



Distribution of Chinese traditional villages and influencing factors for regionalization

Yunong Wu^{1,2}  Mengqi Wu^{1,2} Zhexiao Wang^{1,2} Beiming Zhang^{1,2} Changzuo Li^{1,2} Bin Zhang^{1*,2}

¹Department of Landscape Architecture, College of Horticulture and Forestry, Huazhong Agricultural University, 430070, Wuhan, China. E-mail: zhangbin@mail.hzau.edu.cn. *Corresponding author.

²Key Laboratory of Urban Agriculture in Central China, Ministry of Agriculture and Rural Affairs, Wuhan, China.

ABSTRACT: *Traditional Villages (TVs) are typical and representative of the agricultural civilization in millions of Chinese villages. The distribution of TVs shows spatial heterogeneity, based on the complexity and diversity of several influencing factors. In this study, 6,819 Chinese TVs were identified and the influencing factors that affect their distribution were screened in terms of three indicator groups: climatic, geographic, and humanity-related factors. Additionally, the K-means clustering algorithm clustered the TVs into different distribution regions. The quantitative relationships between the dominant influencing factors of different distribution regions were revealed to ensure a lucid understanding of the regional distribution of TVs. The results indicated that 1) climatic factors have the greatest impact on the spatial distribution of TVs, followed by geographic factors, particularly the elevation, and then by human factors, of which ethnic distribution played a relatively important role. 2) Twenty-one TV clustering distributions were obtained, which were classified into eight regions of TV distribution with different dominant influencing factors. Management and protective strategies were formulated based on the attribute analysis of influencing factors in each region. The obtained results delineated homogeneous TV distribution regions via the clustering method to achieve an accurate statistical analysis of the influencing factors. This study proposes a new perspective and reference for managing and protecting the diversity, continuity, and integrity of TVs across administrative regions.*

Key words: *traditional/ historical rural settlements, dominant factors, cluster analysis, distribution regions, geographic information system.*

Distribuição das aldeias tradicionais chinesas e fatores de influência para a regionalização

RESUMO: *As aldeias tradicionais (TVs) são típicas e representativas da civilização agrícola em milhões de comunidades chinesas. A distribuição das TVs mostra heterogeneidade espacial, baseada na complexidade e diversidade de diversos fatores influenciadores. Neste estudo, 6.819 TVs chinesas foram identificadas e os fatores que influenciam sua distribuição foram analisados em termos de três grupos de indicadores: fatores climáticos, geográficos e relacionados à humanidade. Além disso, o algoritmo de agrupamento K-means agrupou as TVs em diferentes regiões de distribuição. As relações quantitativas entre os fatores de influência dominantes de diferentes regiões de distribuição foram reveladas para garantir uma compreensão lúcida da distribuição regional de TVs. Os resultados indicam que 1) os fatores climáticos têm maior impacto na distribuição espacial das TVs, seguidos dos geográficos, em particular da elevação, e depois dos humanos, nos quais a distribuição étnica teve um papel relativamente importante; 2) foram obtidas as vinte e uma distribuições de agrupamento de TV, as quais foram classificadas em oito regiões de distribuição de TV com diferentes fatores de influência dominante. Estratégias de manejo e proteção foram formuladas com base na análise de atributos dos fatores influenciadores de cada região. Os resultados obtidos delimitaram regiões homogêneas de distribuição de TV por meio do método de agrupamento para obter uma análise estatística precisa dos fatores de influência. Este estudo propõe uma nova perspectiva e referência para a gestão e proteção da diversidade, continuidade e integridade das TVs nas regiões administrativas.*

Palavras-chave: *assentamentos rurais tradicionais / históricos, fatores dominantes, análise de cluster, regiões de distribuição, sistema de informação geográfica.*

INTRODUCTION

Urbanization, producer and consumer behavior, and extensive economic growth have resulted in the serious degradation of the living environment in villages (ANTROP, 2004; TAN & LI, 2013) such that there is a disappearance of important natural and cultural heritages specific to

certain villages (YANG et al., 2018). In China, there are approximately 10 million rural-to-urban migrants each year, and the number of villages has decreased by 900,000 over the past decade (YANG, 2013). In 2012, the Chinese government launched a collaborative investigation into Traditional Villages (TVs) to identify and protect villages with valuable historical and cultural heritage values. TVs, also known as

'ancient villages,' formed early; have rich traditional resources; certain historical, cultural, scientific, artistic, social, and economic values; and still retain relatively well-preserved village style features, architectural structures, and unique folk customs that must be protected (MINISTRY OF HOUSING AND URBAN-RURAL DEVELOPMENT OF THE PEOPLE'S REPUBLIC OF CHINA et al., 2012).

Although, there is no uniform definition of TVs, traditional/historical rural settlements with homology have long been studied (PARSONS, 1972). Previous studies have examined the villages' locations in relation to several factors, such as the geographical environment, economics, and population demographics (GROSSMAN, 1987; DIXON, 1996) as well as immigration (MCLEMAN, 2011; YAWORSKY & CODDING, 2018). These studies were either focused on a single factor or used a qualitative approach. Several studies have evaluated the influencing factors and laws that govern the distribution of ancient settlements from an archaeological perspective (PARSONS, 1972; DEMJÁN & DRESLEROVÁ, 2016; FANTA et al., 2020). These studies have emphasized the relationship between early complex rural settlements and specific environments (GREEN & PETRIE, 2018), often focusing on a fixed historical period, with an emphasis on the protection of relics. With the development of computer and remote sensing technologies, spatial analysis technology based on geographic information systems (GIS) and mathematical modelling have been used to quantitatively analyze the spatial patterns of TVs (SEVENANT & ANTROP, 2007; BRAGA et al., 2016; YANG et al., 2018; FANTA et al., 2020). These studies often adopt the methods of the case study and lack of quantitative analysis of large samples of TVs. Additionally, even macro-scale studies have not considered in-depth analyses of the internal mechanisms that drive the differences and similarities in the TV distribution due to different dominant factors.

In this study, based on the 13 influencing factors of 6,819 TVs in China, we used cluster analysis to divide the TVs into 21 groups. In addition, regression analysis of each cluster further divided the TVs into 8 distribution regions. We focused on three issues: what is the spatial distribution of TVs under the influence of differences and synergy?; what are the dominant factors?; and how do we carry out the scientific management and protection zoning of TVs based on the dominant factors? The purpose of this study was to: 1) further understand the distribution status and formation environment of TVs

and 2) identify the distribution of TV management regions under the influence of similar environmental factors and determine the main influencing factors within each region to guide subsequent planning and policy formulation. This study seeks to overcome the limitations associated with the over-dependence on administrative boundaries and the qualitative method and integrated analysis of environmental factors affecting TV employed in the previous studies. The research framework proposed in this study is conducive to the clarification of regional studies of TVs.

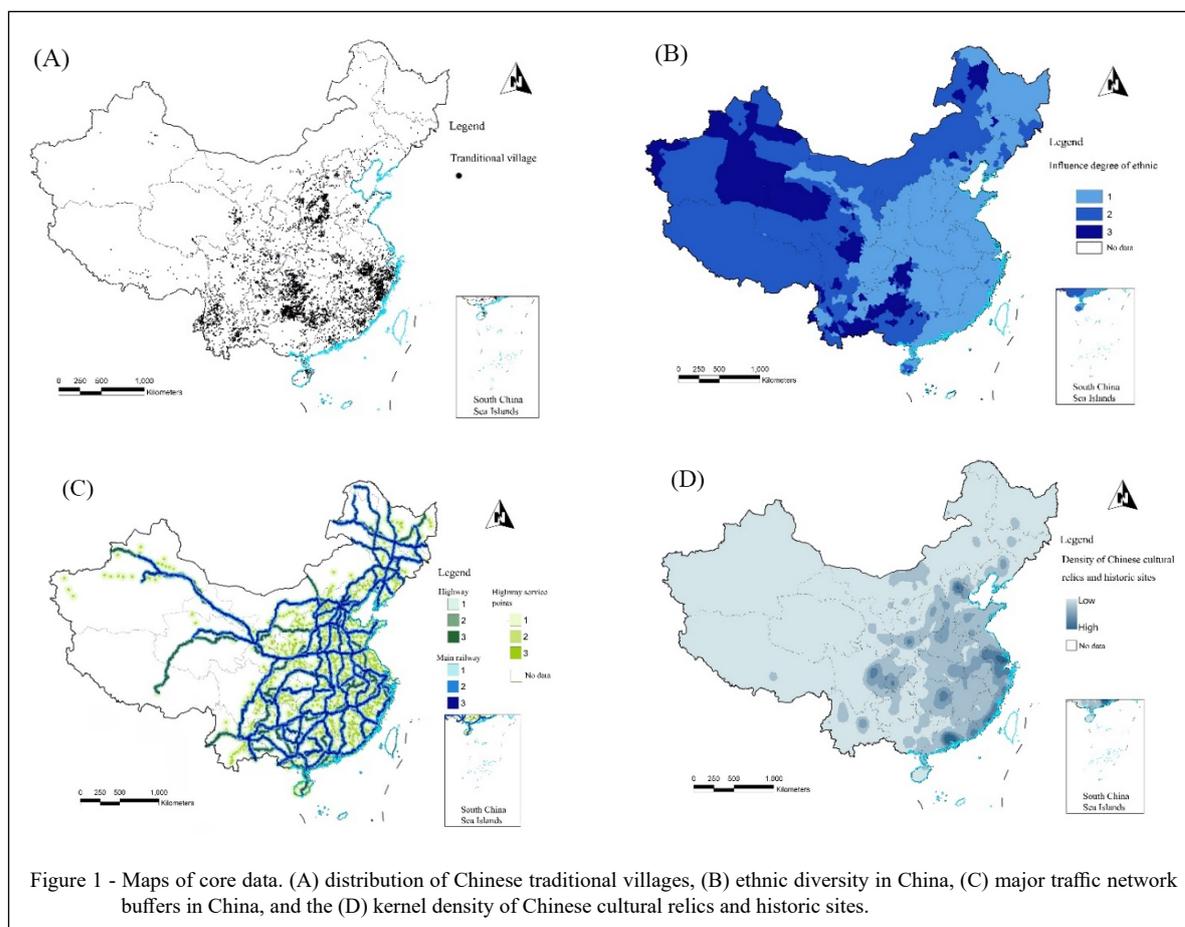
The structure of this article is as follows. First, we introduced the data collection and research methods. We present the results of this study from two perspectives: 21 cluster attributes and the characteristics of 8 distribution regions. Finally, we discussed our results and propose management and protection suggestions for different distribution regions.

MATERIALS AND METHODS

Data collection

According to an evaluation of TV Protection and Development, a total of 6,819 TVs from five approval batches dated up to June 2019 were selected, excluding those in Hong Kong, Macao, and Taiwan (MINISTRY OF HOUSING AND URBAN-RURAL DEVELOPMENT OF THE PEOPLE'S REPUBLIC OF CHINA et al., 2019). Spatial attributes such as the geographical coordinates of TVs were obtained manually using Google Earth to draw a spatial distribution map of the selected TV points (Figure 1).

The influencing factors for the TV distributions were divided into three categories: climate, geography, and humanities. The 13 factors meeting the scale requirements were as follows. A. Climatic factors: climatic zone (vectorized 'distribution map of climate types in China' [<http://map.ps123.net/>], with a total of five climatic zones), annual average temperature, annual average precipitation, and wetness index (China meteorological background dataset, 500 m × 500 m resolution, downloaded from the Resource and Environment Data Cloud Platform (XINLIANG XU, 2017)). B. Geographic factors: elevation, slope, aspect (determined using SRTM DEM V4 [<http://srtm.csi.cgiar.org>], resampled to 1 km and transformed using spatial analysis tools in ArcGIS), and soil type (obtained from the Resource and Environment Data Cloud Platform [<http://www.resdc.cn/data.aspx?DATAID=145>], and divided into 13 Soil Order categories). C. Human factors: population, economy



(vectorized data obtained from the China Bureau of Statistics and from the China Statistical Yearbook in 2010, with maps downloaded from the Resource and Environment Data Cloud Platform [XU, 2017b; XU, 2017a]), traffic networks (obtained from Google Maps), ethnicity (according to the ‘Atlas of Ethnic Minorities Distribution 2012’ [INSTITUTE OF LINGUISTICS CASS, 2012], in combination with the distribution of the Han nationality, and vectorized using ArcGIS), and cultural relics and historical sites (obtained from Baidu Maps).

Data processing

The 13 influencing factors included continuous and categorical variables. The categorical variables were the climatic zone, aspect, and soil type. Due to the limitations of categorical variables in the regression analysis and K-means clustering analysis, these variables were transformed into multiple dummy binary variables. When a categorical variable

has n categories, $n - 1$ dummy variables were defined. If a dummy variable corresponds to a category, its value was set to 1; otherwise, the value was set to 0 (TSUNODA et al., 2012).

Ethnicity is representative of religion, beliefs and customs, and it has numerous classifications. Therefore, identifying and processing this variable is difficult. We adopted the sixth census statistics from 2010 and set the county scale as the calculation unit. The number of dominant ethnic groups determined the influence of ethnic diversity, which was then classified as ‘high,’ ‘medium,’ or ‘low’ diversity, assigned as values of 3, 2, and 1, respectively by ArcGIS, which were then transformed into a map (Figure 1-b).

The special continuous variables were the traffic network and history relics, which required data processing. Using the macro traffic system of the main Chinese railways, highways, and highway service points and the multiple ring buffer module of

the analysis toolbox in ArcGIS, a three-layer valuation buffer zone was established (Figure 1-c) in terms of the radiation attenuation of traffic influence. Buffer zones closer to the traffic network resulted in higher values for the traffic influence, which were then assigned to the TV points. We used a kernel density analysis tool to establish a map of the cultural relics and historic sites and determine their influence, which ranged from high to low based on the distance from the sites (Figure 1-d). The obtained values were also assigned to the TV location points. For the remaining continuous variables, we only removed the outliers and missing values. Finally, a database of the 13 influencing factors of 6,834 TV points was constructed.

Cluster analysis of influencing factors

K-means clustering, one of the classic and most commonly used clustering methods, is a simple unsupervised learning algorithm (MACQUEEN, 1967). This algorithm is widely used to cluster large-scale data (JAIN & DUBES, 1988). The sum square error (SSE) value of the specified cluster was calculated using the 'elbow method.' When the SSE value dropped sharply to a specific point and remained unchanged significantly, this point was designated as the elbow point, which represented the optimal K value of the clustering number (KODINARIYA & MAKWANA, 2013), calculated as follows:

$$SSE = \sum_{i=1}^k \sum_{p \in L_i} |P - q_i|^2$$

where L_i is the i -th cluster, P is the sample point in L_i , q_i is the centroid of L_i (mean of all samples in L_i), and k represents the number of groups.

In this study, the 13 factors and geographic coordinate information of TVs were selected as the clustering database, where the optimal cluster number range was identified using the 'elbow method.' Within this range, the optimal K value was determined through experiments and combined with manual selection. The obtained optimal K value and the clustering database were input into the grouping analysis module of the spatial statistical tool in ArcGIS. The K-means clustering algorithm was selected to obtain the clustering distribution diagram of TVs' influencing factors.

Analysis of the spatial density distribution of TVs using the kernel density tool

In kernel density estimation, a moving cell was mainly used to estimate the density of point or line patterns (ALLEN, 2016). A continuous density

surface was generated by calculating the discrete point elements in the region to visually show the agglomeration status of points within the entire region as follows:

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

where x and x_i are the spatial points, $f_n(x)$ is the estimated kernel density at a certain point, and h is the bandwidth. Bandwidth was set according to the default of the system.

Multiple regression analysis

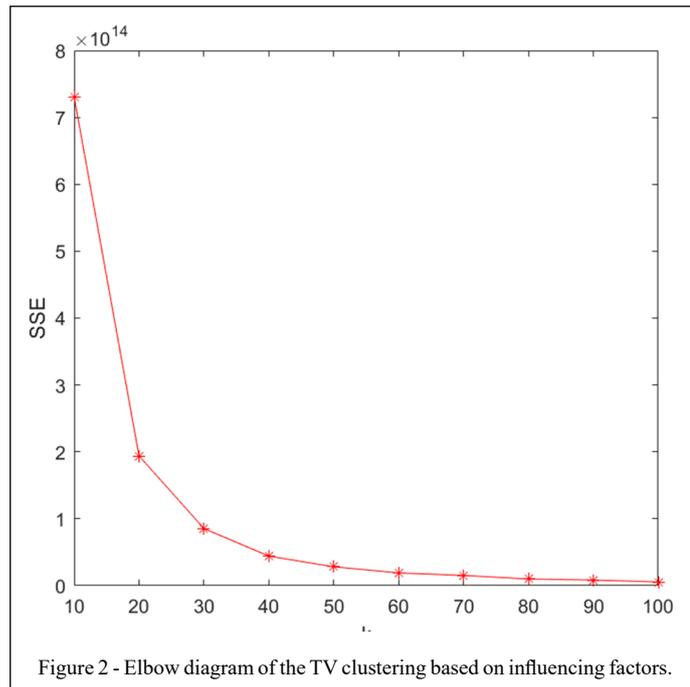
SPSS (v25.0) was used to conduct multiple linear regression analysis, where the kernel density value of TV points in each cluster was the dependent variable and the 13 influencing factors were the independent variables. The statistically significant factors influencing the TV density distribution of each clustering were determined. The areal units in close proximity to each other (GUO et al., 2000) were grouped into eight regions representing different dominant factors.

RESULTS

Cluster analysis of TVs with similar influencing factors

According to the 13 influencing factors and the longitude and latitude information of the TV points, the optimal cluster number k value was determined to be 21 using the elbow method (Figure 2). The grouping analysis tool in ArcGIS was used to generate the TV cluster distribution map, which is shown in figure 3.

The majority of the cluster areas showed strong agglomeration in space, including clusters 2, 3, 5, 8, 9, 10, 11, 12, 14, 20, and 21. For example, in cluster 10, there were 459 TVs that were predominantly distributed in the temperate monsoon climate zone and at high elevations in north-central China. Dense areas of cultural relics ($p = 0.000$, $\beta = 0.205$) had a positive influence on TV distribution, whereas the wetness index ($p = 0.008$, $\beta = -0.222$) had significantly negative effects. Additionally, cluster 11 had a medium density of TVs, but exhibited strong characteristics. TVs were mainly distributed in eastern Yunnan and had a high density of cultural relics and historic sites, which can be attributed to the high ethnic diversity ($p = 0.000$, $\beta = 0.537$), as Yunnan is the province with the most ethnic minorities. In addition, cluster 11 had low population and traffic densities and it was easily distributed in plain areas



with high elevations and slopes. These findings suggested that this cluster was typically protected by the natural environment from the adverse effects of social factors, such as the economy and traffic network.

Certain clusters were multi-regional, including clusters 15, 18, and 19, which were widely distributed in central and southern China. Notably, cluster 17 had 193 TVs with no significant influence from all 13 factors.

TV regions with different dominant influencing factors

Combining clusters and our interpretation of clustering attributes, we summarized the eight regions of TV distribution (Figure 4). Region I, climate-dominant type, included clusters 9 and 5. This region, located in southern Tibet and the northeast of Sichuan, has low precipitation, wetness, and temperature, but high elevations. TVs in Region I tend to be distributed in areas with suitable climatic conditions.

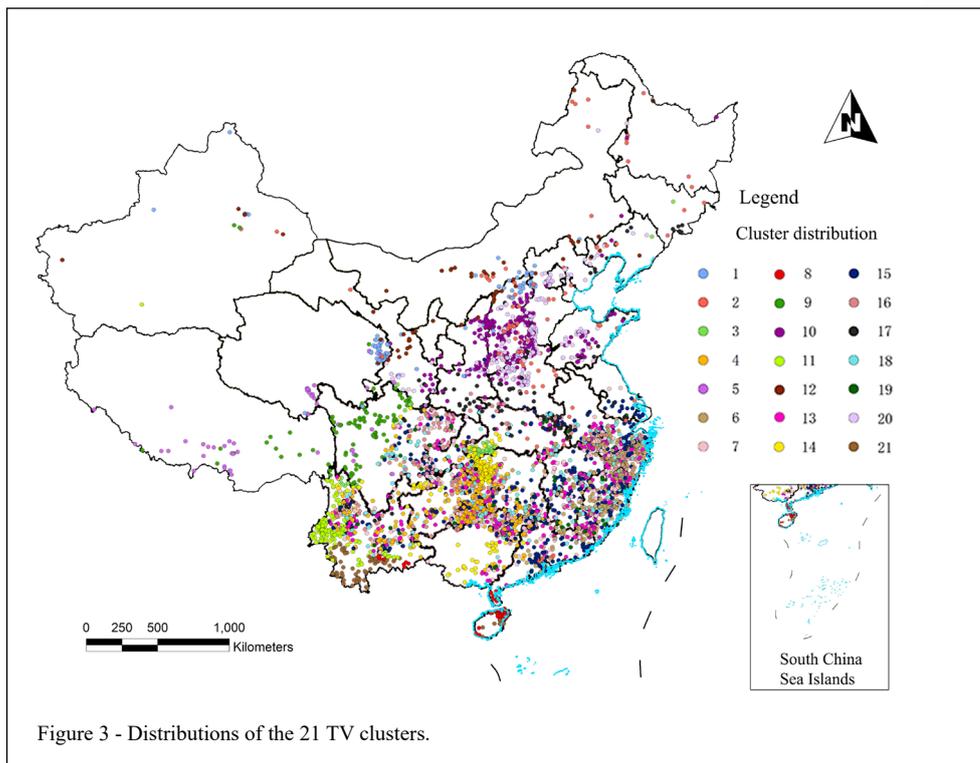
Region II, humanity/climate-dominant type, includes clusters 1, 2, and 12. Region II is located in northern China, southern Inner Mongolia, and Gansu, has a low population, and few cultural relics. TVs are typically distributed in places with a relatively high concentration of cultural relics and historical sites, and they are significantly influenced by ethnic groups.

Region III, humanity-dominant type, includes clusters 10 and 20. This area is located in Shaanxi, Shanxi, Hebei, Shandong, and Henan, and has a high density of cultural relics and historic sites, with the strongest influence on the distribution of TVs. In addition, TVs are distributed in areas with improved economic conditions.

Region IV, geography/humanity-dominant type, includes clusters 11 and 21. This region is mainly located in Yunnan. As one of the provinces with the highest density of TVs, Yunnan exhibits typical ethnic diversity characteristics. TVs are mostly distributed in high elevation areas, which are strongly correlated with a high density of cultural relics and historic sites.

Region V, geography/climate-dominant type, includes clusters 8 and 14. This region is mainly located in Guangxi, Hainan, and southern Yunnan provinces. The distribution of TVs was evidently polarized according to elevation. Owing to the high temperature in this region, TVs are distributed in areas with high rainfall and relatively low temperature.

Region VI, climate/humanity-dominant comprehensive type, is mainly represented by cluster 6, but also includes clusters 13, 15, and 18. This region is located in the eastern coastal region of China, and is significantly affected by the distribution density of cultural relics and historic sites, as well as climate



factors. The TV distribution density has a positive correlation with a high wetness index.

Region VII, geography/climate-dominant comprehensive type, is mainly represented by cluster 4, but also includes clusters 3, 13, and 14. This region is mainly located in Guizhou, Yunnan, and Hunan, with minor regions in additional places. TVs are mainly distributed in areas with low temperature and high humidity, which is typical of residential settlement distribution under similar climatic conditions.

Region VIII, the hybrid type, is primarily represented by cluster 16 but also included clusters 7, 15, 18, and 19. It is predominantly located in Sichuan and Chongqing. TV distribution of this region is significantly affected by cultural relics and historic sites. TVs are distributed in the areas with high values of climate factors, slope, and economy, which may be closely related to the unique micro-climate environment of the mountainous terrain in this region.

DISCUSSION

Different countries have different definitions of historic rural settlements. For example, in Scotland, historic rural settlements included all

habitations from the fifth to the twentieth century (HARRISON et al., 2008). In China, villages must meet the requirements of 20 indicators in three aspects: architecture, site selection and layout, as well intangible cultural heritage in the 'Traditional Village Evaluation and Identification Index System' to be rated as a traditional village (Traditional village evaluation and identification index system (trial implementation), 2012). This provided a clearer definition of the concept of traditional Chinese villages. Recently, discourses promoting the enhancement and development of villages have become a recurrent theme across the globe; the experience in China presents unique characteristics and also has a certain reference significance (POLA, 2019).

An improved understanding of how environmental factors contributed to regional variation in historic settlement organization can improve our understanding of the significance of variations in rural settlements and their historic character as well as the overall setting of numerous types of rural buildings (LOWERRE, 2014). Generally, in China, the climate factors have the most significant influence on a TV's distribution, followed by geographic factors, among which elevation has the

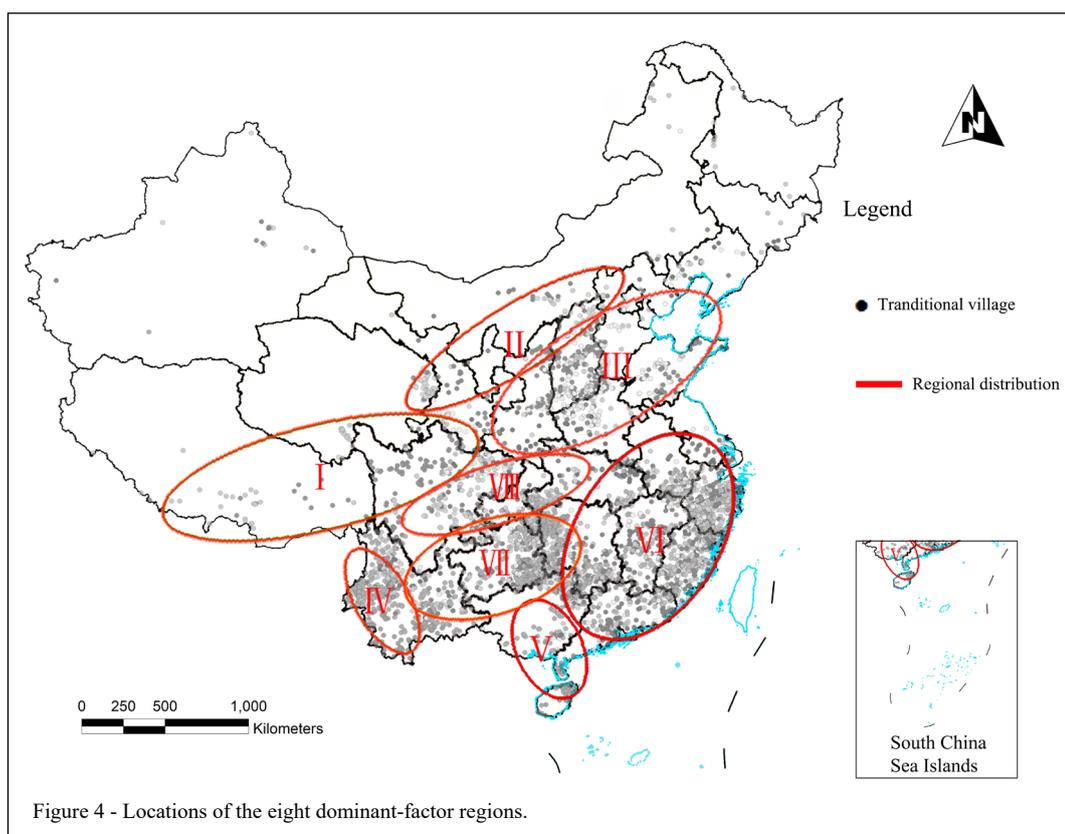


Figure 4 - Locations of the eight dominant-factor regions.

greatest impact, and then by human factors, among which ethnic diversity plays a relatively important role. These three factors are spatio-temporally inter-related, and all of them contribute to the living environment in TVs. Our intent was not to advocate a completely 'environmentally determinist' approach determining the variations in historic settlement organization. Rather, it was to explore the strengths (or weaknesses) of the relationships between environmental factors and settlement organization and how those relationships vary across the country (LOWERRE, 2014). After clustering and drawing the set of settlements and environmental variables, it is possible to more directly explore the relationship between the environmental factors and TVs, and determine the environmental variables that have the greatest impact on the regional changes of TVs.

Although, there are numerous spatial clustering methods, they are not applicable in regionalization (GUO, 2008). The K-means statistical method has proven to be an effective method to determine settlement groups (ÄYRÄMÖ & KÄRKKÄINEN, 2006; KOWALEWSKI, 2008). The

resulting 21 types of clusters clarified the relationship between the TVs and their environments by reducing the statistical scale. Results from the clustering showed that the TV groups with similar environmental factors were distributed across administrative boundaries. Cluster types located close together had more similar dominant influencing factors than those further apart (TSUNODA et al., 2012). For the 21 clusters, the fitting effect of multiple regression was reported to be better than the overall analysis. The significant impact factors of each group are different, which also confirms the complexity of TV distribution factors, and hence, unified protection standards may not apply. Interestingly, some groups distributed in south-eastern China have relatively low differences in natural factors such as topography and geomorphology, but owing to the wide-ranging exchanges of culture in this region and the depth of their influences, the complicated interaction between socio-economic factors and changes in the bio-physical environment, present some clusters with a complex state of large-scale cohesion and partial dispersion.

To uncover a pattern within the data and render the problem size manageable, it is common in most spatial studies to further aggregate the data into larger areal units or zones (GUO et al., 2000). We summarized the results of multiple linear regressions of cluster types, which provided a further characteristic basis for the regional construction of TVs and considered their geographical proximity. The boundaries of the eight regions were somewhat permeable or vague, but they are internally meaningful (KOWALEWSKI, 2008). This can facilitate the adoption of a rational and prudent approach to space and strategic planning.

Crop and livestock production are heavily dependent on rainfall. Temperature and rainfall in general are the most important climatic factors influencing the availability of natural resources and livelihood strategies (ZAMPALIGRÉ et al., 2014). For the TVs within the climate-dominant region, it is necessary to strengthen the long-term and dynamic monitoring of climatic factors, meanwhile the local residents' perception of climatic factor should also be considered (BYG & SALICK, 2009). Additionally, some secondary factors of climate, which may cause floods, soil and water loss, and drought, should be emphasized and corresponding measures adopted over time.

Early settlements produced complex interactions between culture and nature (MCGOVERN et al., 2007). These complex interactions are reflected in ethnic diversity as influenced by geographic diversity (PARSONS, 1972; PAIK & SHAWA, 2013). Chinese TVs dominated by humanity factors have a strong ethnic diversity and mature cultural landscapes, necessitating their protection along with natural territories. Additionally, economic development and convenient transportation are also necessary to maintain the authenticity and sustainability of TVs.

TVs dominated by geographic factors are mainly located in areas with high elevation, large slope, and an evident topographic relief. Restricted by topography, TVs were scattered on multi-directional slopes where soil is suitable for cultivation (YANG et al., 2016). Based on these results, we suggested regular maintenance of high-quality agricultural land and the living conditions of TVs. These regions are more conducive to the preservation of multi-ethnic TVs and have highly prominent regional characteristics (PAIK & SHAWA, 2013).

TVs dominated by hybrid factors exhibited distinct geographic characteristics and a diversified distribution that correspondingly made their landscapes diverse. Therefore, it is necessary to strengthen the perception of TV landscape characteristics and further investigate the interactions among various influencing factors of the TV

distribution to formulate more detailed management and protection measures (ANTROP, 2004).

CONCLUSION

The 13 influencing factors used in this study were closely related to the distribution of TVs. For different regions, there are differences in the significance of the influencing factors and their positive and negative effects, but there are also significant similarities in these factors, which form the 21 TV clusters and 8 types of regions based on the dominant factors. Based on the characteristics of these 8 TV regions, we can formulate diversified, targeted, and possibly even cross-regional joint management and protection strategies. These results further indicated that the combination of unsupervised classification clustering technique and multiple regression is an effective regional analysis method, and it has been possible to develop new, national-scale characterizations of historic settlement organization and their key environmental variables (LOWERRE, 2014) and effectively manage and protect rural settlements and landscapes with a long history (HARRISON et al., 2008).

This study has the following limitations: 1) the selection of the influencing factors was limited based on data availability, human resources, and technological conditions; 2) during data processing, avoiding the intervention of subjective factors such as definition of the traffic network radiation range, determination of ethnic value, and division of the eight regions was difficult. In future studies, the influencing factors should be further enriched such that the content of the analysis will not be solely limited to the influence of the environmental factors on TVs. Instead, they will also include the interactions between the factors that influence the distribution of the TVs. The environmental tendencies of TVs distribution can be used as a basis for the classification of TV typology and used as a reference for the identification and mining of TVs.

ACKNOWLEDGEMENTS

This study was funded by the National Natural Science Foundation of China, grant number 51678269, and the Chinese Fundamental Research Funds for the Central Universities (HZAU: 2662018YJ017).

DECLARATION OF CONFLICT OF INTEREST

The authors declare no conflicts of interest. The funding sponsors had no role in the design of this study; in the

collection, analyses, or interpretation of the data; in the writing of the manuscript, and in the decision to publish the results.

AUTHORS' CONTRIBUTION

Yunong Wu contributed to conceptualization, methodology, software, validation, formal analysis, data curation, and writing of the original draft. Mengqi Wu, Zhexiao Wang, Beiming Zhang, and Changzuo Li contributed to data curation and software. Bin Zhang contributed to the conceptualization, methodology, supervision, and writing through reviewing and editing. All authors have read and agreed to the published version of the manuscript.

REFERENCES

- ALLEN, D. W. **GIS tutorial 2: spatial analysis workbook**. Redlands: Esri Press. 2016. 424 p.
- ANTROP, M. Landscape change and the urbanization process in Europe. **Landscape and Urban Planning**, v.67, n.1-4, p.9-26, 2004. Available from: <<https://www.sciencedirect.com/science/article/pii/S0169204603000264>>. Accessed: Feb. 13, 2019. doi: 10.1016/s0169-2046(03)00026-4.
- ÄYRÄMÖ, S.; T. KÄRKKÄINEN. **Introduction to partitioning-based clustering methods with a robust example**. Department of Mathematical Information Technology, University of Jyväskylä 2006. (C. 1/2006).
- BRAGA, G. B., et al. A methodology for definition of rural spaces: an implementation in Brazil. **Ciencia Rural**, v.46, n.2, p.375-380, 2016. Available from: <https://www.scielo.br/scielo.php?pid=S0103-84782015005050464&script=sci_arttext>. Accessed: Aug. 23, 2019. doi: 10.1590/0103-8478cr20150464.
- BYG, A.; J. SALICK. Local perspectives on a global phenomenon—climate change in Eastern Tibetan villages. **Global Environmental Change**, v.19, n.2, p.156-166, 2009. Available from: <<https://www.sciencedirect.com/science/article/abs/pii/S0959378009000193>>. Accessed: Nov. 21, 2019. doi: 10.1016/j.gloenvcha.2009.01.010.
- DEMJÁN, P.; D. DRESLEROVÁ. Modelling distribution of archaeological settlement evidence based on heterogeneous spatial and temporal data. **Journal of Archaeological Science**, v.69, p.100-109, 2016. Available from: <<https://doi.org/10.1016/j.jas.2016.04.003>>. Accessed: Aug. 10, 2019. doi: 10.1016/j.jas.2016.04.003.
- DIXON, D. RURAL GEOGRAPHY AND POLITICAL ECONOMY. **The North Carolina Geographer**, v.5, p.65-73, 1996. Available from: <<https://ncgeography.org/journal/index.php/NCGeographer/article/view/127/127>>. Accessed: Aug. 10, 2019.
- FANTA, V., et al. How old are the towns and villages in Central Europe? Archaeological data reveal the size of bias in dating obtained from traditional historical sources. **Journal of Archaeological Science**, v.113, p.105044, 2020. Available from: <<https://doi.org/10.1016/j.jas.2019.105044>>. Accessed: Jun. 22, 2019. doi: 10.1016/j.jas.2019.105044.
- GREEN, A. S.; C. A. PETRIE. Landscapes of Urbanization and De-Urbanization: A Large-Scale Approach to Investigating the Indus Civilization's Settlement Distributions in Northwest India. **Journal of Field Archaeology**, v.43, n.4, p.284-299, 2018. Available from: <<https://doi.org/10.1080/00934690.2018.1464332>>. Accessed: Jun. 22, 2019. doi: 10.1080/00934690.2018.1464332.
- GROSSMAN, D. Rural polarization—the relation between population, spatial patterns and socio-economic characteristics: the case of southwestern Cheshire, England. **Tijdschrift Voor Economische En Sociale Geografie**, v.78, n.4, p.276-289, 1987. Available from: <<https://doi.org/10.1111/j.1467-9663.1987.tb01876.x>>. Accessed: Jun. 22, 2019. doi: 10.1111/j.1467-9663.1987.tb01876.x.
- GUO, D. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). **International Journal of Geographical Information Science**, v.22, n.7, p.801-823, 2008. Available from: <<https://doi.org/10.1080/13658810701674970>>. Accessed: Dec. 22, 2019. doi: 10.1080/13658810701674970.
- GUO, J., et al. MOZART: A multi-objective zoning and aggregation tool. **Proceedings of the Philippine Computing Science Congress (PCSC)**. Manila. Nov 29-Dec 1, 2000, 2000. p.197-201. Available from: <<https://pdfs.semanticscholar.org/3bf1/4a1c997404f1f37c98204e1fe96624c398ae.pdf>>. Accessed: Feb. 15, 2020.
- HARRISON, J., et al. A Research framework for historic rural settlement studies in Scotland. 2008. Available from: <https://www.researchgate.net/publication/237402361_A_Research_Framework_for_Historic_Rural_Settlement_Studies_in_Scotland>. Accessed: Dec. 10, 2019.
- INSTITUTE OF LINGUISTICS CASS, T. I. O. E. A. A., CHINESE ACADEMY OF SOCIAL SCIENCE, EDITED BY THE CENTER FOR LANGUAGE AND INFORMATION SCIENCE, CITY UNIVERSITY OF HONG KONG. **Chinese dialect volume of Chinese Language Atlas (2nd Edition)**. Beijing: The Commercial Press. 2012. 232 / 291 p.
- JAIN, A. K.; R. C. DUBES. **Algorithms for clustering data**. United States: Prentice-Hall, Inc. 1988
- KODINARIYA, T. M.; P. R. MAKWANA. Review on determining number of Cluster in K-Means Clustering. **International Journal**, v.1, n.6, p.90-95, 2013. Available from: <https://www.researchgate.net/publication/313554124_Review_on_Determining_of_Cluster_in_K-means_Clustering>. Accessed: Jun. 22, 2019.
- KOWALEWSKI, S. A. Regional settlement pattern studies. **Journal of Archaeological Research**, v.16, n.3, p.225-285, 2008. Available from: <<https://link.springer.com/article/10.1007/s10814-008-9020-8>>. Accessed: May. 15, 2020. doi: 10.1007/s10814-008-9020-8.
- LOWERRE, A. **Rural Settlement in England: Analysing Environmental Factors and Regional Variation in Historic Rural Settlement Organisation Using Regression and Clustering Techniques**. English Heritage. London. 2014. (72).
- MACQUEEN, J. Some methods for classification and analysis of multivariate observations. **Proceedings of the fifth Berkeley symposium on mathematical statistics and probability**: Oakland, CA, USA, 1967. p.281-297. Available from: <<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.308.8619>>. Accessed: Sep. 07, 2019.

- MCGOVERN, T. H., et al. **Landscapes of settlement in northern Iceland: Historical ecology of human impact and climate fluctuation on the millennial scale.** v.109, n.1, p.27-51, 2007. Available from: <<https://doi.org/10.1525/aa.2007.109.1.27>>. Accessed: Jun. 22, 2019. doi: 10.1525/aa.2007.109.1.27.
- MCLEMAN, R. A. Settlement abandonment in the context of global environmental change. **Global Environmental Change-Human and Policy Dimensions**, v.21, p.S108-S120, 2011. Available from: <<https://doi.org/10.1016/j.gloenvcha.2011.08.004>>. Accessed: Jun. 22, 2019. doi: 10.1016/j.gloenvcha.2011.08.004.
- MINISTRY OF HOUSING AND URBAN-RURAL DEVELOPMENT OF THE PEOPLE'S REPUBLIC OF CHINA, et al. **the Ministry of Housing and Urban-Rural Development, the Ministry of Culture and Tourism, National Cultural Heritage Administration and the Ministry of Finance on conducting surveys of traditional villages notice** (Jiancun [2012] No. 58). 2012. Available from: <http://www.gov.cn/zwqk/2012-04/24/content_2121340.htm>. Accessed: Jun. 10, 2019.
- MINISTRY OF HOUSING AND URBAN-RURAL DEVELOPMENT OF THE PEOPLE'S REPUBLIC OF CHINA, et al. **Ministry of Housing and Urban-Rural Development and other departments on announcing the fifth batch of village lists included in the list of traditional villages in China.** 2019. Available from: <http://www.mohurd.gov.cn/wjfb/201906/t20190620_240922.html>. Accessed: Jun. 10, 2019.
- PAIK, C.; T. W. SHAWA. Altitude and adaptation: A study of geography and ethnic division. **Applied geography**, v.40, p.212-221, 2013. Available from: <<https://doi.org/10.1016/j.apgeog.2013.02.003>>. Accessed: Jun. 22, 2019. doi: 10.1016/j.apgeog.2013.02.003.
- PARSONS, J. R. Archaeological settlement patterns. **Annual Review of Anthropology**, v.1, n.1, p.127-150, 1972. Available from: <<https://www.jstor.org/stable/2949239?seq=1>>. Accessed: Jun. 22, 2019. doi: 10.1146/annurev.an.01.100172.001015.
- POLA, A.-P. Reframing Chinese villages. **Modu magazine**, China, Aug 21, 2019. Available from: <<https://www.modumag.com/focus/reframing-chinese-villages/>>. Accessed: Feb. 05, 2020.
- SEVENANT, M.; M. ANTRON. Settlement models, land use and visibility in rural landscapes: Two case studies in Greece. **Landscape urban planning**, v.80, n.4, p.362-374, 2007. Available from: <<https://doi.org/10.1016/j.landurbplan.2006.09.004>>. Accessed: Nov. 18, 2019. doi: 10.1016/j.landurbplan.2006.09.004.
- TAN, M. H.; X. B. LI. The changing settlements in rural areas under urban pressure in China: Patterns, driving forces and policy implications. **Landscape and Urban Planning**, v.120, p.170-177, 2013. Available from: <<https://doi.org/10.1016/j.landurbplan.2013.08.016>>. Accessed: Nov. 18, 2019. doi: 10.1016/j.landurbplan.2013.08.016.
- TRADITIONAL VILLAGE EVALUATION AND IDENTIFICATION INDEX SYSTEM (trial implementation). **Ministry of Housing and Urban-Rural Development of the People's Republic of China Ministry of Culture of the People's Republic of China National Cultural Heritage Administration**
- Ministry of Finance of the People's Republic of China**, China, 2012. Available from: <http://www.mohurd.gov.cn/wjfb/201208/t20120831_211267.html>. Accessed: Jun. 10, 2019.
- TSUNODA, M., et al. Handling categorical variables in effort estimation. **Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement**, 2012. p.99-102. Available from: <<https://dl.acm.org/doi/abs/10.1145/2372251.2372267>>. Accessed: May. 15, 2020.
- XINLIANG XU, Y. Z. China Meteorological background data set. **Data registration and publishing system of data center of resources and Environmental Sciences, Chinese Academy of Sciences**, Resource and Environment Data Cloud Platform, 2017. Available from: <<http://www.resdc.cn/DOI>>. Accessed: Oct. 08, 2019. doi: 10.12078/2017121301.
- XU, X. China's GDP spatial distribution kilometer grid data set. **Data registration and publishing system of the resource and environment science data center of the Chinese Academy of Sciences**, Resource and Environment Data Cloud Platform, 2017a. Available from: <<http://www.resdc.cn/DOI>>. Accessed: Oct. 14, 2019. doi: 10.12078/2017121102.
- XU, X. China population spatial distribution kilometer grid data set. **Data registration and publishing system of data center of resources and environment science of Chinese Academy of Sciences** Resource and Environment Data Cloud Platform, 2017b. Available from: <<http://www.resdc.cn/doi>>. Accessed: Oct. 14, 2019. doi: 10.12078/2017121101.
- YANG, R., et al. Spatial distribution characteristics and optimized reconstruction analysis of China's rural settlements during the process of rapid urbanization. **Journal of Rural Studies**, v.47, p.413-424, 2016. Available from: <<https://doi.org/10.1016/j.jrurstud.2016.05.013>>. Accessed: Aug. 11, 2019. doi: 10.1016/j.jrurstud.2016.05.013.
- YANG, X., et al. An Intelligent Analysis for Rural Settlement Distribution Based on Gaussian Mixture Models: A Case Study of Kengzi Village. **2018 IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC): IEEE**, 2018. p.478-484. Available from: <<https://ieeexplore.ieee.org/abstract/document/8482033>>. Accessed: Jun. 10, 2019.
- YANG, X. J. China's rapid urbanization. **Science**, v.342, n.6156, p.310-310, 2013. Available from: <<https://science.sciencemag.org/content/342/6156/310>>. Accessed: Jun. 10, 2019. doi: 10.1126/science.342.6156.310-a.
- YAWORSKY, P. M.; B. F. CODDING. THE IDEAL DISTRIBUTION OF FARMERS: EXPLAINING THE EURO-AMERICAN SETTLEMENT OF UTAH. **American Antiquity**, v.83, n.1, p.75-90, 2018. Available from: <<https://doi.org/10.1017/aaq.2017.46>>. Accessed: Jun. 10, 2019. doi: 10.1017/aaq.2017.46.
- ZAMPALIGRÉ, N., et al. Climate change and variability: perception and adaptation strategies of pastoralists and agro-pastoralists across different zones of Burkina Faso. **Regional Environmental Change**, v.14, n.2, p.769-783, 2014. Available from: <<https://link.springer.com/article/10.1007/s10113-013-0532-5>>. Accessed: Jun. 10, 2019. doi: 10.1007/s10113-013-0532-5.