An Entropy-based Method to Evaluate the Appropriate Spatial Resolution

Nicholas A. S. Hamm

Faculty of Science and Engineering, University of Nottingham, Ningbo, China - nicholas.hamm@nottingham.edu.cn

Keywords: Spatial Resolution, Pixel Size, ELSA, Entropy, Local Indicator of Spatial Association (LISA).

Abstract

Geospatial data is available at increasingly finer spatial resolutions. However, such data can also be problematic because it may result in increased financial, resource and time cost. It might also be unnecessary if the phenomenon of interest does not vary at this scale or if the scientific application does not require it. This paper explores the scale effects associated with increased spatial resolution using the entropy-based local indicator of spatial association (LISA-ELSA). Results are illustrated for two datasets a digital elevation model (30 m ASTER GDEM) and for a raster PM2.5 air pollution map (1000 m). It is shown that increasing the size of the local window for the LISA can be used to explore the effect of reducing the spatial resolution. It is possible to identify areas where the spatial association is invariant with increasing window size and which could be represented with coarser pixels.

1. Introduction

Spatial resolution is possibly the most widely quoted characteristic for geospatial datasets. However, producing fine spatial resolution datasets (i.e., small pixel size) can also be difficult and problematic. In order to achieve a finer resolution more data is required, leading to increased financial, resource and time costs. The resulting dataset may also be more difficult to work with. A fine spatial resolution may also not be necessary if the phenomenon of interest does not vary at this scale or if the scientific application does not require it. Hence there is an interest in methods to evaluate different spatial resolutions and to identify the appropriate ones. Published approaches include those based on entropy (Kodl et al., 2024), spatially stratified heterogeneity (Guo et al., 2022) and the Pareto boundary (Waldner et al., 2018).

Predictive mapping of environmental phenomena aims to produce a map with complete spatial coverage based on measurements of the target variable that are sparse in space. This may be based on (i) statistical or machine-learning models that predict the target variable based on covariates that have complete spatial coverage, (ii) interpolation between the sparse observations, e.g., using geostatistics, or (iii) some combination of both. Hence the sample density of the target variable, the spatial resolution of the covariates and the model prediction accuracy may all affect that spatial resolution that can be achieved. Therefore spatial data quality is also a concern.

In order to identify the appropriate spatial resolution we first need to understand the effects of decreasing the spatial resolution. The research presented here evaluates the effect of reducing the spatial resolution using the entropy-based local indicator of spatial association (ELSA) (Naimi et al., 2019). The use of ELSA to help identify appropriate spatial resolutions is then discussed.

2. Methods

The entropy-based local indicator of spatial association (ELSA) was developed by Naimi et al. (2019) and further illustrated by Hamm et al. (2024) and Zheng et al. (2024). It is a local-indicator of spatial association (LISA) as described by Getis and Ord (1996), Anselin (1995) and others.

ELSA, $E_i(r)$, is defined as:

$$E_i(r) = E_{ai}(r) \times E_{ci}(r) \tag{1}$$

and is calculated within a local neighbourhood centred on location, i, and ω_{ij} is a weight for an adjacent feature at location j. The local neighbourhood (r) is typically defined by a radius, adjacency matrix or moving window The weight ω_{ij} takes a value of 1 when the feature is within the local neighbourhood and zero otherwise.

$E_{ai}(r)$ is the dissimilarity within the local neighbourhood

$$E_{ai}(r) = \frac{\sum_j \omega_{ij} d_{ij}}{\max d \sum_j \omega_{ij}}, j \neq i \tag{2}$$

where d_{ij} is the dissimilarity between two categories or numbers at locations i and j and $\max d$ is the maximum dissimilarity in the data. The calculation of dissimilarity will depend on the data. For continuous data it would be the difference between the attribute values at i and j, but for categorical data it depends on the hierarchical structure and ordering of the categories.

$E_{ci}(r)$ is the composition or diversity, evaluated using the Shannon entropy:

$$E_{ci}(r) = -\frac{\sum_{k=1}^{m_\omega} p_k \log_2 p_k}{\log_2 m_i}, j \neq i \tag{3}$$

where m is the total number of categories in the entire dataset and m_ω is the actual number of categories within r. The m_i is the maximum possible number of categories within r and $m_i = m$ if $\sum \omega_{ij} > m$ or $\sum \omega_{ij}$ otherwise. The p_k indicates the probability of obtaining category k from the m_ω categories within r.

The presentation above makes clear that $E_i(r)$ can be applied to categorical data; however, it can also be applied to continuous data by dividing the data into ordered categories in a manner designed to minimize the information loss (Naimi et al., 2019). There is a R package `elsa`, which is available on CRAN (https://cran.r-project.org/package=elsa) and can be used to implement ELSA. This was used for the experiments presented in this paper.

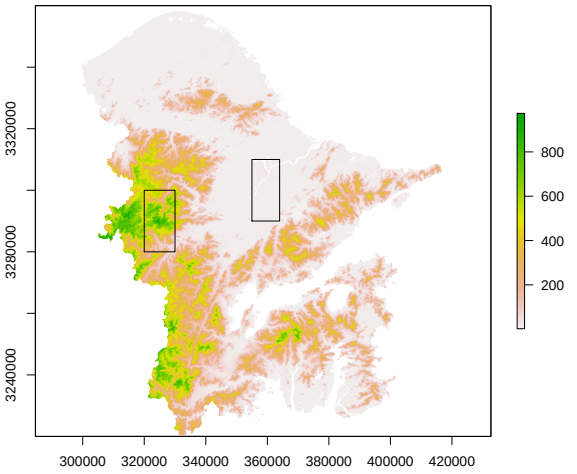

Figure 1. ASTER GDEM (version 3) 30 m digital elevation model (DEM) for Ningbo, Zhejiang, China (units: m a.s.l., UTM Zone 51N). The two boxes are subsets covering the plain (right) near the urban area and the hills (left).

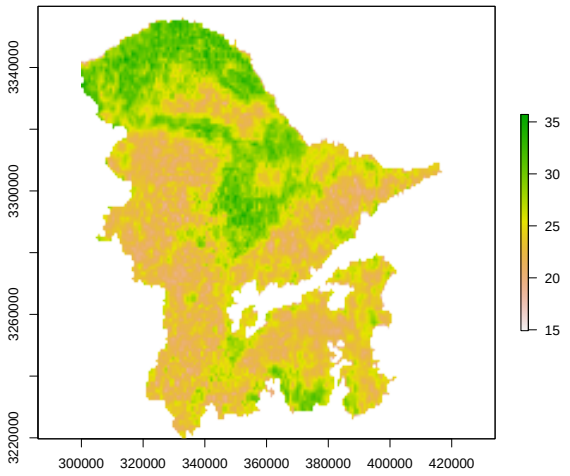

Figure 2. Annual PM2.5 concentration for 2020 (Chi et al., 2023) at 1000 m spatial resolution (units: μg m^{-3}, UTM Zone 51N). The areas covered is the same as in Figure 1.

It is proposed that the change in ELSA with decreased spatial resolution depends on the spatial structure in the data, the window size as well as data quality. This is expected from theory (Hamm et al., 2024; Naimi et al., 2019) and this study further elaborates it through experimental results. This paper presents structured experiments based on datasets of elevation and PM2.5 air quality that cover different landscapes, pixel sizes and window sizes. Hence evaluating ELSA for increasing window sizes would allow us investigation of the effect of aggregation or ’upscaling’ and to identify areas with high spatial association at different scales.

3. Data

This study used publically available data for (i) elevation (ASTER GDEM version 3) ASTER Science Team (2019) (`https://doi.org/10.5067/ASTER/ASTGTM.003`) (Figure 1) and (ii) air quality (PM2.5) (Chi et al., 2023) (Figure 2) for Ningbo in Zhejiang Province, People’s Republic of China (Tang et al., 2015). Ningbo (total population 9.6 million) is made up of an urban centre (population 2 million) and surrounding urban and rural areas. It is made up of six districts, two rural southern counties and two county-level cities (Cixi and Yuyao). The central urban area is on a flat plain bordered to the north by the sea and is surrounded by hills. The altitude ranges from sea-level to approximately 1000 m, as shown in Figure 1. The air pollution data is annual PM2.5 data for 2020, shown in Figure 2. Two subsets were identified covering the plains and hilly areas (Figure 1).

The steps are given as follows:

1. Calculate ELSA for the base pixel size u and a window size with a radius of $r = 2u$. For the air pollution example the base pixel size is $u = 1000$ (1000 m $\times$ 1000 m) and the radius, $r = 2000$ m. For the DEM example $u = 30$ m (30 m $\times$ 30 m)and the radius is $r = 60$ m.
2. Calculate ELSA for increasing window sizes, $r = 4u, 6u, 8u, \ldots$. This evaluated the spatial association within progressively larger local windows. The logic is that this can help to identify the scale of spatial variation (Getis and Ord, 1996), which supports the identification of suitable spatial resolutions.
3. Evaluate the maps, histograms and tabulated results. Note that the ELSA value is dependent on the data so the results need to be interpreted for each dataset specifically. The aim here is to identify patterns related to the increased window size and associated spatial scale.
4. Aggregate (upscale) the pixel size by a factor of 2, 3 and 4 and present the results. For the DEM example this leads to a pixel size of 60, 90 and 120 m.
5. Evaluate ELSA for the new pixel sizes.

4. Results and discussion

In this section the results of experiments for the DEM dataset and PM2.5 air pollution dataset are presented.

4.1 Digital elevation model (DEM)

The ELSA calculations are shown in Figure 3. For $r = 2u$ and $r = 4u$ the maps show a very high degree of spatial association, although this decreases for larger values of r. Hence, even in the hilly areas, the spatial association is high for small r whereas for larger r there could be more variability within the local window. The result is different in the low-lying flat areas in the north and north east where the central urban area is located. That shows consistently high spatial association across all window sizes. In Figure 4 the pixels were aggregated to larger window sizes (30, 60, 90, 120m) while maintaining $r = 2u$. A similar pattern is observed to Figure 4.

The ELSA values from Figure 3 were divided into bins of width 0.01. The proportion of ELSA values in each bin is shown in

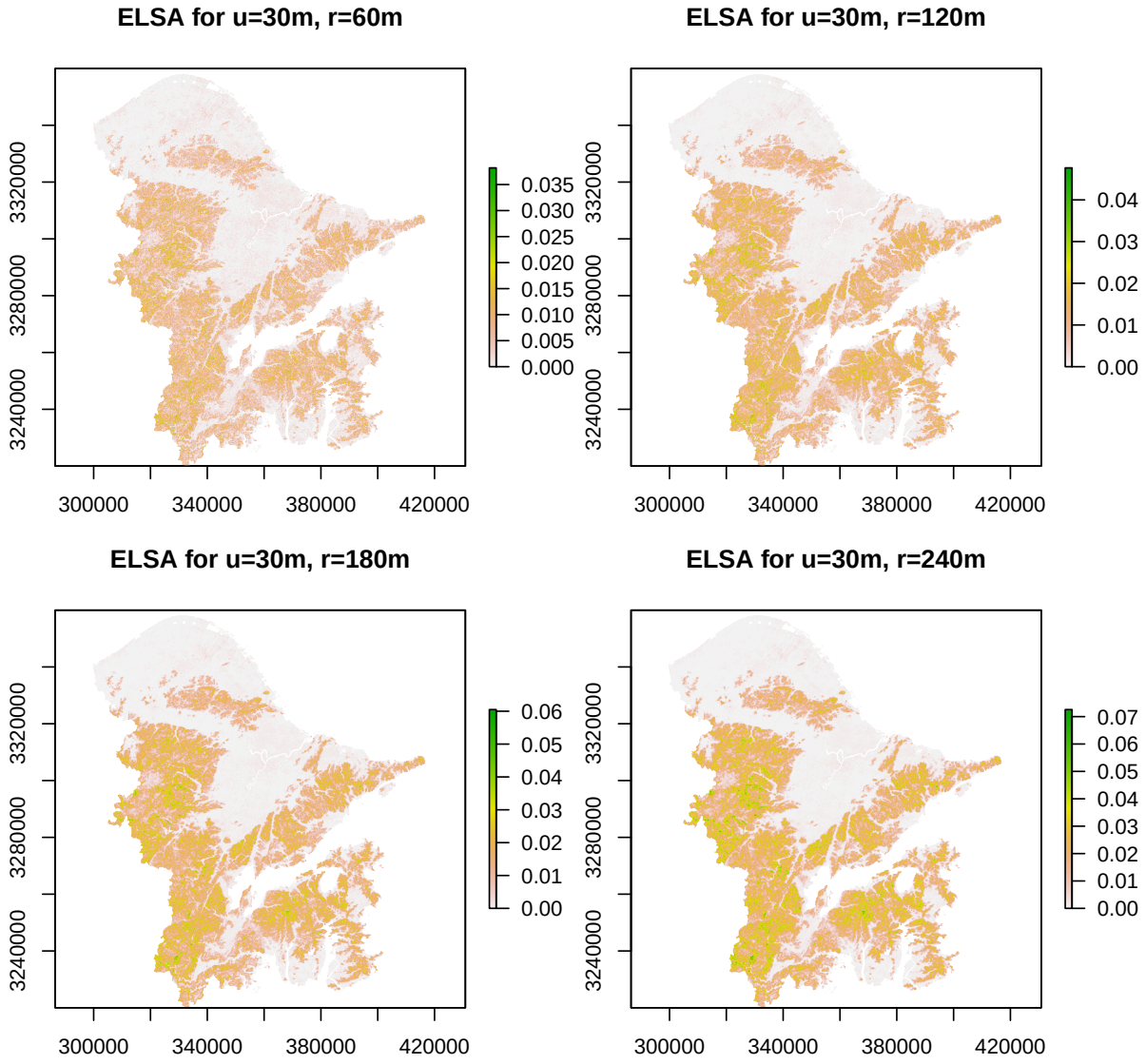

Figure 3. ELSA for the DEM (Figure 1) for fixed $u = 30$m and increasing window size, $r = 2u, 4u, 6u, 8u$.

ELSA	60	120	180	240	Plain	Hills
0.01	88.2	62.7	54.3	50.7	99.9	9.5
0.02	11.1	23.2	16.9	12.5		16.7
0.03	0.5	11.4	16.2	14.2		23.3
0.04		2.2	8.8	11.6		20.9
0.05		0.2	2.9	6.9		15.4
0.06			0.5	2.7		8.9
0.07			0.1	0.8		4.2
0.08				0.2		0.9
0.09				0.1		0.1
0.10						
NA	0.0	0.2	0.3	0.4	0.0	0.0

Table 1. Percentage ELSA values in different bins for $u = 30$m and $r = 2u, 4u, 6u, 8u$ (60m, 120m, 180m, 240m) for the DEM dataset (Figure 1). For $r = 8u$ the ELSA distribution for plain and hilly areas is shown (see Figure 1). A blank means that $< 0.1\%$ of ELSA values fall in this bin.

Table 1. There is a clear shift towards larger ELSA values as r increases. This reflects the increased dissimilarity and diversity within the local window. Table 1 also shows the results for $r = 4u$ for the flat and hilly areas illustrated in Figure 1. Even for the largest window size the ELSA values remained very low indicating high spatial association and suggesting that these areas could be represented using larger pixels. The result is different for the hilly areas which show increased ELSA, due to increased dissimilarity and diversity, for larger window sizes.

4.2 PM2.5 air pollution

The results for the air pollution data are shown in Figure 5 and Table 2. The range of ELSA values is larger than for the DEM example. The heterogeneity across the study area is also less clear, in particular for lower values of r – possibly reflecting the fact that the PM2.5 data appears to be more noisy. For larger values of r we see clearer patterns with higher spatial association (low ELSA) over the hills and lower spatial association in the plains. Hilly areas are predominantly rural whereas the plains include urban areas, roads, industry and a major port. It is possible that the air pollution over rural areas is mainly background pollution whereas this is augmented by local sources on the plains, which show higher heterogeneity. It should also be noted that there are large flat areas (e.g., the north-west coast) which show high PM2.5 concentration (Figure 2) and also high spatial association. Hence both high and low pollution areas can exhibit high spatial association.

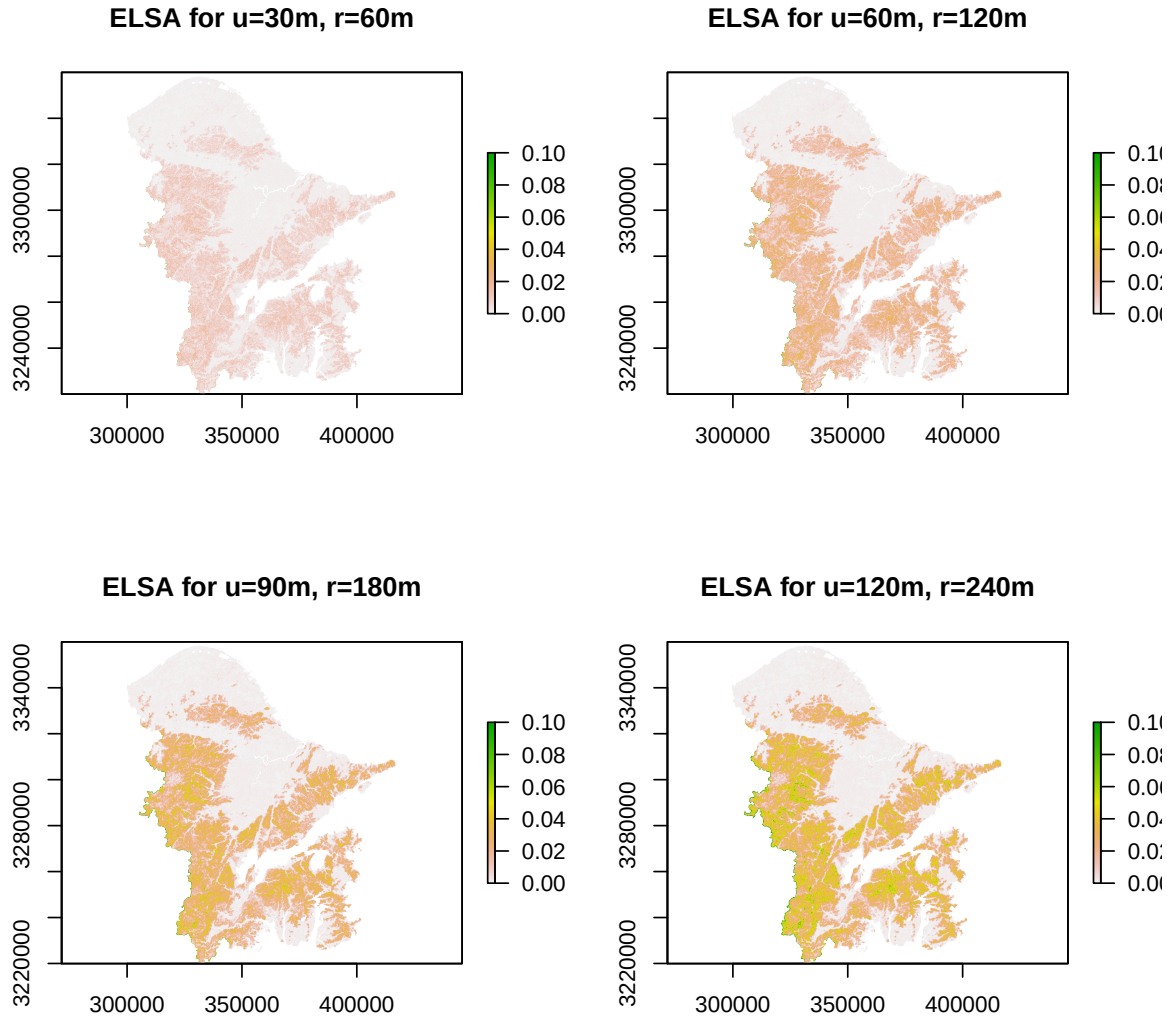

Figure 4. ELSA for the DEM (Figure 1) for aggregated pixels, 30, 60, 90, 120 m and consistent $r = 2u$. Note that the legend is the same for all four maps.

ELSA	2000	4000	6000	8000	Plain	Hills
0.02	24.3	4.5	1.3	0.1		
0.04	39.2	30.0	20.8	13.8		54.5
0.06	21.2	39.2	29.2	26.5	4.7	34.4
0.08	8.4	17.2	20.5	22.2	33.7	7.0
0.10	3.3	9.6	13.2	15.5	24.3	2.3
0.12	1.4	4.6	6.8	9.4	22.8	1.7
0.14	0.6	2.1	3.6	5.4	6.5	
0.16	0.5	0.9	2.1	2.9	3.6	
0.18	0.3	0.5	0.9	1.7	4.0	
0.20	0.3	0.3	0.6	0.9	0.4	
> 0.20	0.6	1.0	1.1	1.6		

Table 2. Percentage ELSA values in different bins for $u = 1000$m and $r = 2u, 4u, 6u, 8u$ (2000m, 4000m, 6000m, 8000m) for the air pollution data (Figure 2). For $r = 8u$ the ELSA distribution for plain and hilly areas is shown (see Figure 1). A blank means that $< 0.1\%$ of ELSA values fall in this bin.

5. Conclusions

The research presented in this paper used the ELSA statistic to quantify local spatial association for elevation and air pollution (annual PM2.5) in the city of Ningbo in Zhejiang Province, China. By calculating and mapping ELSA the heterogeneity in the spatial association across the study area could be evaluated. It is proposed that this could be used to help identify the appropriate spatial resolution for a given environmental variable. For the DEM dataset the high level of spatial association (low ELSA) over the plains for $r = 4u$ (Figure 3, Table 1) as well as for large pixel sizes ($4u$) suggested that a pixel size of 120 m or larger would be sufficient. Over the hilly areas the rapidly changing relief meant that a finer resolution was required to represent the topography. The ELSA statistic allows to identify these areas. Depending on the application, and on the location, the spatial resolution could be adjusted for different areas.

The PM2.5 dataset was more difficult to evaluate. The dataset was already provided at a relatively coarse resolution of 1000m $\times$ 1000 m (Chi et al., 2023) and the data appears to be noisy. This conclusion is made based on the salt and pepper effect in Figure 2 and on the mottled appearance of Figure 5 (top left). Nevertheless the results did indicate a higher degree of spatial association over the hills, in particular for larger values of r and also for some flat areas. The tentative conclusion is that

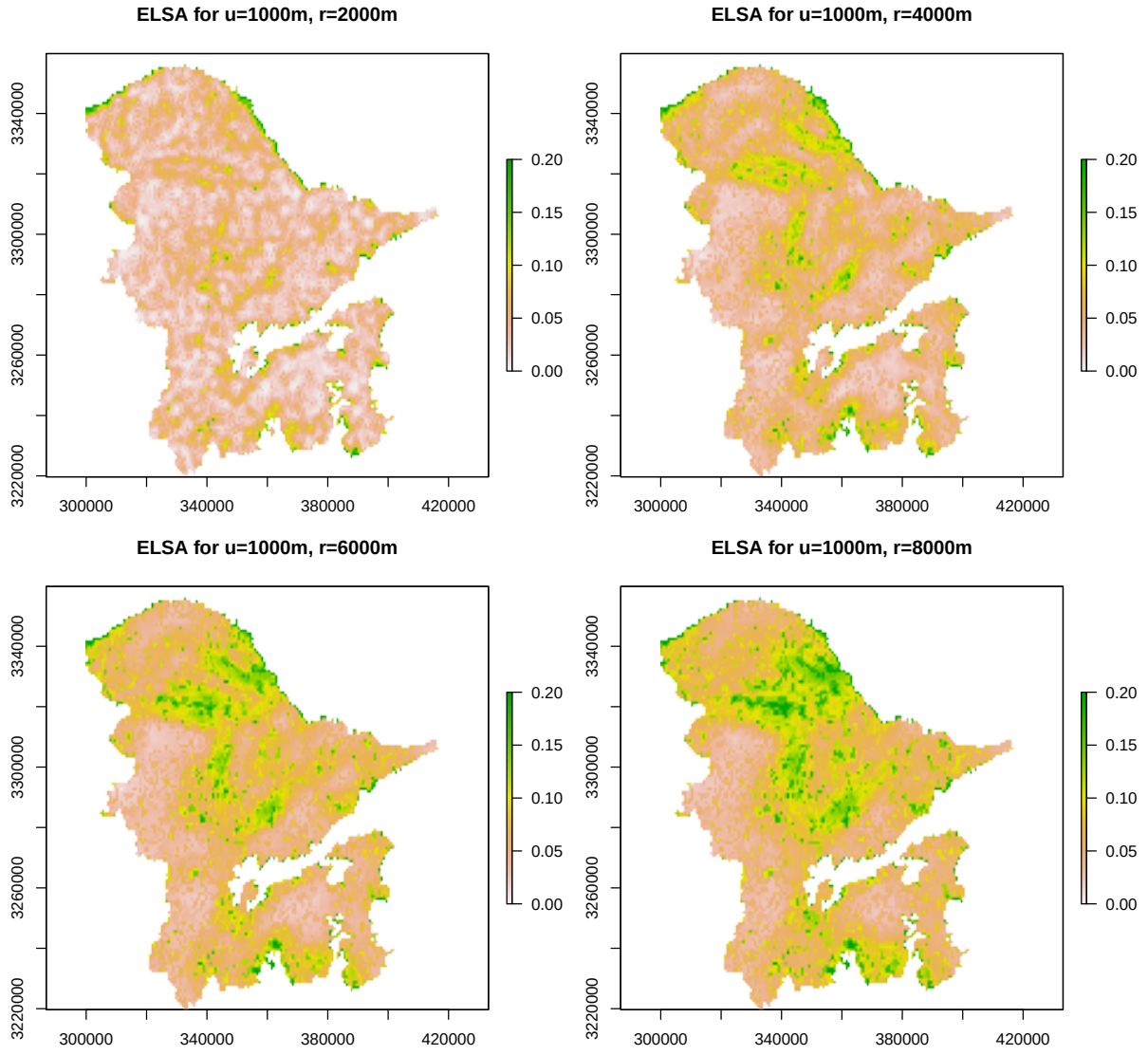

Figure 5. ELSA for the PM2.5 dataset (Figure 2) for fixed $u = 1000$m and increasing window size, $r = 2u, 4u, 6u, 8u$. Note that the legend is the same for all four maps.

these areas could be represented by coarser spatial resolutions. Possibly of greater interest is what these patterns might inform us about PM2.5 air pollution and how it might relate both to the physical and urban environment and to pollution sources. Although not directly related to the original research objective, it does suggest that evaluating scale can help to gain insight into environmental data.

Further research will aim to establish whether thresholds can be identified to support representing some areas by coarser resolution pixels. Other steps will be to link to the idea of spatially stratified heterogeneity and to address the multidimensional issue of data quality, including noise, accuracy and spatial sampling density. Hence, an important task is to evaluate the spatial resolution that can be obtained considering spatial data quality.

Acknowledgements

This research was supported by the Ningbo Natural Science Foundation (Project ID 2022J173) and also by the Ningbo Commonweal Science and Technology Planning Project (Grant No. 2021S081).

References

Anselin, L., 1995. Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93-115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.

ASTER Science Team, 2019. *ASTER Global Digital Elevation Model V003*. NASA EOSDIS Land Processes Distributed Active Archive Center.

Chi, Y., Zhan, Y., Wang, K., Ye, H., 2023. Spatial Distribution of Multiple Atmospheric Pollutants in China from 2015 to 2020. *Remote Sensing*, 15(24). https://doi.org/10.3390/rs15245705.

Getis, A., Ord, J. K., 1996. Local spatial statistics: an overview. P. A. Longley, M. Batty (eds), *Spatial Analysis: Modelling in a GIS Environment*, Taylor and Francis, Oxford, 261–277.

Guo, J., Wang, J., Xu, C., Song, Y., 2022. Modeling of spatial stratified heterogeneity. *GIScience & Remote Sensing*, 59(1), 1660-1677. https://doi.org/10.1080/15481603.2022.2126375.

Hamm, N. A. S., Naimi, B., Groen, T. A., 2024. ELSA: a new local indicator for spatial association. *29th Annual GIS Research UK (GISRUK) Conference*. https://doi.org/10.5281/zenodo.4665865.

Kodl, G., Streeter, R., Cutler, N., Bolch, T., 2024. Arctic tundra shrubification can obscure increasing levels of soil erosion in NDVI assessments of land cover derived from satellite imagery. *Remote Sensing of Environment*, 301, 113935. https://doi.org/10.1016/j.rse.2023.113935.

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., Toxopeus, A. G., Alibakhshi, S., 2019. ELSA: Entropy-based local indicator of spatial association. *Spatial Statistics*, 29, 66-88. https://doi.org/10.1016/j.spasta.2018.10.001.

Tang, Y.-T., Chan, F. K. S., Griffiths, J., 2015. City profile: Ningbo. *Cities*, 42, 97-108. https://doi.org/10.1016/j.cities.2014.10.001.

Waldner, F., Duveiller, G., Defourny, P., 2018. Local adjustments of image spatial resolution to optimize large-area mapping in the era of big data. *International Journal of Applied Earth Observation and Geoinformation*, 73, 374-385. https://doi.org/10.1016/j.jag.2018.07.009.

Zheng, Y., Lin, T., Hamm, N. A. S., Liu, J., Zhou, T., Geng, H., Zhang, J., Ye, H., Zhang, G., Wang, X., Chen, T., 2024. Quantitative evaluation of urban green exposure and its impact on human health: A case study on the 3–30-300 green space rule. *Science of The Total Environment*, 924, 171461. https://doi.org/10.1016/j.scitotenv.2024.171461.