

## 顾及地图内容表达稳定性的时序地理信息统计数据分级评价方法

An approach for classification evaluation of geographic information data considering the stability of map contents

**Abstract:** Evaluation on geographic information data classification is important for their based applications and visualizations. But existing evaluation approaches only consider the statistical characteristics and spatial distribution characteristics of the classification results, ignore their temporal characteristics. These approaches are not suitable for the evaluation of spatio-temporal data classification. This paper first proposed an approach to evaluate the classification result considering their statistical characteristics, spatial distribution characteristics and temporal characteristics together. First, the thematic map information is introduced to quantify statistical characteristics and spatial distribution characteristics of the classification results at a specific time. Second, two parameters describing the correlation of the thematic map information and the consistency of the map contents at different times are defined. Finally, the three parameters are combined to evaluate classification results of spatio-temporal data. The COVID-19 deaths of the United States month by month is selected for experiment. The classification results of the experiment data obtained by seven commonly used map classification approaches are evaluated with proposed approach. Experiments show that the proposed approach can indicate the best classification approach according to the needs of different map content stability, and meet the needs of practical applications. It also can analyze the differences in map contents at different times.

**Keywords:** Time series data; geographic information data classification; cartographic information; temporal correlation; choropleth maps.

**摘要:** 合理的分级结果评价是地理信息统计数据分级展示的前提。已有评价方法多关注分级结果的统计特征和空间分布特征,忽略了其时序特征,不适宜时序地理信息统计数据分级结果的评价。本文综合考虑分级结果的统计特征、空间分布特征和时序特征,评价分级结果的合理性。首先,利用专题地图信息量指标量化特定时刻分级结果的数据统计特征和空间分布特征;其次,构建专题地图信息量的时间相关性和地图内容表达一致性指标,评价不同时刻数据分级结果随时序变

化的稳定性；最后，综合专题地图信息量、专题地图信息量时间相关性和地图内容表达一致性评价时序地理信息统计数据分级结果的合理性。选择美国新冠疫情数据逐月死亡人数，基于本文方法评价 7 种常用地图分级方法获取的分级结果，实验表明本文方法能依据不同地图内容表达稳定性的需要指示出最优分级方法，满足实际应用需求；同时，能分析不同时刻地图内容表达差异。

**关键字：**时序统计数据；地理信息数据分级；地图信息量；时间相关性；质底图。

## 0. 引言

抽象和概括是科学认知的基本手段<sup>[1]</sup>。地理信息统计数据分级是地理信息抽象表达的重要方式，能有效辅助认知地理现象的时空发展规律，广泛应用于专题地图制图等 GIS 领域<sup>[2]</sup>。目前已发展有较多成熟的地理信息统计数据分级方法，如等差分级、等比分级、标准差分级、分位数分级、自然裂点分级、基于优化方法的裂点分级、头尾分割分级和顾及数据不确定性的启发式分级等<sup>[2-7]</sup>。不同数据分级方法适用于不同时空分布特征的数据，依据分级结果的时空特征评价不同数据分级方法，辅助分级方法的选择是当前地理信息统计数据分级研究的重点。

统计数据分级需在最大限度保留原始数据主要特征的同时，尽可能准确地反映地理现象的时空分布<sup>[8]</sup>。因此，基于分级结果的统计特征和时空分布特征能评价分级结果的优劣。基于分级结果统计特征，可依据分级结果的级内一致性、级间差异性和级间均匀性等进行评价<sup>[9]</sup>；基于分级结果空间分布特征，可以依据分级结果的区域边界误差、地理面积均等程度和空间自相关系数等进行评价<sup>[10]</sup>。同时，不同指标可以组合表达，如文献[11]和文献[12]分别利用了 3 种和 5 种统计特征指标的线性组合评价分级结果；文献[13]则综合了统计特征指标和空间分布特征指标评价癌症数据的分级结果。面向特定应用，文献[14]研究了流行病数据制图中的分级结果评价；文献[15]研究了人口密度分级统计图中的分级结果评价，文献[16]研究了不同分级方法相对于数据不确定性的敏感性。地图表达的核心功能是传输空间信息，可以基于地图信息量评价分级结果。文献[2]依据级间观测数量差异构建了分级结果评价的信息熵模型；文献[17]依据级内数据聚集水平和级间数据离散水平构建了分级结果评价的相对指数熵模型；文献[18]则提出了能同时顾及地图统计特征和空间分布特征的地图几何信息量和地图专题信息量等指

标。

但是,上述分级结果评价方法只考虑了分级结果的统计特征和空间分布特征,忽略了其时间特征。随着城市化的飞速发展,具有时空维度的地理大数据呈现爆炸式增长,催生了大量面向时序地理信息统计数据的分级表达任务,如新冠疫情的逐周、月演化,通常是利用分级统计图或质底法表达特定时刻的分级结果,并在此基础上利用时间轴、邮票法和动画等形式表达疫情的动态变化<sup>[19]</sup>。一方面,地理现象随时间的演化往往表现出较强的时间相关性,邻近时刻的分级结果需要有效保持这种相关性;另一方面,需要保证邻近时刻地图内容表达的一致性,即强调邻近时刻变化的同时,需保证邻近时刻地图内容不宜差异过大<sup>[20]</sup>。因此,本文试图综合考虑分级结果的统计特征、空间分布特征和时间特征评价分级结果,而不只是考虑分级结果的统计特征和空间分布特征。

## 1. 方法

给定包含  $N$  个时刻的时序标签统计数据样本集合,记为  $TS=\{subTS_1,subTS_2,...,subTS_n,...subTS_N\}$ ,  $subTS_n$  表示时刻  $T_n$  的样本集合,记为  $subT_n=\{x^n_1,x^n_2,...,x^n_m\}$ 。利用分级方法获取  $TS$  的一个有序划分,记为  $D=\{d_1,d_2,...,d_k\}$ ,  $\bigcup_{d_k \in D} d_k = TS \wedge \bigcap_{d_k \in D} d_k \neq \emptyset$ ;依据  $D$  可获得  $T_n$  时刻样本集合  $subTS_n$  的有序划分  $subD_n=\{d_{n1},d_{n2},...,d_{nk}\}$ 。依据前文分析,可依据分级结果的统计特征、空间分布特征和时间特征评价分级效果;其中,专题地图信息量可有效描述分级结果的统计特征和空间分布特征,即评价  $T_1-T_N$  时刻样本集合在划分  $subD_1-subD_N$  下的专题地图信息量;其次,评价分级结果的时间特征,如邻近时刻地图内容表达的稳定性,包括专题地图信息量的相关性和地图内容表达一致性。因此,本文分级结果评价包括:(1)  $T_n$  时刻的专题地图信息量;(2)  $T_1-T_N$  时刻专题地图信息量的时间相关性;(3)  $T_n$  和  $T_{n+1}$  时刻间的地图内容表达一致性;(4) 综合评价。

### 1.1 专题地图信息量评价

地理信息统计数据的分级通常服务于相应的制图表达任务,不同地图区域对应的分级数值可以利用不同符号表示,如以色相区分<sup>[21]</sup>。专题地图信息量基于概率频次,利用不同类符号在空间分布上的数量差异度量地图信息<sup>[2,18]</sup>。假设地图区域数量为  $M$ ,  $T_n$  时刻对应的统计数据划分为  $subD_n=\{d_{n1},d_{n2},...,d_{nk}\}$ ;区域  $M_i$  的

一阶邻近区域数量为  $C_i$ ,  $M_i$  一阶邻近区域中对应划分为  $d_{nj}$  的区域数量为  $subC_{ij}$ , 则区域  $M_i$  一阶邻近区域中对应划分  $d_{nj}$  的区域频率  $P_{ij}$  见式(1)。

$$P_{ij} = \frac{subC_{ij}}{C_i} \quad (1)$$

累加不同区域对应不同划分的区域频次表达的地图专题信息即为地图专题信息量。其中, 若采用经典的香农熵计算地图专题信息量, 会导致无穷大问题。参考文献[17], 本文利用对数熵计算  $T_n$  时刻分级结果的专题地图信息量, 见式(2)。

$$H_{T_n} = \frac{\sum_{i=1}^M \sum_{j=1}^k \frac{subC_{ij}}{C_i} \times e^{\frac{1-subC_{ij}}{C_i}}}{M} \quad (2)$$

## 1.2 专题地图信息量时间相关性评价

地理现象随时间和空间上的演变往往呈现出较强的相关性, 表现为邻近时刻或邻近空间的数据呈现出一定的依赖关系。描述空间自相关的方法有很多, 如 Moran's I、Geary's C 和 Getis 等, 其核心是度量邻近区域数据间的依赖关系<sup>[23]</sup>。因此, 利用空间自相关原理可以度量邻近时刻数据间的依赖关系。其中, Moran's I 指数应用广泛, 本文利用全局 Moran's I 指数表示专题地图信息量的全局时间相关性。其中, 邻近时刻的定义需要考虑统计数据可能存在的周期性变化, 如气象数据。依据全局 Moran's I 指数计算原理, 专题地图信息量的全局时间相关性( $I_G$ )定义见式(3)。

$$I_G = \frac{N}{S_0} \times \frac{\sum_{i=0}^N \sum_{j=0}^N w_{ij} (H_{T_i} - \bar{H})(H_{T_j} - \bar{H})}{\sum_{i=0}^N (H_{T_i} - \bar{H})^2} \quad (3)$$

$N$  表示存在  $N$  个不同时刻的数据,  $S_0 = \sum_{i=0}^N \sum_{j=0}^N w_{ij}$ ,  $H_{T_i}$  和  $H_{T_j}$  表示  $i$  时刻和  $j$  时

刻的专题地图信息量,  $\bar{H} = \sum_{i=0}^n H_{T_i}$ ;  $w_{ij}$  为权重, 由  $i$  和  $j$  间邻近程度( $A_{ij}$ )确定,

$A_{ij}$  表示时刻  $T_i$  和  $T_j$  的时间间隔。其中,  $A_{ij}$  定义需考虑统计数据可能存在的周期性, 假设统计数据存在周期  $T$ ,  $A_{ij}$  定义见式(4)。

$$A_{ij} = \begin{cases} |i-j| & (|i-j| < T) \\ |i-j| \div T & (|i-j| \geq T \wedge |i-j| \% T = 0) \\ +\infty & (Else) \end{cases} \quad (4)$$

$T_i$  和  $T_j$  时刻越邻近, 则权重  $w_{ij}$  取值越大, 可依据高斯权重表示, 见式(5)<sup>[24]</sup>。

$$w_{ij} = \begin{cases} 0.5 & (A_{ij} = 0) \\ \int_{A_{ij}-1}^{A_{ij}} \frac{1}{\sqrt{2\pi}} e^{-\frac{|A_{ij}|^2}{2}} & (A_{ij} > 0) \end{cases} \quad (5)$$

另外, 考虑到地理现象周期性变化的强弱, 可为  $w_{ij}$  设置不同权重, 见式(6)。

$$w_{ij} = \begin{cases} 0.5 & (A_{ij} = 0) \\ w_{ij} \times \alpha_1 & (A_{ij} < T \wedge A_{ij} \neq 0) \\ w_{ij} \times \alpha_2 & (|i - j| \geq T) \end{cases} \quad (6)$$

其中,  $\alpha_1 + \alpha_2 = 1$ 。若统计数据无周期性,  $T = +\infty$ ,  $\alpha_2 = 0$ ; 若统计数据存在明显的周期性,  $\alpha_2$  可适当取较大值。

### 1.3 邻近时刻地图内容表达的一致性评价

时序地理信息的表达需要保证地图内容随时间变化的一致性, 即保持邻近时刻地图内容差异的同时, 该差异也不宜过大<sup>[20]</sup>。给定邻近时刻  $T_n$  和  $T_{n+1}$ , 基于  $T_n$  和  $T_{n+1}$  时刻分级结果的制图表达内容差异可以用  $T_n$  和  $T_{n+1}$  时刻同一区域分级数值变化的区域比( $Cv_n$ )表示, 定义见式(7)。

$$Cv_n = \frac{\sum_{i=0}^M AC_i}{M}$$

$M$  为地图区域个数,  $AC_i = \begin{cases} 1 & (v_p \neq v_q) \\ 0 & (v_p = v_q) \end{cases}$ ,  $v_p$  与  $v_q$  表示  $T_n$  和  $T_{n+1}$  时刻同一区域的分级数值。因此,  $T_n$  和  $T_{n+1}$  时刻地图内容表达一致性( $C_{Gn}$ )定义见式(8)。

$$C_{Gn} = 1 - \left| \frac{\sum_{i=0}^M Cv_n}{M} - \beta \right| \quad (8)$$

$\beta$  为常数,  $\beta \in [0, 1]$ , 可依据制图目的和数据特征设置, 如若需表达邻近时刻间地图内容差异,  $\beta$  可取较大值, 如  $\beta = 0.8$ ; 若需表达邻近时刻地图内容一致性,  $\beta$  可取较小值, 如  $\beta = 0$ 。

### 1.4 综合评价

分级结果需保证:  $T_1 - T_N$  时刻划分  $subD_1 - subD_N$  的专题地图信息量大, 可用

$\overline{H} = \sum_{i=0}^N H_{Ti}$  表示；同时，需保证专题地图信息量的时间相关性高，可用  $I_G$  表示；

另外，需保证邻近时刻地图内容表达一致性高，可用  $\overline{C_G} = \sum_{i=0}^N C_{Gi}$  表示。因此，分

级结果的评价需综合考虑  $\overline{H}$ 、 $I_G$  和  $\overline{C_G}$ 。其中， $\overline{H}$ 、 $I_G$  和  $\overline{C_G}$  存在一定的相关性，如  $\overline{H}$  增大时， $\overline{C_G}$  通常会降低。因此，分级结果的优劣程度  $G$  适宜表达为  $\overline{H}$ 、 $I_G$  和  $\overline{C_G}$  的乘数组合，定义见式(9)<sup>[22]</sup>。

$$G = \sqrt[3]{I_G \times \overline{C_G} \times \overline{H}} \tag{9}$$

## 2. 实验

### (1) 实验数据与分级方法

本文利用美国 2020/3/1-2021/9/1 逐月(共 19 个月，以 T\_0-T\_18 标识)、49 个州的新冠疫情死亡人数 (DC) 为实验数据，下载地址：<https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.html><sup>[25]</sup>。针对该数据的箱线图统计分析结果见图 1，由图可知：T<sub>1</sub>、T<sub>2</sub>时刻和 T<sub>9</sub>、T<sub>10</sub>时刻为 DC 的峰值，其余时刻则随时间变化 DC 呈现出明显的递增或递减趋势。尽管新冠疫情死亡人数受防疫政策、医疗条件等多方面因素的影响，该统计分析结果显示，新冠疫情的传播亦呈现出明显的随时间演化特征。利用等差分级、标准差分级、分位数分级、头尾分割分级和自然裂点分级(包含 3 个不同变种)共 7 种常用的统计数据分级方法将实验数据分为 5 级，分级结果见表 1<sup>[26]</sup>。

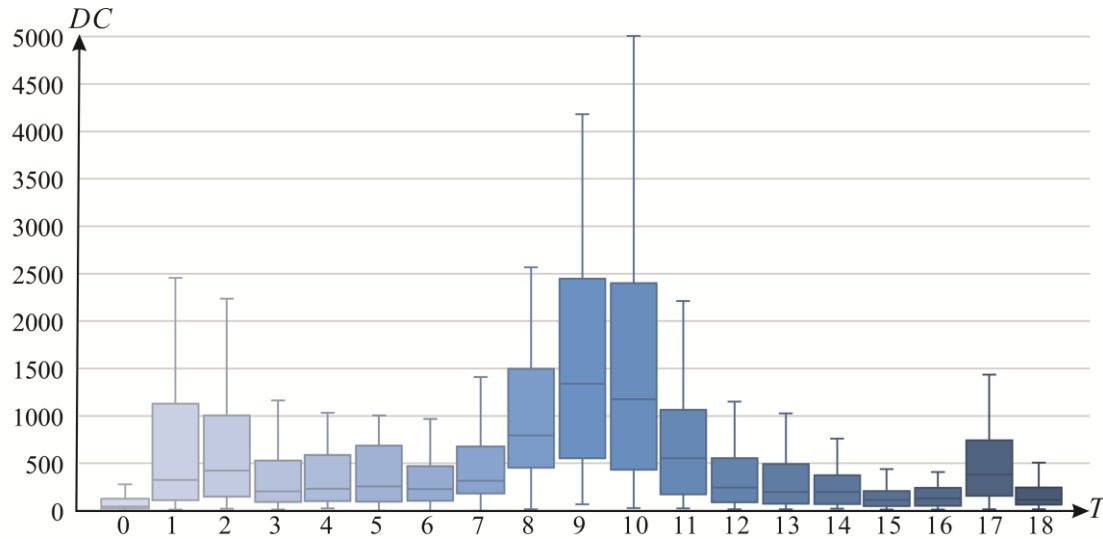


图 1.美国新冠疫情逐月(2020/3/1-2021/9/1，49 个州)死亡人数箱线图

表 1. 数据分级结果

分级方法	第 1 级	第 2 级	第 3 级	第 4 级	第 5 级
等差分级	0-4001	4001-8003	8003-12004	12004-16005	16005-20007
标准差分级	0-384	384-523	523-800	800-938	938-20007
分位数分级	0-52	52-164	164-332	332-879	879-20007
头尾分割分级	0-661	661-2018	2018-4146	4146-7551	7551-20007
自然裂点分级 I	0-719	719-2236	2236-5223	5223-10700	10700-20007
自然裂点分级 II	0-188	188-605	605-1477	1477-3502	3502-20007
自然裂点分级 III	0-765	765-2341	2341-5223	5223-10700	10700-20007

## (2) 实验结果与分析

利用本文方法对所有分级结果进行评价。实验参数设置如下：考虑到实验数据具有较强的时空演化特征，周期性特征相对不明显，公式(6)中参数设置为 $\alpha_1=0.9, \alpha_2=0.1$ ；同时，需一定程度上表达新冠疫情按月的变化，公式(8)参数设置为 $\beta=0.4$ 。实验结果见表 2，由表 2 可知，分位数分级具有较高的平均专题地图信息量；同时，其专题地图信息量的时间相关性和邻近时刻间地图内容表达一致性相对较高，故分位数分级是最佳分级方法。对比美国疾病控制与预防中心以州为单位的新冠疫情可视化(<https://coronavirus.jhu.edu/us-map>，应用 1)和美国约翰霍普金斯大学以县为单位的新冠疫情可视化([https://covid.cdc.gov/covid-data-tracker/#cases\\_casesper100klast7days](https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days)，应用 2)中采用的分级方法(表 3)可知，实际应用中新冠疫情数据的制图均采用了近似分位数分级的方法，与本文评价方法确定的最佳分级方法一致。

表 2. 分级结果统计分析

指标	等差 分级	标准差 分级	分位数 分级	头尾分割 分级	自然裂点 分级 I	自然裂点 分级 II	自然裂点 分级 III
$\overline{H} \uparrow$	1.038	1.419	1.615	1.312	1.279	1.559	1.264
$I_G \uparrow$	0.509	0.381	0.512	0.581	0.602	0.383	0.600
$\overline{CG} \uparrow$	0.634	0.915	0.938	0.837	0.813	0.989	0.802
$G \uparrow$	0.694	0.791	0.919	0.861	0.855	0.839	0.847

表 3. 实际应用中的数据分级结果

第 1 级	第 2 级	第 3 级	第 4 级	第 5 级	第 6 级	第 7 级	第 8 级
-------	-------	-------	-------	-------	-------	-------	-------

应用 1	0-67	67-156	156-209	209-269	269-369	369-754	754-	
应用 2	0-7	7-21	21-38	38-59	59-91	91-148	148-324	324-

其次，分位数分级结果下不同时刻制图表达的专题地图信息量( $H$ )和邻近时刻地图内容变化 ( $C_v$ )统计结果见图 2，依据分位数分级结果生成的时序可视化结果见图 3。由图可知，分位数分级方法获取的制图结果不同时刻的专题地图信息量均较大，平均值为 1.615；同时，不同时刻专题地图信息量(除时刻  $T_9$ )外均保持在 1.615 附近。其次，邻近时刻地图内容的变化平均值为 0.462，有效表达了邻近时刻地图内容随时间的变化，与设定的  $\beta=0.4$  接近，故邻近时刻地图内容表达一致性均值( $\overline{CG}=0.919$ )较大。另外，可以分析特定时刻专题地图信息量和地图内容表达的变化，如  $T_1$  时刻相对于  $T_0$  时刻  $C_v$  较大，表现为 2020/4 相对于 2020/3 新冠疫情死亡人数呈现出明显增长。

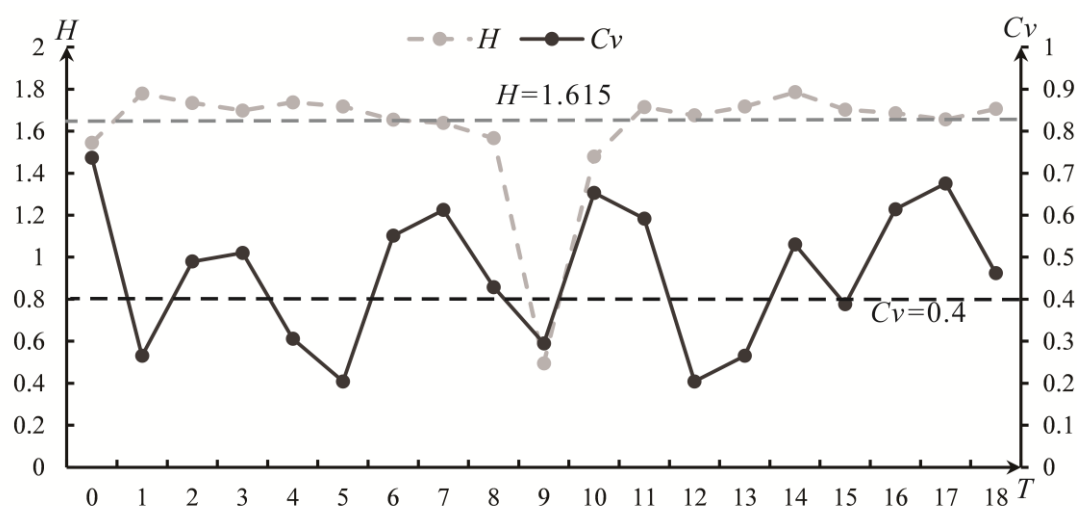


图 2. 不同时刻制图表达的专题地图信息量( $H$ )和邻近时刻地图内容变化 ( $C_v$ )统计结果



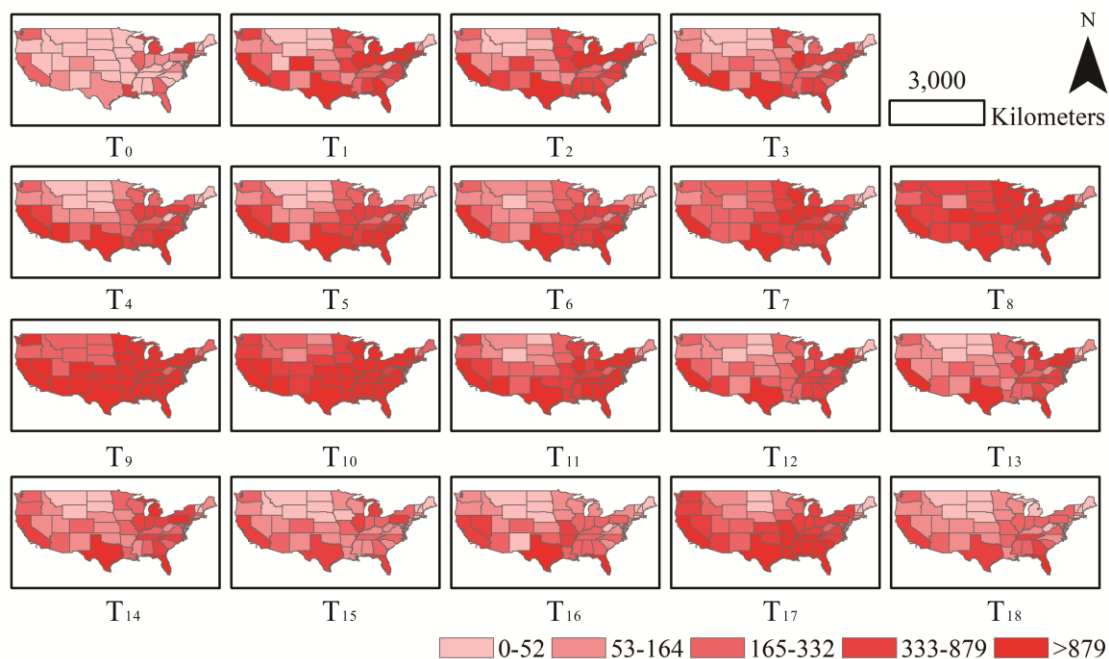


图 3. 基于分位数分级的美国新冠疫情逐月(2020/3/1-2021/9/1, 49 个州)死亡人数时序制图

### 3. 讨论

#### 3.1 对比分析

为验证本文方法有效性,对比只利用专题地图信息量评价统计数据分级结果的方法,见表 2。由表 2 可知,若制图表达需一定程度上保持邻近时刻间地图内容变化,即公式(8)中  $\beta=0.4$ ,利用专题地图信息量评价分级结果与本文方法评价结果一致,均认为分位数分级结果最佳。但是,若制图表达需保证邻近时刻地图内容变化小,如若制图结果能较少反映邻近时刻因疫情死亡人数的增长可以有效缓解公众恐慌,即公式(8)中  $\beta=0$ ,结果见表 4。利用专题地图信息量评价分级结果仍认为分位数分级方法为最佳,而利用本文方法评价分级结果则认为自然裂点分级 I 或自然裂点分级 III 方法最佳,本文方法更符合制图表达需要。由上述对比可知,本文方法相比已有统计数据分级评价方法,能更有效地顾及时序统计地理信息数据制图过程中不同地图内容表达一致性的需要。

另外,若取公式(8)中  $\beta=0.4$ ,见表 2,在只考虑专题地图信息量的时间相关性前提下,自然裂点分级 I 方法为最佳分级方法;但是,其平均专题地图信息量较低,故其实际上不是最佳分级方法。若取公式(8)中  $\beta=0$ ,见表 4,在只考虑邻近时刻地图内容表达一致性前提下,等差分级方法为最佳分级方法;但是,其平均专题地图信息量较低,故其实际上亦不是最佳分级方法。因此,上述分析说明

本文综合考虑专题地图信息量、专题地图信息量时间相关性和邻近时刻地图内容表达一致性综合评价时序地理信息统计数据分级结果是有效的。

表 4. 分级结果统计分析( $\beta=0$ )

	等差 分级	标准差 分级	分位数 分级	头尾分 割分级	自然裂点 分级 I	自然裂点 分级 II	自然裂点 分级 III
$\overline{CG} \uparrow (\beta=0)$	0.966	0.685	0.549	0.763	0.787	0.589	0.798
$G \uparrow (\beta=0)$	0.799	0.718	0.768	0.835	0.846	0.706	0.846

3.2 参数敏感性分析

公式(6)和公式(8)中参数需要设置,前文实验依据数据特征和制图需要设定了相应的经验阈值。公式(6)计算专题地图信息量的时间相关性( $I_G$ ),考虑到新冠疫情死亡人数呈现出明显的随时间变化特点,且周期性变化相对不明显,取 $\alpha_1=0.9, \alpha_2=0.1$ 。针对表 2 中  $I_G$  最大和最小的标准差分级结果和自然裂点分级 I 分级结果,分别设置 $\alpha_1=0.0, 0.1, \dots, 0.9, 1.0$ , 对应设置 $\alpha_2=1.0, 0.9, \dots, 0.1, 0.0$ , 分析  $I_G$  随参数的变化,结果见图 4(a)。由图 4(a)可知,  $I_G$  随着 $\alpha_1$ 取值的增加逐渐减小。但是,基于不同分级方法获取的不同分级结果的  $I_G$  差异随着 $\alpha_1$ 取值的增加而增加,这表明随时间变化特征明显的统计数据,  $\alpha_1$ 取值较大时能更有效的区分不同分级结果专题地图信息量的时间相关性。实际应用中需要依据数据本身的数据特征确定 $\alpha_1$ 和 $\alpha_2$ 取值,若统计数据周期性变化明显,  $\alpha_2$ 可适当取较大值;否则,可取较小值;数据随时间变化的特征可以利用时间序列方法获取。

公式(8)用于计算邻近时刻地图内容一致性的平均值( $\overline{CG}$ ),考虑到新冠疫情死亡人数的制图表达需一定程度表达数据按月的变化特征,设置 $\beta=0.4$ 。针对表 2 中  $\overline{CG}$  最大和最小的等差分级和分位数分级结果,分别设置 $\beta=0.0, 0.1, \dots, 0.9, 1.0$ , 分析  $\overline{CG}$  随参数的变化,结果见图 4(b)。由图 4(b)可知,平均专题地图信息量低的分级结果,在 $\beta$ 取值较小时,其 $\overline{CG}$ 较高,这是因为专题地图信息量低的分级结果邻近时刻变化相对较小。随着 $\beta$ 取值的增大,  $\overline{CG}$ 先增大后减小,这是因为实际制图过程中邻近时刻间数据本身存在差异,若该差异接近 $\beta$ 取值,  $\overline{CG}$ 会增大;反之则减小。实际应用过程中,需依据制图需要和数据本身的特征设置 $\beta$ 取值,若需表达邻近时刻地图内容的变化,则 $\beta$ 可适当取值较大;反之,取较小值。

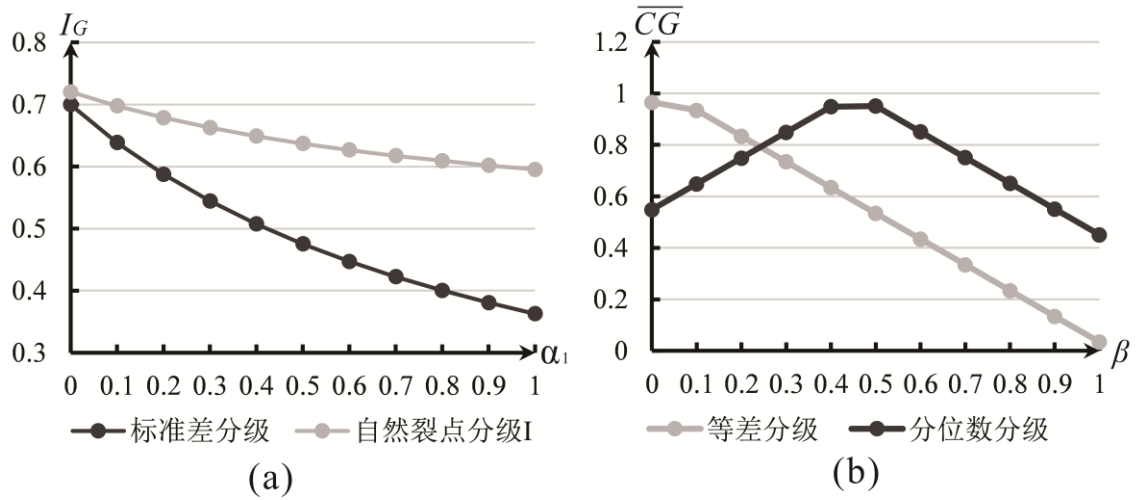


图 4. 参数分析。(a)  $I_G$  随参数  $\alpha_1$  和  $\alpha_2$  的变化；(b)  $\overline{CG}$  随参数  $\beta$  的变化。

### 3.3 其它

实际应用中，时序地理信息统计数据的表达依据数据分布特点和用户需要，有时需要将时间序列划分成多段进行制图，不同时间段的制图需要可能不一样；如新冠疫情初期可能需强调表达数据的增量变化，疫情平稳期可能需削弱这种变化的表达。另外，尽管地理现象随时间变化通常具有规律性(如每天的早晚高峰)；这些现象有时会因客观因素呈现出异常，如早晚高峰易受极端天气等外界事件影响，制图表达时可能需要突出时间序列中存在的异常<sup>[27]</sup>。因此，利用本文方法针对这一类制图需要进行分级结果评价时，需要综合不同时间段、不同时间点的特定需要进行综合评价。

另外，时序地理信息统计数据具有明显的时间尺度效应，如新冠疫情死亡人数按月和按周统计的死亡人数呈现出的规律可能不一样。因此，同一时间段内不同尺度上的最佳分级策略往往并不相同。如基于本文方法针对美国 2020/3/1-2021/9/1 逐月(共 19 月、49 个州)的新冠疫情影响死亡人数的分级结果进行评价，显示分位数分级为最佳分级方法；但是，针对 2020/3/7-2021/2/13 逐周(共 50 周、49 个州)的新冠疫情影响死亡人数分级结果进行评价，显示自然裂点分级 II 为最佳分级方法。

## 4. 结论

为有效评价时序地理信息统计数据的分级质量，本文综合考虑分级结果的统计特征、空间分布特征和时序特征，利用专题地图信息量、专题地图信息量的时间相关性和邻近时刻内容表达一致性评价分级结果，相比已有统计数据分级结果

评价方法,本文方法能较好适应时序统计地理信息数据制图过程中不同地图内容表达一致性的需要。实际应用中数据分布特征多样,未来结合时间序列分析方法,设计适应数据分布特征和用户需求的统计数据分级算法和分级结果评价方法是未来研究的重点。

## 参考文献

- [1] Wu H. Research on Fundamental Theory and Technical Approaches of Automated Map Generalization[J], 2000, 25(5):377-386. (毋河海. 地图信息自动综合基本问题研究[J]. 武汉测绘科技大学学报, 2000, 25(5): 377-386.)
- [2] 何宗宜. 地图数据处理模型的原理与方法[M]. 武汉: 武汉大学出版社, 2004. (He Z. Elements and methods of model for cartographical data processing[M]. Wuhan: Wuhan University Press, 2004.)
- [3] Jenks G F. Generalization in statistical mapping[J]. Annals of the Association of American Geographers, 1963, 53(1):15-26.
- [4] Jiang Bin. Head/Tail breaks: a new classification scheme for data with a heavy-tailed distribution[J]. The Professional Geographer, 2013. 65(3): 482-494.
- [5] Sun M, Wong D W, Kronenfeld B J. A classification method for choropleth maps incorporating data reliability information[J]. The Professional Geographer, 2015, 67(1): 72-83.
- [6] Sun M, Wong D, Kronenfeld B. A heuristic multi-criteria classification approach incorporating data quality information for choropleth mapping[J]. Cartography and geographic information science, 2017, 44(3): 246-258.
- [7] Sergio J. Rey, Philip Stephens, and Jason Laura. An evaluation of sampling and full enumeration strategies for Fisher Jenks classification in big data settings. Transactions in GIS, 21(4):796–810.
- [8] 陆效中. 统计地图的分级表示法[M]. 北京: 解放军出版社, 1989. (Lu X. Classification representation of statistic maps[M]. Beijing: The PLA Press, 1989.)
- [9] 孙亚梅, 王如云. 专题要素分级的新方法及其应用[J]. 测绘学报, 1994, 23(1): 59-66. Sun Y, Wang R. A new grading method and its application in grading of thematic

- elements[J]. *Acta Geodaetica et Cartographica Sinica*, 1994, 23(1): 59-66.
- [10] 郭庆胜, 李留所, 贾玉明, 等. 顾及空间自相关的统计数据分级质量评价[J]. *武汉大学学报 (信息科学版)*, 2006, 31(3): 240-243. Guo Q, Li L, Jia Y, etc. Quality Evaluation of Statistical Data Classification Considering Spatial Autocorrelation[J]. *Geomatics and Information Science of Wuhan University*, 2006, 31(3): 240-243.
- [11] 江南, 白小双, 孙娟娟. 基于多属性决策的统计数据分级评价模型[J]. *测绘学报*, 2007, 36(2): 198-202. Jiang N, Bai X, Sun J. Classification evaluation model of statistic data based on multi-attribute decision making[J]. *Acta Geodaetica et Cartographica Sinica*, 2007, 36(2): 198-202.
- [12] 姚宇婕, 陈毓芬. 引导型专题数据分级处理研究[J]. *测绘工程*, 2012, 21(1): 25-29. Yao Y, Chen Y. Research on guiding thematic data classification [J]. *Engineering of Surveying and Mapping*, 2012, 21(1): 25-29.
- [13] Cromley R G, Mrozinski R D. An Evaluation of Classification Schemes Based on the Statistical Versus the Spatial Structure: Properties of Geographic Distributions in Choropleth Mapping[C]//ACSM ASPRS Annual Convention and Exposition. 1997, 5: 76-85.
- [14] Brewer C A, Pickle L. Evaluation of methods for classifying epidemiological data on choropleth maps in series[J]. *Annals of the Association of American Geographers*, 2002, 92(4): 662-681.
- [15] Eicher C L, Brewer C A. Dasymetric mapping and areal interpolation: Implementation and evaluation[J]. *Cartography and Geographic Information Science*, 2001, 28(2): 125-138.
- [16] Huang Z, Lu Y, Mack E A, et al. Exploring the sensitivity of choropleths under attribute uncertainty[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2019, 26(8): 2576-2590.
- [17] 肖佳, 田沁, 何宗宜. 地理信息数据分级评价的相对指数熵模型[J]. *测绘学报*, 2020, 49(11): 1497-1505. XIAO Jia, TIAN Qin, HE Zongyi. Relative exponential entropy model on classification evaluation of geographic information data[J]. *Acta Geodaetica et Cartographica Sinica*, 2020, 49(11): 1497-1505.

- [18] 李志林, 刘启亮, 高培超. 地图信息论: 从狭义到广义的发展回顾[J]. 测绘学报, 2016, 45(7): 757-767. LI Zhilin, LIU Qiliang, GAO Peichao. Entropy-based Cartographic Communication Models: Evolution from Special to General Cartographic Information Theory[J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(7): 757-767.
- [19] 应申, 窦小影, 徐雅洁,等.新型冠状病毒肺炎疫情可视化进展与分析[J].地球信息科学学报,2021,23(2):211-221. Ying S, Dou X Y, Xu Y J, et al. Visualization of the epidemic situation of COVID-19[J]. Journal of Geo-information Science, 2021,23(2):211-221.
- [20] Chi M T, Lin S S, Chen S Y, et al. Morphable word clouds for time-varying text data visualization[J]. IEEE transactions on visualization and computer graphics, 2015, 21(12): 1415-1426.
- [21] 何宗宜, 宋鹰. 普通地图编制[M]. 武汉: 武汉大学出版, 2015: 61-78. (He Z, Song Y. General Map Compilation [M]. Wuhan: Wuhan University Press, 2015:61-78.)
- [22] 苏为华. 多指标综合评价理论与方法问题研究[D]. 厦门大学, 2000. (Su W. Research on Multiple-Objective-Comprehensive-Evaluation[D]. Xiamen University, 2000.)
- [23] 王劲峰, 廖一兰, 刘鑫. 空间数据分析教程[M]. 北京: 科学出版社(第 2 版), 2019, 30-38. (Wang J, Liao Y, Liu X. Spatial Data Analysis Tutorial (The second edition) [M]. Beijing: Science Press, 2019, 30-38.)
- [24] Williams C K, Rasmussen C E. Gaussian processes for machine learning[M]. Cambridge, MA: MIT press, 2006.
- [25] National Center for Health Statistics. Provisional Death Counts and COVID-19, 2021. Available from: <https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm>
- [26] Rey S J, Stephens P, Laura J. An evaluation of sampling and full enumeration strategies for Fisher Jenks classification in big data settings[J]. Transactions in GIS, 2017, 21(4): 796-810.
- [27] Cheng X, Wang Z, Yang X, et al. Multi-scale detection and interpretation of spatio-temporal anomalies of human activities represented by time-series[J]. Computers, Environment and Urban Systems, 2021, 88: 101627.