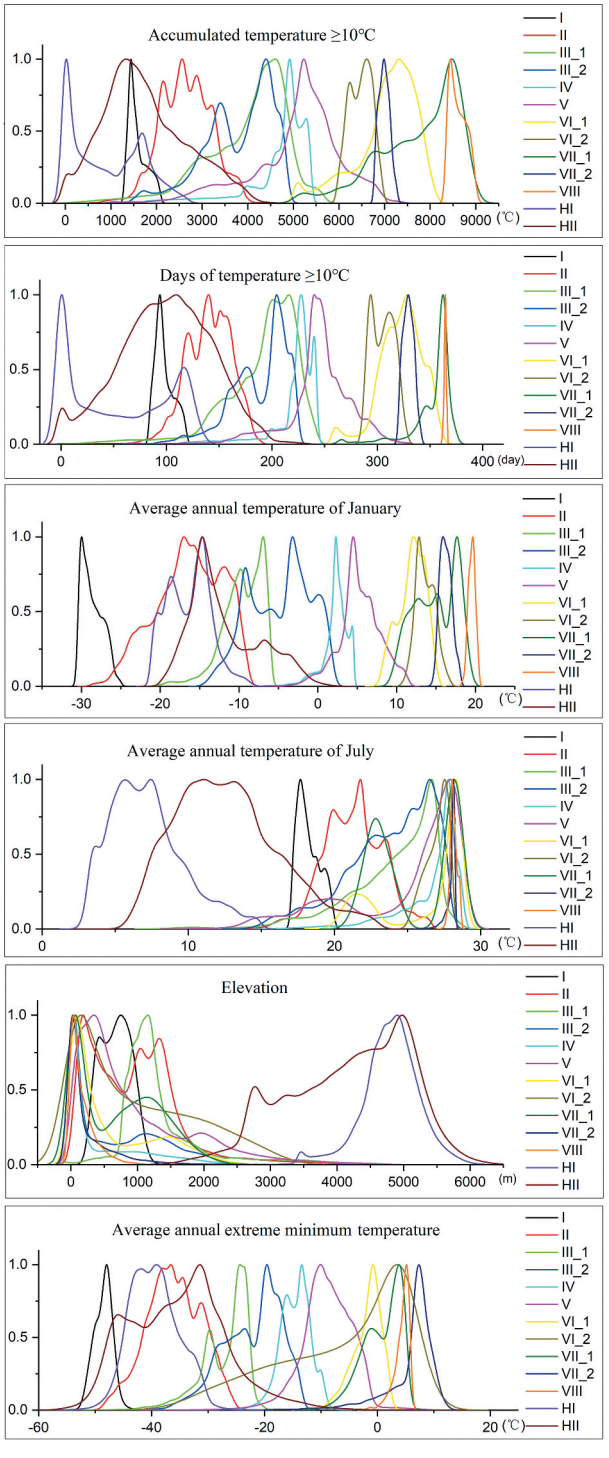

Figure 3. The probability density curves of environmental variables for all climatic zones.

Table 2. The characteristic values extracted from the probability density curves of annual accumulated temperature ≥10°C for all climatic zones.

Zone ID	Minimum	Maximum	$P1X_{peak}$	$P1X_{0.5\text{-left}}$	$P1X_{0.5\text{-right}}$	$P2X_{peak}$	$P2X_{0.5\text{-left}}$	$P2X_{0.5\text{-right}}$
I	1230.0	2142.0	1440.5	1352.3	1578.8			
II	800.4	4097.2	2151.9–2874.9	1953.9	3115.9			
III-1	339.3	5815.1	4599.7	3875.3	4943.4			
III-2	1380.3	5027.7	3401.5	2944.6	3765.4	4400.4	3974.5	4779.9
IV	1717.3	5492.4	4915.9	4773.7	5129.1	5279.2	5129.1	5381.9
V	1051.3	7370.1	5227.2	4883.8	5776.7			
VI-1	4882.2	8258.4	7318.6	6761.9	7846.3			
VI-2	5867.8	7051.7	6238.8	6075.7	6367.7	6607.1	6367.7	6830.7
VII-1	4842.8	9247.0	8479.4	7846.1	8796.0			
VII-2	6710.5	7334.4	6985.8	6806.5	7143.9			
VIII	8231.9	9103.1	8439.0	8350.0	8859.2			
HI	0.0	2781.0	21.2	0.0	181.6	1683.5	1259.9	1908.1
HII	0.0	4348.1	1335.8	658.5	2316.3			

Note: $P1X_{peak}$ is the environmental value (annual accumulated temperature ≥10°C in this table) at the curve peak for the only curve or the first sub-curve, $P1X_{0.5\text{-left}}$ and $P1X_{0.5\text{-right}}$ are the environmental values when the fuzzy memberships equate 0.5 at left and right sides, respectively for the only curve or the first sub-curve. And $P2X_{peak}$, $P2X_{0.5\text{-left}}$ and $P2X_{0.5\text{-right}}$ are for the second sub-curve.

PDF curves of the two sub-regions are different for climatic zones III, VI and VII, indicating the necessity of dealing with the two sub-regions separately.

4.2. Extraction of characteristic values from the probability density curves

We extract the characteristic values from the probability density curves and present the characteristic values for annual accumulated temperature ≥10°C (as example) in Table 2.

4.3. The generated climatic zone maps and its comparisons with the original climatic zone map

By transforming the PDFs to fuzzy membership function curves through normalization, we can generate climatic zone maps based on the fuzzy membership function curves (see Figure 4). Whilst the spatial distributions of the predicted climatic zones are generally similar to the original climatic zone map, there are still some differences between the two maps. First, the predicted climatic zones HI and IV are smaller than the original zones. This is primarily because the environmental condition of HI has a large overlap with that of HII but HI has a smaller area size and narrower fuzzy membership curves compared to HII. When predicted the climatic zones according to their fuzzy membership functions, the area of the predicted HI would be smaller meanwhile the predicted HII would be larger. The same reason was for a smaller area of the predicted IV compared with its original distribution. The other major difference is that the predicted II contains smaller parts of III. This is because the environmental characteristics of some areas inside II are closer to those of III. Regarding the regional integrity of the climatic zone, these small parts of III inside II should be labeled as II in the post-processing step.

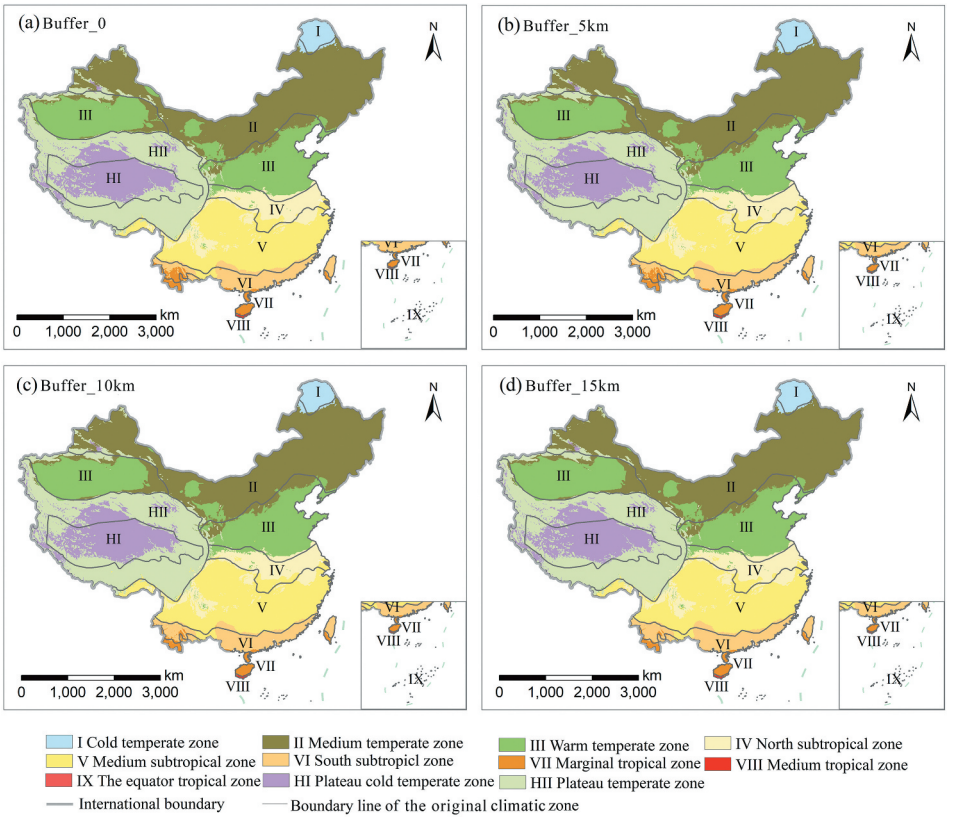

Figure 4. The predicted climatic zone maps based on the extracted fuzzy membership functions, the grey lines in the maps are the original boundary lines in Figure 1. IJGIS remains strictly neutral with respect to jurisdictional claims on disputed territories and the naming conventions used in the maps.

Although the maps generated using the extracted knowledge from the original climatic zones and zones without different buffer zones exhibit very similar spatial patterns, they are not completely indistinguishable. According to the boundary lines of VI and VII and area sizes of the predicted III inside II, we recommend a buffer zone with 10–15 km.

Based on Arcinfo 10.1 software, we eliminate the small polygons of the generated map with a 10 km buffer and present the final climatic zone map in Figure 5. The key difference between this map and the original map is that the area of the predicted zone HI is smaller. This is because of the big overlap of environmental distributions for climatic zones HI and HII. Besides, the smaller area of the predicted zone IV is a result of its relatively narrow fuzzy membership curve compared to those of IV and V.

4.4. The calculated q indices of the generated and the original climatic zone maps

Table 3 displays the calculated q values for the original climatic zone map and the generated climatic zone map of Figure 5. Apparently, the q value of the generated climatic zone map in our study is higher than that of the original climatic zone map for all the

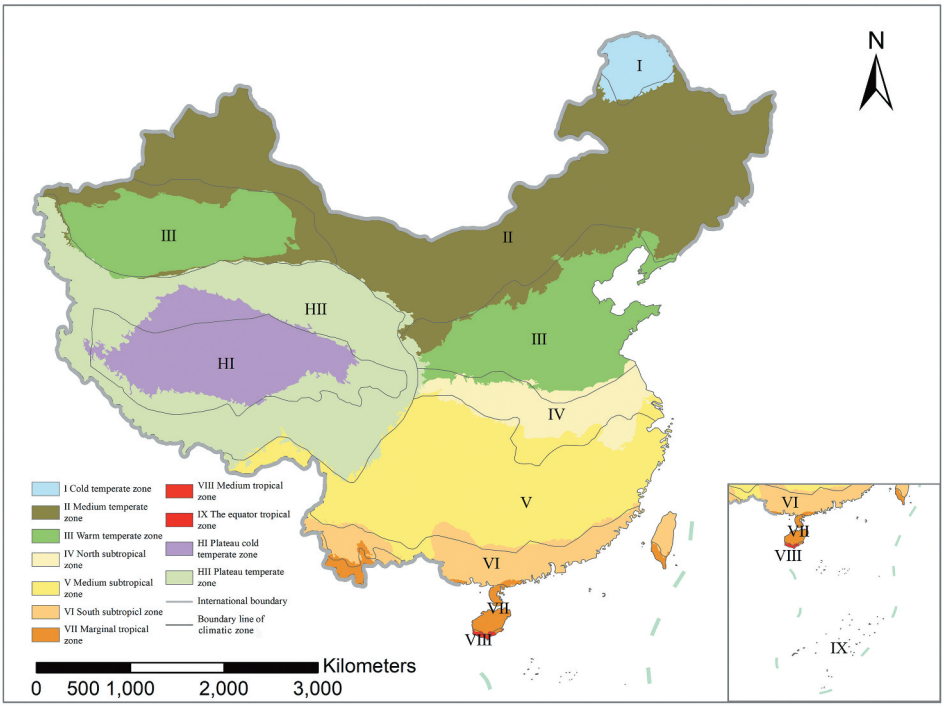

Figure 5. The generated climatic zone map after post-processing. The grey lines in the maps are the original boundary lines in Figure 1. IJGIS remains strictly neutral with respect to jurisdictional claims on disputed territories and the naming conventions used in the maps.

environmental variables. The average *q* value of the generated climatic zone map is also higher. This indicates a better spatial-stratified heterogeneity of the generated climatic zone map of our study compared to the original climatic zone map.

4.5. The updated climatic zone map using climate data of 1961-2015

Based on the extracted knowledge, climate data of the period 1961–2015 and post-processing procedure, we present an updated climate zone as shown in Figure 6. By comparing this map with the original map (for the period 1961–1995), we observe several

Table 3. The calculated *q* values for the original climatic zone map and the generated climatic zone map.

Environmental variables	The original climatic zone map	The generated climatic zone map of our study
DEM	0.820	0.835
Annual accumulated temperature ≥10°C	0.804	0.833
Days of temperature ≥ 10°C	0.816	0.836
Average annual temperature of January	0.869	0.897
Average annual temperature of July	0.802	0.814
Average annual extremely lowest temperature	0.833	0.860
The average *q* value	0.824	0.846

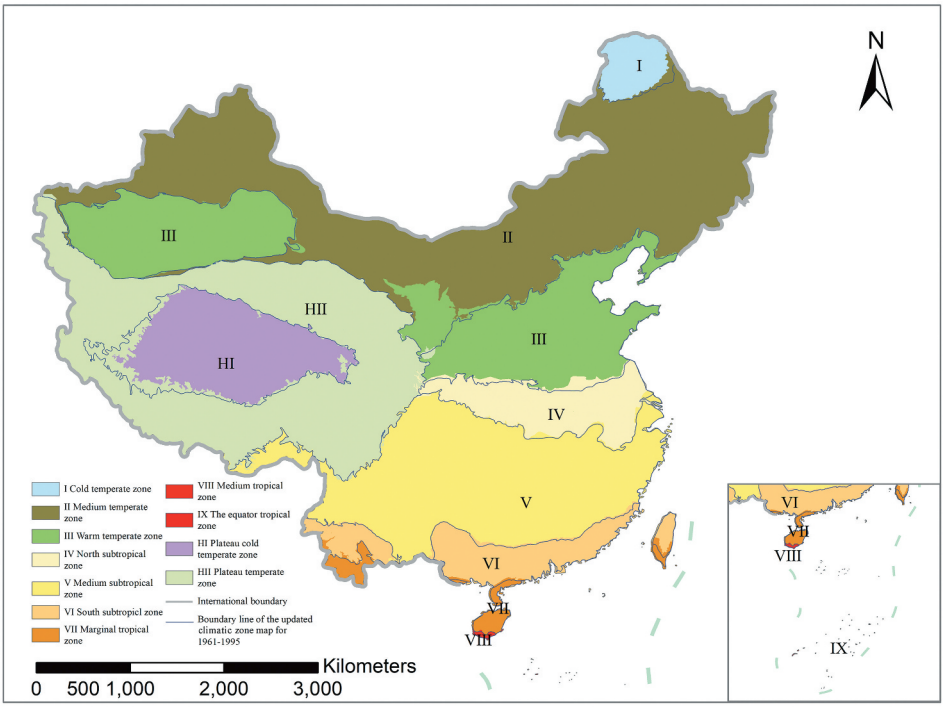

Figure 6. The climatic zone map for the period 1961–2015. The dark blue lines in the maps are the boundary lines of the generated climatic zone map of Figure 5. IJGIS remains strictly neutral with respect to jurisdictional claims on disputed territories and the naming conventions used in the maps.

differences. First, boundaries of most of the climatic zones in the eastern China shifted to the north to some extent, and the warm temperate zone expanded. Second, Plateau temperate zone expanded a little and its boundary shifted to the east, while Plateau cold temperate zone shrank and slightly shifted to the north.

5. Discussions

5.1. Comparisons between the extracted knowledge with the criteria constructed by experts

When comparing the knowledge extracted from the climatic zone map, including the normalized probability density curves and the characteristic values (Table 2) with the knowledge generalized by experts (Table 1), we could recognize some obvious differences (see below).

First, the knowledge extracted from the original climatic zone map, especially the probability density curves, provided more information than the expert knowledge (refer to Table 1). Whilst there were only ranges of climatic variables for each climatic zone in Table 1, more detailed information could be generated from the probability density curves in addition to the ranges, including the climatic variable values of peaks, the climatic variable values when memberships equal to 0.5, number of peaks, etc (see Table 2).

Second, the extracted characteristic values and criteria in Table 1 were different for most climatic zones. There was almost no overlap between ranges of the climatic variables for the climatic zones in Table 1. Nonetheless, the overlaps of the distribution of climatic variables for different climatic zones were clear in Figure 3. The ranges of climatic variables for most climatic zones in Tables 1 and 2 were different. Moreover, the climatic variable values of peaks were different from the middle values of the ranges (which can be taken as the central concept or typical climatic conditions for certain climatic zone) in Table 1. Part of the reason for these differences (above) is because experts highlighted the differences between climatic zones when generalizing the criteria in Table 1. Thus, the ranges of variables for climatic zones were in order with hardly overlapping values. Furthermore, the generalization of knowledge was partly due to the limitations of technologies and environmental data resolution back then. Although the expert-generalized knowledge captured the main characteristics of different climatic zones, the generated characteristics were not comprehensive enough.

According to the spatial distributions of the environmental variables within each climatic zone (Figure 1), the changing of environmental variables from one climatic zone to its spatially adjacent zones was not abrupt but gradual. This is consistent with the natural environmental characteristics of climatic zones. With only 11 climatic zones in China, nearly each climatic zone occupied a large area size. With large areas of some climatic zones, their probability density curves possessed two or multiple peaks (such as climatic zone II and III-2), indicating that several aggregations existed over space within those climatic zones. This cannot be seen from the expert knowledge (displayed in Table 1) due to the generalized expression. In this sense, we deduce that the knowledge extracted from the climatic zone map is more reasonable and usable.

5.2. The applicability of the knowledge extraction method

Expert knowledge is often neither precisely defined nor well formulated (Zhu 1999). This study provided a tool to extract the knowledge hidden in legacy maps and expressed the knowledge in membership functions. One advantage of our method (Section 3) is that the extracted knowledge is quantified and structured, and can be used to store expert knowledge and update dynamic ecoregion maps. A portion of expertise is lost each year as experienced domain scientists retire. It is desirable to retain this expertise to maintain knowledge continuity between different generations of scientists. In this study, we explicitly preserved the extracted knowledge and presented it as an important source for new generations of scientists to build their cognition models. This would shorten the time for new scientists to understand terrestrial systems, and increase the consistency of ecoregion–environment models between generations of scientists.

Another advantage of our method is that researchers can adjust the extracted fuzzy member functions once they obtain a better understanding of land surface differentiation through more field works or studies. Then, an improved climatic zone map could be generated.

The proposed method to extract quantitative knowledge is useful for knowledge-driven regionalization methods, and also applicable to the machine-learning-based regionalization maps. When using machine-learning methods to generate a regionalization map, people are usually required to interpret the underlying characteristics of the generated

ecoregions. In this case, our proposed method can still be employed to extract and represent the relationships between ecoregions and their environmental factors. The fuzzy membership functions for knowledge representation could serve as a bridge for connecting knowledge and the modern regionalization methods.

One challenge in applying data mining algorithms to knowledge discovery from legacy maps is that the existing errors in those maps would affect the accuracy of the extracted knowledge. Previous research (Qi and Zhu 2003, Zhang and Zhu 2018) showed that data preprocessing could play an important role in knowledge discovery. In this paper, we assumed possible uncertainties in transitional areas and used a buffer zone approach to reduce those uncertainties in boundaries between climatic zones on the map. When transferring our method to other areas, the optimal buffer distance can be determined based on context-related experiments and field surveys.

The generated climatic zone map with the period 1961–2015 (Figure 6) is consistent with the results of Bian *et al*. (2013), and Wu *et al*. (2016). The phenology change in China (Bai *et al*. 2011, Wang *et al*. 2012, Dai *et al*. 2013, Chen *et al*. 2015) also proved the northward shift of warm temperate zone and north tropical. This indicates that the extracted knowledge can be used to update climatic zone maps.

6. Conclusions

This study proposed a method to extract knowledge from the legacy regionalization map and presented the knowledge as a set of fuzzy membership functions (curves). The extracted knowledge showed more details than the legacy descriptive knowledge generalized by experts. We used the extracted knowledge to generate a new climatic zone map automatically. By comparing the spatial distribution of the newly generated climatic zone map to the original map, we confirmed the effectiveness of the proposed method. The results showed the effectiveness of the buffer zone approach for reducing the uncertainties of boundaries between climatic zones. We recommended a buffer distance of 10–15 km for the case study. Based on the extracted knowledge and updated climate data (1961–2015), we prepared an updated climatic zone map.

More generally, this study presented a knowledge extraction method for understanding the spatial heterogeneity of land surface elements. The knowledge extraction method can be easily adapted for other applications with an aim to obtain knowledge from area-class maps. The extracted knowledge can also be used to produce dynamic zoning maps with updated environmental data, which can play an important role in climate change and ecological modeling research.

Data and codes availability statement

The data and codes that support the findings of this study are available in ['figshare.com'] with the identifier at the public link (https://doi.org/10.6084/m9.figshare.c.4833234).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China [41530749,41971054]; Open grant from Key Laboratory of Land Surface Pattern and Simulation, CAS [LBKF201506].

Notes on contributors

Lin Yang received the Ph.D degree in Cartography and Geographical Information System from Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. She is currently an associate professor in School of Geography and Ocean Science, Nanjing University. Her research interests focus on spatial sampling, spatial prediction, and data mining.

Xinming Li received the Master's degree in Cartography and Geography Information System from University of Chinese Academy of Sciences. His research interests are automated geoscience calculation and spatial analysis.

Qinye Yang is currently a researcher in the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests focus on regionalization.

Lei Zhang is currently a Ph.D. candidate at Nanjing University. His research interests include GIScience, spatio-temporal analytics, machine learning, and spatial predictive mapping.

Shujie Zhang is currently a senior engineer at China Academy of Urban Planning & Design, and her interests are the application of big data mining and spatial analysis in urban planning.

Shaohong Wu is currently a professor in the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests include climate change impact and adaptation, natural disaster risk, land surface pattern and process, and regionalization.

Chenghu Zhou is an academician of the Chinese Academy of Sciences, and currently a professor in the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests are broadly situated in research on the application of GIS and remote sensing.

ORCID

Lei Zhang http://orcid.org/0000-0002-1090-6338
Shaohong Wu http://orcid.org/0000-0003-3011-4685

References

Bai, J., Ge, Q.S., and Dao, J.H., 2011. The response of first flowering dates to abrupt climate change in Beijing. *Advances in Atmospheric Sciences*, 28 (3), 564–572. doi:10.1007/s00376-010-9219-8.

Bailey, R.G., 1976. *Ecoregions of the United States (1: 7,500,000-scale map)*. Ogden, Utah: Intermountain Region, USDA Forest Service.

Bailey, R.G., 1983. Delineation of ecosystem regions. *Environmental Management*, 7 (4), 365–373. doi:10.1007/BF01866919.

Betz, F., Lauermann, M., and Cyffka, B., 2018. Delineation of the riparian zone in data-scarce regions using fuzzy membership functions: an evaluation based on the case of the Naryn River in Kyrgyzstan. *Geomorphology*, 306, 170–181. doi:10.1016/j.geomorph.2018.01.024

Beucher, A., *et al*., 2014. Fuzzy logic for acid sulfate soil mapping: application to the southern part of the Finnish coastal areas. *Geoderma*, 226–227, 21–30. doi:10.1016/j.geoderma.2014.03.004

Bian, J.J., *et al*., 2013. The shift on boundary of climate regionalization in China from 1951 to 2010. *Geographical Research*, 32 (7), 1179–1187.

Bonfante, A., *et al*., 2018. A dynamic viticultural zoning to explore the resilience of terroir concept under climate change. *Science of the Total Environment*, .624, 294–308. doi:10.1016/j.scitotenv.2017.12.035

Chai, H.X., *et al*., 2009. Digital regionalization of geomorphology in Xinjiang. *Journal of Geographical Sciences*, 19 (5), 600–614. doi:10.1007/s11442-009-0600-4.

Chen, X.Q., *et al*., 2015. Spatiotemporal response of Salix matsudana's phenophases to climate change in China's temperate zone. *Acta Ecologica Sinica*, 35, 3625–3635. in Chinese.

Chen, Z.Y., *et al*., 2019. Spatial self-aggregation effects and national division of city-level PM2.5 concentrations in China based on spatio-temporal clustering. *Journal of Cleaner Production*, 207 (10), 875–881. doi:10.1016/j.jclepro.2018.10.080.

Chen, Z.Y., *et al*., 2020. Influence of meteorological conditions on PM2. 5 concentrations across China: A review of methodology and mechanism. *Environment International*, 139, 105558. doi:10.1016/j.envint.2020.105558

Dai, J.H., Wang, H.J., and Ge, Q.S., 2013. Changes of spring frost risks during the flowering period of woody plants in temperate monsoon area of China over the past 50 years. *Di Li Xue Bao / Chung-kuo ti li hsueh hui pien chi*, 68 in Chinese, 593–601.

Dokuchaev, V.V., 1951. *On the theory of natural zones//Sochineniya (Collected Works)*. Moscow-Leningrad: Academy of Sciences of the USSR.

Du, S.J., *et al*., 2020. Large- scale urban functional zone mapping by integrating remote sensing images and open social data. *GIScience & Remote Sensing*, 57 (3), 411–430. doi:10.1080/15481603.2020.1724707.

Ecological Stratification Working Group, 1995. *A National Ecological Framework for Canada*. Ottawa, ON: Agriculture and Agri-Food Canada, Research Branch, Centre for Lake and Biological Research, and Environment Canada, State of the Environment Directorate, Ecozone Analysis Branch.

Editorial Committee of Physical Geography of China of the Chinese Academy of Sciences, 1982. *Physical geography of china: pedological Geography*. Beijing: Science Press.

Ellis, E.C. and Ramankutty, N., 2008. Putting people in the map: anthropogenicbiomes of the world. *Frontiers in Ecology and the Environment*, 6 (8), 439–447. doi:10.1890/070062.

Fan, J. and Li, P.X., 2009. The scientific foundation of major function oriented zoning in China. *Journal of Geographical Sciences*, 19 (5), 515–531. doi:10.1007/s11442-009-0515-0.

Fei, D.Q., *et al*., 2017. Land use zoning using a coupled gridding-self-organizing feature maps method: A case study in China. *Journal of Cleaner Production*, 161, 1162–1170. doi:10.1016/j.jclepro.2017.05.028

Fu, B.J., *et al*., 2001. Scheme of ecological regionalization in China. *Acta Ecologica Sinica*, 21 (1), 1–6. [in Chinese].

Fu, B.J., *et al*., 2004. Ecoregions and ecosystem management in China. *International Journal of Sustainable Development & World Ecology*, 11, 397–409. doi:10.1080/13504500409469842

Gao, J., Li, S., and Zhao, Z., 2010. Validating the demarcation of eco-geographical regions: a geostatistical application. *Environmental Earth Sciences*, .59, 1327–1336. doi:10.1007/s12665-009-0120-7

Gelcer, E., *et al*., 2018. Influence of El Niño-Southern oscillation (ENSO) on agroclimatic zoning for tomato in Mozambique. *Agricultural and Forest Meteorology*, 248, 316–328. doi:10.1016/j.agrformet.2017.10.002

Gruijter, J.J., Walvoort, D.J.J., and Bragato, G., 2011. Application of fuzzy logic to Boolean models for digital soil assessment. *Geoderma*, 166, 15–33. doi:10.1016/j.geoderma.2011.06.003

Guo, X.D., *et al*., 2017. Regional mapping of vegetation structure for biodiversity monitoring using airborne lidar data. *Ecological Informatics*, 38, 50–61. doi:10.1016/j.ecoinf.2017.01.005

Guo, X.D., *et al*., 2018. Multi-dimensional eco-land classification and management for implementing the ecological redline policy in China. *Land Use Policy*, .74, 15–31. doi:10.1016/j.landusepol.2017.09.033

Hargrove, W.W. and Hoffman, F.M., 2004. Potential for multivariate quantitative methods for delineation and visualization of ecore- gions. *Environmental Management*, 34 (Suppl. 1), S39–S60. doi:10.1007/s00267-003-1084-0.

Herbertson, A.J., 1905. The major natural regions: an essay in systematic geography. *The Geographical Journal*, 25 (3), 300–310. doi:10.2307/1776338.

Hijmans, R.J., *et al*., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. doi:10.1002/joc.1276

Hou, X.Y., *et al*., 2001. *Vegetation Map of China (1:1, 000, 000)*. Beijing: Science Press.

Huang, B.W., 1989. Map of China's comprehensive natural regionalization. *In*: Changchun Institute of geography, Chinese Academy of Sciences, ed. *Atlas of natural conservation of China*. Vols. 20–21. Beijing: Science Press, 155–157. in Chinese.

Jasiewicz, J., Netzel, P., and Stepinski, T.F., 2014. Landscape similarity, retrieval, and machine mapping of physiographic units. *Geomorphology*, 221, 104–112. doi:10.1016/j.geomorph.2014.06.011

Jiang, Y., *et al*., 2018. A geogrid-based framework of agricultural zoning for planning and management of water & land resources: A case study of northwest arid region of China. *Ecological Indicators*, 89, 874–879. doi:10.1016/j.ecolind.2017.12.022

Kaufmann, M., Tobias, S., and Schulin, R., 2009. Quality evaluation of restored soils with a fuzzy logic expert system. *Geoderma*, 151, 290–302. doi:10.1016/j.geoderma.2009.04.018

Kearney, S.P., *et al*., 2019. EcoAnthromes of Alberta: an example of disturbance-informed ecological regionalization using remote sensing. *Journal of Environmental Management*, 234, 297–310. doi:10.1016/j.jenvman.2018.12.076

Köppen, W., 1931. *Grundriss der Klimakunde*. Berlin: Walter de Gruyter.

Li, J., *et al*., 2019. Regional differences and spatial patterns of health status of the member states in the "Belt and Road" initiative. *PLoS ONE*, 14 (1), e0211264. doi:10.1371/journal.pone.0211264.

Li, S.C., Zhao, Z.Q., and Gao, J.B., 2008. Identifying eco-geographical boundary using spatial wavelet transform. *Acta Ecologica Sinica*, 28 (9), 4313–4322.

Li, S.C. and Zheng, D., 2003. Applications of artificial neural networks to geosciences: review and prospect. *Advance in Earth Science*, 18 (1), 68–76.

Lin, J.Y. and Li, X., 2019. Large-scale ecological red line planning in urban agglomerations using a semi-automatic intelligent zoning method. *Sustainable Cities and Society*, 46, 101410. doi:10.1016/j.scs.2018.12.038

Liu, K.D., Yang, G.L., and Yang, D.G., 2020. Industrial water-use efficiency in China: regional heterogeneity and incentives identification. *Journal of Cleaner Production*, 258, 120828. doi:10.1016/j.jclepro.2020.120828

Liu, X.H., Liu, L., and Peng, Y., 2017. Ecological zoning for regional sustainable development using an integrated modeling approach in the Bohai Rim, China. *Ecological Modelling*, 353, 158–166. doi:10.1016/j.ecolmodel.2016.09.027

Liu, Y.X., *et al*., 2018. A solution to the conflicts of multiple planning boundaries: landscape functional zoning in a resource-based city in China. *Habitat International*, 77, 43–55. doi:10.1016/j.habitatint.2018.01.004

Liu, Y.X., *et al*., 2019. Landscape functional zoning at a county level based on ecosystem services bundle: methods comparison and management indication. *Journal of Environmental Management*, 249, 109315. doi:10.1016/j.jenvman.2019.109315

Loveland, T.R. and Merchant, J.M., 2004. Ecoregions and ecoregionalization: geographical and ecological perspectives. *Environmental Management*, 34 (S1). doi:10.1007/s00267-003-5181-x.

McBratney, A.B., Mendonça Santos, M., and Minasny, B., 2003. On digital soil mapping. *Geoderma*, 117 (1–2), 3–52. doi:10.1016/S0016-7061(03)00223-4.

McMahon, G., *et al*., 2001. Developing a spatial framework of common ecological regions for the conterminous United States. *Environmental Management*, .28, 293–316. doi:10.1007/s0026702429

Medasani, S., Kim, J., and Krishnapuram, R., 1998. An overview of membership function generation techniques for pattern recognition. *International Journal of Approximate Reasoning*, 19, 391–417. doi:10.1016/S0888-613X(98)10017-8

Miliaresis, G.C., 2013. Terrain analysis for active tectonic zone characterization: a new application for MODIS night LST (MYD11C3) data set. *International Journal of Geographical Information Science*, 27, 1417–1432. doi:10.1080/13658816.2012.685172

Olson, D.M., *et al*., 2001. Terrestrial ecoregions of the world: a new map of life on earth: anew global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 51 (11), 933–938. doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.

Omernik, J.M., 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geographers*, 77, 118–125. doi:10.1111/j.1467-8306.1987.tb00149.x

Omernik, J.M. and Griffith, G.E., 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. *Environmental Management*, 54, 1249–1266. doi:10.1007/s00267-014-0364-1

Peng, J., *et al*., 2019. Multifunctional landscapes identification and associated development zoning in mountainous area. *Science of the Total Environment*, 660, 765–775. doi:10.1016/j.scitotenv.2019.01.023

Powers, R.P., *et al*., 2013. A remote sensing approach to biodiversity assessment and regionalization of the Canadian boreal forest. *Progress in Physical Geography: Earth and Environment*, 37, 36–62. doi:10.1177/0309133312457405

Praene, J.P., *et al*., 2019. GIS-based approach to identify climatic zoning: A hierarchical clustering on principal component analysis. *Building and Environment*, 164, 106330. doi:10.1016/j.buildenv.2019.106330

Qi, F., *et al*., 2006. Fuzzy soil mapping based on prototype category theory. *Geoderma*, 136, 774–787. doi:10.1016/j.geoderma.2006.06.001

Qi, F., *et al*., 2008. Knowledge discovery from area–class resource maps: capturing prototype ects. *Cartography and Geographic Information Science*, 35, 223–237. doi:10.1559/152304008786140533

Qi, F. and Zhu, A.X., 2003. Knowledge discovery from soil maps using inductive learning. *International Journal of Geographical Information Science*, 17, 771–795. doi:10.1080/13658810310001596049

Sheather, S.J., 2004. Density estimation. *Statistical Science*, 19, 588–597. doi:10.1214/088342304000000297

Sheather, S.J. and Jones, M., 1991. A reliable data - based bandwidth selection method for kernel density estimation. *Royal Statistical Society*, B53, 683–690.

Silverman, B.W., 1986. *Density estimation for statistics and data analysis*. London, United Kingdom: Chapman and Hall, 175.

USDA, 1981. *Land resource regions and major land resource areas of the United States*. Washington, DC: Agriculture Handbook 296, USDA Soil Conservation Service, Government Printing Office.

Verichev, K., Zamorano, M., and Carpio, M., 2019. Assessing the applicability of various climatic zoning methods for building construction: case study from the extreme southern part of Chile. *Building and Environment*, 160, 106165. doi:10.1016/j.buildenv.2019.106165

Walsh, A., Cóstola, D., and Labaki, L.C., 2017. Comparison of three climatic zoning methodologies for building energy efficiency applications. *Energy and Buildings*, 146, 111–121. doi:10.1016/j.enbuild.2017.04.044

Wang, C.Y. and Pan, D.L., 2017. Zoning of Hangzhou Bay ecological red line using GIS-based multi-criteria decision analysis. *Ocean & Coastal Management*, 139, 42–50. doi:10.1016/j.ocecoaman.2017.01.013

Wang, H.J., Dai, J.H., and Ge, Q.S., 2012. The spatiotemporal characteristics of spring phenophase changes of Fraxinus chinensis in China from 1952 to 2007. *Science China Earth Sciences*, 55, 991–1000. doi:10.1007/s11430-011-4349-0

Wang, J.F., Zhang, T.L., and Fu, B.J., 2016. A measure of spatial stratified heterogeneity. *Ecological Indicators*, 67, 250–256. doi:10.1016/j.ecolind.2016.02.052

Wu, S.H., *et al*., 2016. Amplitude and velocity of the shifts in the Chinese terrestrial surface regions from 1960 to 2011. *Chinese Science Bulletin*, 61 (19), 2187–2197. in Chinese. doi:10.1360/N972016-00051.

Wu, S.H., Yang, Q.Y., and Zheng, D., 2003a. Comparative study on eco-geographical regional systems between China and USA. *Acta Geographic Sinica*, 58 (5), 686–694.

Wu, S.H., Yang, Q.Y., and Zheng, D., 2003b. Delineation of eco-geographic regional system of China. *Journal of Geographical Sciences*, 13, 309–315. doi:10.1007/BF02837505

Wu, S.H. and Zheng, D., 2000. New recognition on boundary between tropical and subtropical zone in the middle section of ecogeographical system. *Acta Geographic Sinica*, 55 (6), 689–697.

Wu, X., *et al*., 2018. A regional strategy for ecological sustainability: A case study in Southwest China. *Science of the Total Environment*, 616–617, 1224–1234. doi:10.1016/j.scitotenv.2017.10.196

Xu, L.X., *et al*., 2019. An ecosystem services zoning framework for the permafrost regions of China. *Advances in Climate Change Research*, 10, 92–98. doi:10.1016/j.accre.2019.06.007

Xu, X.G., *et al*., 2006. Zoning of sustainable agricultural development in China. *Agricultural Systems*, 87, 38–62. doi:10.1016/j.agsy.2004.11.003

Yang, L., *et al*., 2017. Regional soil mapping using multi-grade representative sampling and a fuzzy membership based mapping approach. *Pedosphere*, 27 (2), 344–357. doi:10.1016/S1002-0160(17)60322-9.

Yu, J.K., Ma, J.Q., and Liu, D., 2020. Historical evolution of marine functional zoning in China since its reform and opening up in 1978. *Ocean & Coastal Management*, 189, 105157. doi:10.1016/j.ocecoaman.2020.105157

Zhang, G.M. and Zhu, A.X., 2018. The representativeness and spatial bias of volunteered geographic information: a review. *Annals of GIS*, 24 (3), 1–12. doi:10.1080/19475683.2018.1501607.

Zhang, X.Y. and Du, S.H., 2015. A linear Dirichlet mixture model for decomposing scenes: application to analyzing urban functional zonings. *Remote Sensing of Environment*, 169, 37–49. doi:10.1016/j.rse.2015.07.017

Zhang, X.Y., Du, S.H., and Zhang, Z.J., 2020. Heuristic sample learning for complex urban scenes: application to urban functional-zone mapping with VHR images and POI data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161, 1–12. doi:10.1016/j.isprsjprs.2020.01.005

Zheng, D., 1998. A study on the regionality and regional differentiation of geography. *Geographical Research*, 17 (1), 4–9.

Zheng, D., et al., 2008. *China's ecogeographical regionalization research*. in Chinese. Beijing: The Commercial Press

Zheng, D., *et al*., 2008. *China's ecogeographical regionalization research*. in Chinese. Beijing: The Commercial Press.

Zheng, D. and Fu, X.F., 1998. A preliminary study on issues of integrated geographical regionalization. *Scientia Geographica Sinica*, 19 (3), 193–197.

Zhou, Y., *et al*., 2003. A GIS-based spatial pattern analysis model for eco-region mapping and characterization. *International Journal of Geographical Information Science*, 17, 445–462. doi:10.1080/1365881031000086983

Zhu, A.X., 1997. A similarity model for representing soil spatial information. *Geoderma*, 77, 217–242. doi:10.1016/S0016-7061(97)00023-2

Zhu, A.X., 1999. A personal construct-based knowledge acquisition process for natural resource mapping using GIS. *International Journal of Geographic Information Science*, 13 (22), 119–141. doi:10.1080/136588199241382.

Zhu, A.X., *et al*., 2010. Construction of membership functions for predictive soil mapping under fuzzy logic. *Geoderma*, 155, 164–174. doi:10.1016/j.geoderma.2009.05.024

Zhu, A.X., *et al*., 2018. A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping. *Catena*, 166, 317–327. doi:10.1016/j.catena.2018.04.003

Zhu, A.X. and Band, L.E., 1994. A knowledge-based approach to data integration for soil mapping. *Canadian Journal of Remote Sensing*, 20, 408–418. doi:10.1080/07038992.1994.10874583