Improvements to Information Entropy for Raster Spatial Data: A Thermodynamic-based Evaluation

Hong Zhang [a, *], Peichao Gao [b], Zhilin Li [c]

[a] *Faculty of Geosciences & Environmental Engineering, Southwest Jiaotong University, Chengdu, Sichuan, China;* zhangh2011@swjtu.edu.cn
[b] *Faculty of Geographical Science, Beijing Normal University, Beijing, China;* gaopc@bnu.edu.cn
[c] *Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong;* zl.li@polyu.edu.hk

* Corresponding author

Keywords: Shannon Entropy; Information Entropy; Raster Data; Information Content

Abstract:

Spatial information is fundamentally important to our daily life. It has been estimated by many scholars that almost 80 percent or more of all information in this world are spatially referenced and can be regarded as spatial information. Given such importance, a discipline called spatial information theory has been formed since the late 20th century. In addition, international conferences on spatial information have been frequently held. For example, COSIT (Conference on Spatial Information Theory) was established in 1993 and are held every two years all over the world.

In spatial information theory, one fundamental question is how to measure the amount of information (i.e., information content) of a spatial dataset. A widely used method is to employ entropy, which is proposed by the American mathematician Claude Shannon in 1948 and usually referred to as Shannon entropy or information entropy. This information entropy was originally designed to measure the statistical information content of a telegraph message. However, a spatial dataset such as a map or a remote sensing image contains not only statistical information but also spatial information, which cannot be measured by using the information entropy.

As a consequence, considerable efforts have been made to improve the information entropy for spatial datasets in either a vector format of a raster format. There are two basic lines of thought. The first is to improve the information entropy by defining how to calculate its probability parameters, and the other is to introduce new parameters into the formula of the information entropy. The former results in a number of improved information entropies, while the latter leads to a series of variants of the information entropy. Both seem to be capable of distinguishing different spatial datasets, but there is a lack of comprehensive evaluation of their performance in measuring spatial information.

This study first presents a state-of-the-art review of the improvements to the information entropy for the information content of spatial datasets in a raster format (i.e., raster spatial data, such as a grey image and a digital elevation model). Then, it presents a comprehensive evaluation of the resultant measures (either improved information entropies or variants of the information entropy) according to the Second Law of Thermodynamics. A set of evaluation criteria were proposed, as well as corresponding measures. All resultant measures were ranked accordingly.

The results reported in this study should be useful for entropic spatial data analysis. For example, in image fusion, a crucial question is how to evaluate the performance of a fusion algorithm. This evaluation is usually achieved by using the information entropy to measure the increase in the information content during the fusion. It can now be performed by the best-improved information entropy reported in this study.