

利用 AQL 的逐批海洋大数据质量检验模型

黄冬梅 周雪楠 王振华

上海海洋大学信息学院 上海 201306

(dmhuang@shou.edu.cn)

A based-on-AQL quality inspection model for different lots of big marine data

Huang Dongmei, Zhou Xuenan, and Wang Zhenhua

(College of Information, Shanghai Ocean University, Shanghai 201306, China)

Abstract Data quality guarantees the marine data processing and application. How to design an optimal quality inspection plan fast and control the marine data quality efficiently grows more and more important for the application of big marine data. A quality inspection plan contains three parameters, which are lot size, sample size and acceptance number. Recent studies mainly focus on developing relationships between sample size and acceptance number, while ignoring the influence of lot size which is not suitable for inspecting big marine data. Hence, Hypergeometric Distribution and the concept of Acceptance Quality Limit (AQL, for short) are led up to the optimal quality inspection plan for big marine data to solve the problem. By comparing the results of the proposed sampling plan with those calculated in terms of the traditional standard, the reliability of the proposed model is validated.

Key words big marine data; quality management; quality inspection plan; hypergeometric distribution; acceptance quality limit(AQL)

摘要 海洋数据的质量是数据处理和应用的基础,如何准确高效的评价海洋数据的质量,是制约其精确有效应用的关键问题之一.质量检验方案主要涉及三个参数,即批量、样本量和接收数,而现有的质量检验方案大多集中于样本量与接收数之间的关系推导,忽略了数据批量对于质量检验方案的影响.此类方案不适用于批量大小不固定的海洋大数据的质量检验.针对该问题,通过基于接收质量限(Acceptance Quality Limit, AQL),提出了符合超几何分布的海洋大数据优化质量检验模型,建立了批量和样本量之间的联系,平衡了数据生产方和使用方对于数据精确度的需求.最后,通过与传统质量检验模型比较,验证了其对于海洋大数据质量检验的有效性.

关键词 海洋大数据; 质量控制; 质量检验方案; 超几何分布; 接收质量限

中图法分类号 TP393

空间数据的质量问题一直是国际关注的地理信息科学的问题之一,数据质量的优劣将直接影响数据分析结果的可靠程度和应用目标的实现.作为一类典型的空间数据,海洋数据由于获取手段的多样化,呈现出多源、异构的特点,同时立体化、准实时海洋监测系统的建立,使得海洋数据的量不断增加,使之成

为大数据的典范^[15].

随着时空地理信息科学产业的兴起,对海洋大数据进行科学的质量检验和精度评估已迫在眉睫.从海洋数据生产方的角度出发,若对采集到海洋大数据逐记录、逐属性进行质量检验,判定每条记录是否合格,将花费大量的人力、物力、财力,这显然是不科学的.

基金项目: 国家自然科学基金项目(61272098);上海市自然科学基金(13ZR1455800);国家973项目(2012CB316200)

通信作者: 周雪楠, xuenanzhou@163.com

所以需要海洋大数据用抽样的方式进行质量检验,通过评估抽取样本的质量程度来判断总体批量是否符合数据使用方的精度要求

二十世纪以来,许多学者针对空间地理数据的特点,设计了数据的质量检验模型,包括 S. Balamurali 提出的基于多重参数的抽样检验优化方案^[1], Chien-Wei Wu 提出的一种基于过程能力指数的抽样方案^[2]以及 Duarte 和 Saraiva 在提出的一种抽样检验最优整数规划模型^[3]。但是这些抽样检验模型重点研究了样本量和接收数之间的关系优化,较少涉及批量对整个检验方案影响的研究。显然不适用于来源多样化的海洋数据的质量检验。

针对以往研究的不足,本文从生产者的精度需求出发,基于接收质量限(AQL),利用超几何分布模型,设计了海洋大数据逐批次的优化质量检验模型,建立了批量与样本量之间的联系,保证了质量检验方案在不同批次海洋大数据的质量检验中具有一致的严谨度和辨别能力。

1 相关工作

海洋时空大数据的质量问题是伴随着数据的采集、处理和应用的过程而产生并表现出来的。在 GIS 的相关图书和文献中,与空间数据质量相关的一般围绕着误差(error)、准确度(accuracy)、偏差(bias)、精密度(precision)以及不确定性(uncertainty)几个问题展开^[16-18],但是对于空间数据质量评价方法上,空间数据质量的检验和控制方面研究较少。

二十世纪以来,我国地理信息界的许多学者针对空间地理数据的特点,利用抽样检验的模型,控制和检验空间数据的质量。S. Balamurali 在文章中讨论了基于 AQL 和 LQL 的计量抽样检验下多重参数的抽样检验优化方案^[1]。Chien-Wei Wu 提出了一种基于过程能力指数(Process Capability Indices)的抽样方案,立足于平衡生产方和使用方利益的基础上通过控制接收质量限(AQL)和极限质量限(Limited Quality Level, LQL)的接收概率保证达到良好的抽样检验效果^[2]。Duarte 和 Saraiva 在文献中提出了一种抽样检验最优整数规划模型^[3],在文章中分别构建了基于海洋数据接收质量限(AQL)的最优化抽样检验模型和同时控制两种风险的最优化抽样检验模型,还构建了两种情况的二次抽样检验模型。刘大杰等分析了百分比抽样方案在空间数据抽样中的缺陷并提出了挑选型一次极限质量抽样检验模型^[8]。王振华在其论文中提出 GIS 二级抽样检验模型,并基于此模型制定《国土资源数据库数据质量检查验收规范》,二级抽样检验模型是基于 AQL 建立的连续批调整型抽样模型^[9]。

传统的抽样检验模型大多基于传统工业产品建

立,作为一种空间数据,海洋数据产品与传统工业产品不同,其典型特征包括:(1)海洋数据具有多尺度性;(2)海洋数据具有多类性;(3)海洋数据具有多源性;(4)海洋数据具有多维性^[7]。不同时间获取的相同类型海洋数据和相同时间获得的不同类型海洋数据由于其繁杂多变的数据量和数据质量元素,均被视为不同批次的海洋大数据。海洋数据涵盖区域广、时空分布不均匀,因此如何根据海洋数据的不同批量范围,快速给出其质量检验的优化方案是海洋大数据质量检验的关键问题。对于此类复杂、多源、多类型的海洋数据,传统的质量检验模型由于其固定的抽样方案,不能灵活地满足不同批次海洋大数据对于检验精度不同的要求,具有局限性,因此需要针对海洋大数据的这些特征,建立基于用户质量需求的逐批海洋大数据质量检验模型。

2 海洋数据抽样检验方案与相关参数

本文以某一时间间隔内收集到的固定海域数据为一个批次,即检验产品批。同时以该批数据产品中每一条包含不同质量元素的海洋数据记录作为一个单位产品,解决不同批次海洋数据中如何确定抽样的样本量和判定系数从而保证抽样检验质量的问题。

2.1 海洋数据的抽样方案

对海洋数据进行质量检验,首先考虑从总体中随机选取一部分数据产品作为样本,对其进行质量检查,根据样本的质量程度来判断其总体数据产品批量的质量水平,其检验结果可能被接收,也可能被拒绝。

定义1 一次海洋数据抽样方案记为三元组 $S(N, n, c)$,其中 N 为一个常量,对应一批海洋数据的批量; n, c 均是变量, n 为对应批量 N 中抽样所取的样本量, c 为对应该批数据样本不合格品的接收数,其中 $c < n < N$ 。

用 d 表示样本 n 中发现的不合格品的个数,若 $d \leq c$ 时,即该批海洋数据中具有质量问题的数据个数小于或等于质量评估的判定参数,则认为该批数据合格;若 $d > c$ 时,即该批海洋数据中具有质量问题的数据个数大于质量评估的判定参数,该批海洋数据视为不合格数据。

对于包含 D 个不合格品的批量 N ,在 n 个样本的检验中,恰好发现 d 个不合格品数的概率服从超几何分布^[11]:

$$h(d, n, D, N) = \frac{\binom{D}{d} \binom{N-D}{n-d}}{\binom{N}{n}} \quad (1)$$

超几何分布是抽样理论的基础,它适用于有限总体的无放回抽样,适用于逐批的海洋大数据质量检验。用 $L(p)$ 表示抽样方案的接收概率,则抽样方案 $S(N, n, c)$ 的接收概率 $L(p)$ 即为样本中不合格品数 d 小于 c 时所有可能取值的概率之和^[11]。

$$L(p) = \sum_{d=0}^c h(d, n, D, N) \quad (2)$$

2.2 生产者风险与接收质量限

对一批产品进行抽样检验的结果,存在两种错误的判断,一种是好的数据批被拒收了,另一种是坏的数据批被接收了,达到可接收质量水平的数据批被部分不接收,这是数据生产方不希望发生的,称为生产方风险或“ α ”风险^[13],而抽样方案给数据生产方和数据使用方提供的保护,就是指减少风险的程度,并可通过计算进行调整。

接收质量限(Acceptance Quality Limit, AQL)是当一个连续系列批被提交验收检验时,可允许的最差过程平均质量水平,它是可以接收和不可以接收的过程平均界限值^[11]。在已有抽样检验模型例如国家标准 GB/T 2828.1(2012)中, AQL 被作为一个检索工具,用来查找质量水平要求不劣于 AQL 的检验批的抽样检验模型^[12]。

从生产者的角度出发,当检验批的质量水平等于或优于 AQL 时,其判为不合格的概率应不大于 α ,即生产方风险小于等于 α ,也就是抽样检验方案的接收概率不小于 $1 - \alpha$ 。

本文以检验批的不合格品率 p_0 表示检验批的 AQL。在抽样检验实施前,生产者和使用者需协商给出该检验过程的 AQL,即检验批可允许的不合格品率 p_0 。如果检验批的不合格品率低于或等于这个值,则该检验批达到质量要求,为合格批。

2.3 OC 曲线

OC 曲线为操作特征曲线,表示的是在假定的质量水平下,批次产品被接收的概率^[13]。横坐标为表示批的质量,纵坐标表示批被接收的概率。OC 曲线描述的是一个抽样方案能力的强弱,一个分辨率强的 OC 曲线是在高质量水平处凸出显示,而在低质量水平处凹进显示^[11],如 Fig.1 所示。

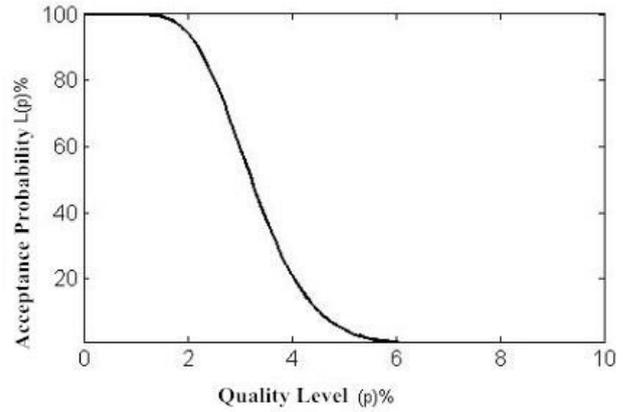


Fig.1 Sample of OC Curve

图 1 抽样检验模型的 OC 曲线

3 设计海洋大数据优化质量检验方案

对一批海洋数据进行质量检验,本文通过海洋大数据的优化质量检验模型来解决如何设定一批数据产品的抽样样本和合格接收数,通过最小化生产方风险,优化海洋数据的抽样方案。

基于以上思想,对于海洋数据质量检验优化模型的部分参数定义如下:

常量:

N - 海洋数据批量;

D - 同一批海洋数据中的不合格品数;

d - 被检验海洋数据样本中检验后不合格品数;

连续变量:

p_0 - 接收质量限(AQL)下的不合格品率;

p - 实际不合格品率;

ε_α - 某批海洋数据在一定接收质量限下的接收概率残差;

α - 在接收质量限下的拒绝概率;

离散变量:

n - 进行质量评估所需要的海洋数据的样本量;

c - 海洋数据样本中的评估判定数,即接收数。

3.1 逐批次海洋数据质量检验模型

针对每一批待检验海洋数据,都存在一个接收质量限 AQL。在对一批海洋数据质量检验前,先根据所需数据质量要求给出该次检验过程的 AQL 值,即检验批可允许的不合格品率 p 。基于超几何分布模型,质量检验方案的接收概率记为 $L(p)$:

$$L(p) = \sum_{d=0}^c \frac{\binom{N-D}{n-d} \binom{D}{d}}{\binom{N}{n}} \quad (3)$$

$$D = N \times p \quad (4)$$

因此，海洋数据质量检验接收概率残差 ε_α ，由下式给出：

$$\varepsilon_\alpha = L(p_\alpha) - (1 - \alpha) \quad (5)$$

其中， α 为生产方风险，当海洋数据的质量达到质量接收限 AQL，其接收概率 $L(p_\alpha)$ 应在 $1 - \alpha$ 附近。所以通过最小化生产方风险，优化海洋数据的抽样方案，通过控制接收概率不小于 $1 - \alpha$ ，且 OC 曲线过点 $(p_0, 1 - \alpha)$ ，本文制定了质量检验方案的非线性规划模型如下：

$$\begin{aligned} & \min_n \varepsilon \\ \text{s.t. } & \varepsilon = \varepsilon_\alpha = \sum_{d=0}^c \frac{\binom{N-D}{n-d} \binom{D}{d}}{\binom{N}{n}} - (1 - \alpha) \quad (6) \\ & (0 < c < n - 1, \varepsilon > 0) \end{aligned}$$

实践已经证明，将要选择的抽样方案，对保证提交批次的平均质量水平优于 AQL 的供方是有利的^[13]，也就是说，在此过程中，质量检验接收概率残差 ε 大于 0。

抽样检验模型中接收数 c 和样本量 n 均为整数，使用迭代的方法来确定，计算海洋大数据的优化质量检验方案集合的迭代步骤如下：

(1) 令 $c=0$;

(2) 令 $n=0$ ，基于公式(3)(4)计算空间数据检验模型的接收概率 $L(p)$ 与规定的最小接收概率 $1 - \alpha$ ，初始状态的残差 ε_0 ;

(3) 接着，使 $n=n+1$ ，基于公式(3)(4)计算此时空间数据检验模型的接收概率 $L(p)$ 与规定的最小接收概率 $1 - \alpha$ 的残差平方和 ε_1 ;

(4) 比较 ε_0 和 ε_1 。若空间数据检验模型的接收概率 $L(p)$ 与规定的最小接收概率 $1 - \alpha$ 的残差平方和 ε_0 小于 ε_1 ，停止运算，即 $n=0$ 为该空间数据质量检验批的最佳接收数；

(5) 若初始状态的残差 ε_0 不小于残差 ε_1 ，继续 $n=n+1$ ，计算样本量 n_i 对应的残差平方和，直至残差平方和 ε_{i-1} 小于残差平方和 ε_i ，停止计算， n_{i-1} 为 $c=0$

最佳样本量，此时跳出，令 $c=c+1$ ，重复步骤(2)；

(6) 当 $c=N \times p_0$ 时停止，此时得到基于海洋大数据的优化质量检验模型的方案集合。

通过对得出的质量检验方案集合的比较发现，随着迭代过程中接收数 c 每增加 1，样本量 n 的增加幅度趋于稳定。当样本量 n 的增加幅度不再变化，也就是抽样方案每增加 1 个不合格品数，相应增加固定个数的样本量，也就是选取更多的样本量已经没有意义，所以认为已经得到最优质量检验方案。

3.2 海洋大数据优化质量检验方案选择算法

根据建立的质量检验模型和迭代的方法，本文设计了海洋大数据优化质量检验方案的选择算法，通过接收质量水平 p_0 ，生产方风险 α ，批量 N ，这三个参数来确定一个最优化的检验方案。算法描述如下：

算法 1: 海洋大数据优化质量检验方案选择算法(p_0, α, N)

输入: 接收质量水平 p_0 ，生产方风险 α ，批量 N ;

输出: 海洋数据质量检验方案 $S(N, n, c)$ 。

① 初始化：令批量海洋数据样本接收数 $c=0$ ，对应的样本量 $n=1$;

② FOR ($c=0$; $c \leq p_0 * N$; $c++$) {

③ FOR ($n=1$; $n < N-1$; $n++$) {

④ 将 (n, c) 代入质量检验模型，利用公式 (6) 计算当前状态的残差平方和 ε_n ;

⑤ IF ($\varepsilon_n < \varepsilon_{n+1} \parallel \varepsilon_n > 0$) BREAK;

⑥ 输出 $S(N, n, c)$;

⑦ IF ($n_{c+1} - n_c \leq n_c - n_{c-1}$) BREAK; }

⑧ 输出优化方案 $S(N, n, c)$ 。

算法分析：

在该算法中，(1)求该批海洋数据在给定批量和接收质量限下的样本量 n 的时间复杂度为 $O(N)$ ；(2)求已知当前样本量 n 的接收数 c 的时间复杂度为 $O(1)$ 。(3)在样本量 n 和接收数 c 都未知的情况下时间复杂度为 $O(N^2)$ ，因此该算法的时间复杂度为 $O(N)$ 。

4 实验结果与分析

本文实验选用数据为环渤海海域秦皇岛监测站获取的某时刻海水增养殖区监测的水质数据，其观测要素包括观测浮标的经纬度，水深，水色，盐度，酸碱度，需氧量，悬浮物，和各种化学物质质量等数据。

由于该批海洋数据具有检测要素种类繁多，而且获取周期短，频率高，数据来源广泛等特点，是一类典型的质量不稳定的海洋大数据。该研究区域内某时刻包括监测点位数据 N 为 5078 条，其每一点位包括三大类数据，分别是空间位置数据，海洋要素数据和

养殖信息数据, Table 1 截取了该批数据的部分监测要素和相关数据记录。

Table 1 Sample of Marine Data Sheets
表 1 实验数据表相关海洋数据示例

St.	Depth m	Color	Salt Mg/L	Ph	Nitra Mg/L	Nitri Mg/L	AN Mg/L
64	21	4.5	31.2	8.22	0.00224	0.00503	0.0608
63	23	4	31.405	8.23	0.00159	0.0172	0.0359
62	29	5.5	31.359	8.34	0.0014	0.0293	0.0314
116	19	5	30.673	8.35	0.00205	0.00158	0.0484
117	16	3	30.782	8.11	0.00495	0.0182	0.0466
87	14	3	30.725	8.19	0.00271	0.0122	0.0295
105	13	2	29.979	8.28	0.0104	0.0951	0.043
94	13	1	29.843	8.14	0.0175	0.167	0.053
84	17	5	31.637	8.21	0.0182	0.0833	0.0218
74	22	5	30.299	8.21	0.0228	0.108	0.0418
83	17	9	31.586	8.21	0.00812	0.087	0.034

对该批海洋数据选用三种不同的接收质量限 0.02、0.03、0.05 进行质量检验, 通过以上算法参照 GB/T 2828.1(2012)中规定的 90%、95%、98% 三种严格程度的抽样方案详细情况如 Table 2 所示:

Table 2 Comparison of Sampling Plans under different AQL
表 2 三种质量限下的抽样方案

p_0	90%		95%		98%	
	n	c	n	c	n	c
0.02	356	10	393	12	619	19
0.03	211	9	316	14	332	16
0.05	159	10	189	14	216	20

Fig.2 所示为 Table 2 基于本文的抽样方案分别在 0.02、0.03、0.05 三种接收质量限下不同严格程度情况下 oc 曲线的对比。从图上可以看出, 接收质量限为 0.02 时 oc 曲线靠左, 说明此时的抽样方案最为严格, 接下来依次是接收质量限为 0.03 和 0.05 的抽样方案。

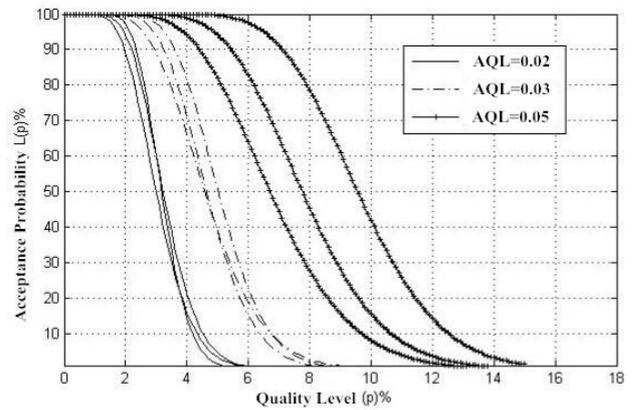


Fig. 2 Comparison of OC curve under different AQL
图 2 三种质量限下的 oc 曲线比较

将本文提出的检验模型与查找国家标准 GB/T 2828.1 所得抽样模型进行比较, 选取了质量限为 0.02 和 0.05 的两种情况进行对比分析。Table 3 列出两类抽样检验模型在 0.02 和 0.05 两种质量限的情况下 90%、95%、98% 三种严格程度的抽样方案详细情况, n_1 , c_1 分别表示海洋数据质量检验模型的样本量、接收数; n_2 , c_2 分别表示查表 GB/T 2828.1 所得抽样检验模型的样本量、接收数。

由 Table 3 可以看出, 在 AQL 相同的情况下, 对于不同的不合格品率 GB/T 2828.1 所得的抽样检验模型的样本量是相同的; 然而基于本文算法的抽样方案在不同的不合格品率的情况下得到的最优样本量是不相同的, 这样保证了在不同质量等级要求下抽样样本量的灵活性。尤其是对于具有多尺度性的海洋数据来说, 在不同比例尺批量数据要求的数据质量等级不同的情况下, 抽样样本量的灵活度直接影响到抽样的效率和准确度。

Table 3 Comparison of two Sampling Plans
表 3 本文抽样检验方案和 GB/T 2828.1 的比较

	p_0 AQL	0.02			0.05		
		90%	95%	98%	90%	95%	98%
Proposed Sampling Plan	n_1	356	393	619	159	189	216
	c_1	10	12	19	10	14	20
GB/T 2828.1	n_2	32	200	315	32	200	315
	c_2	3	9	11	6	17	18

Fig.3、Fig.4 所示为针对 Table 3, 两类抽样检验模型在不合格品率分别在 0.02 和 0.05 的情况下的 OC 曲线比较:

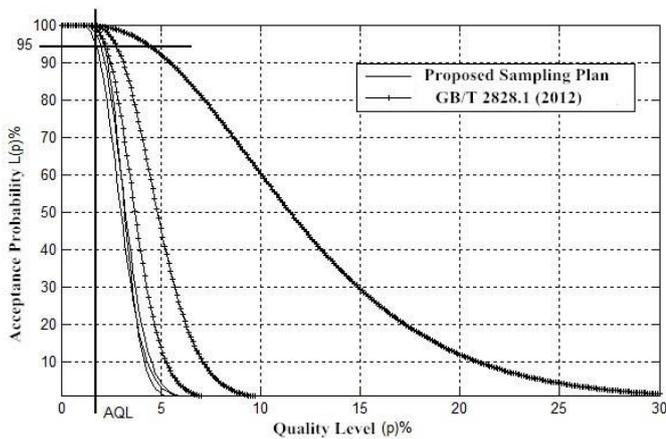


Fig.2 Comparison of OC Curve between Two Sampling Plans($p_0=0.02$)

图3 本文抽样方案和GB/T 2828.1的OC曲线比较($p_0=0.02$)

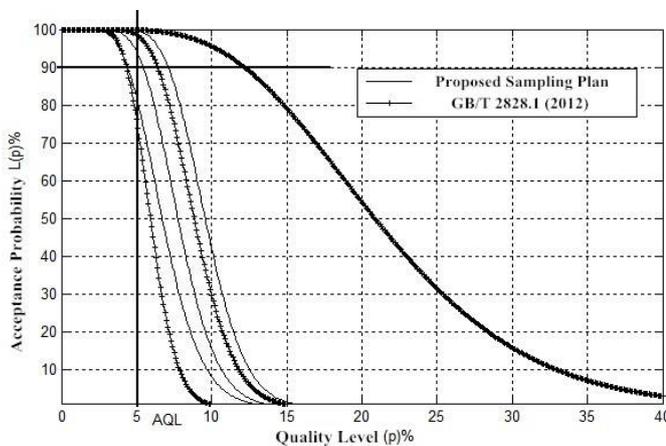


Fig.4 Comparison of OC Curve between Two Sampling Plans($p_0=0.05$)

图4 本文抽样方案和GB/T 2828.1的OC曲线比较($p_0=0.05$)

由 Fig.3、Fig.4 可以看出：

(1) 查表 GB/T2828.1 所得抽样检验模型具有不一致的严格度。如图 4($p_0=0.05$)中带“+”曲线所示，例如，当检验批的不合格品率 p 接近于或等于质量指标 p_0 时，其抽样检验模型(32,6)的接收概率小于 70%，而抽样检验模型(159, 10)的接收概率为 90%；说明在相同的严格度和批量的情况下，本文提出的抽样方案具有更高的接收概率，数字产品更容易被使用方接收，具有良好的分辨率。

(2) 从 Fig.3 和 Fig.4 均可以看出，图中带“+”的曲线要比图中光滑的曲线分散，此现象表明本文提出的方案协调了生产者和使用者的矛盾问题，使得抽样检验模型的严谨度不会随着接收质量限严谨度的变化而变化，保证了抽样方案的科学性和稳定性。

(3) 从 Fig.3 和 Fig.4 均可以看出，图中光滑的曲线比带“+”的曲线更靠近点($p_0, 1-\alpha$)，即(0.05, 90%)，

这说明本文提出的方案比 GB/T 2828.1 更能保证抽样结果接近生产者和使用者的期望，不会导致检验过严或者过松的现象发生，能够更好的协调数据生产方和使用方之间对于数据精度需求的矛盾。

5 总结

本文提出了一种基于接收质量限的符合超几何分布特征的海军大数据质量优化检验模型，通过控制抽样方案的接收概率与接收质量限的期望来选择达标的质量检验方案，建立了批量与样本量之间的关系，并通过实验验证了本文提出的方法在协调了数据生产方和使用方之间对于数据精度需求的矛盾的同时较 GB/T 2828.1 质量检验模型更具有稳定性和分辨性。

参考文献

- [1] Balamurali S, Jun C-H. Multiple dependent state sampling plans for lot acceptance based on measurement data[J]. European Journal of Operational Research, 2007, 180(3): 1221-1230.
- [2] Wu C-W, Pearn W L. A variables sampling plan based on Cpmk for product acceptance determination[J]. European Journal of Operational Research, 2008, 184(2): 549-560.
- [3] Belmiro P.M. Duarte, Pedro M. Saraiva, (2008) "An optimization-based approach for designing attribute acceptance sampling plans"[J], International Journal of Quality & Reliability Management, Vol. 25 Iss: 8, pp.824 – 841.
- [4] Schilling, E.G., 1982. Acceptance Sampling and Quality Control. New York: Dekker.
- [5] Eleftherion, M., Farmakis, N., 2009. Continuous sampling plan under quadratically varying acceptance cost. The XIII international conference "Applied Stochastic Models and Data Analysis", Vilnius, Lithuania, 289-293.
- [6] Kandasamy S., Venugopal H., (2013) "Selection of single sampling attribute plan for given AQL and LQL involving minimum risks using weighted Poisson distribution"[J], International Journal of Quality & Reliability Management, Vol. 30 Iss: 1, pp.47 – 58.
- [7] Dodge, H.F., 1943. A sampling inspection plan for continuous production. The Annals of Mathematical Statistics[J], 14(3):264-279.
- [8] Liu Dajie, Liu Chun. Study on sampling inspections schemes to digital products in GIS [J] . Journal of Wuhan Technical University of Surveying and Mapping, 2000, 4(25): 348.
(刘大杰, 刘春. GIS 数字产品质量抽样检验方案探讨[J]. 武汉测绘科技大学学报, 2000, 4(25) : 348.)
- [9] Tong, X.H., Wang, Z.H., Huan, X, Liang, D, Jiang, Z.Q., Li, J.C., Li, J., 2010. Designing a two-rank acceptance sampling plan for quality inspection of geospatial data products[J].Computers and Geosciences;

2010,(01) ; 37:1570-1583.

- [10] Zhang Yuzhu. Theoretical and Applied adjustable sampling system[M]. Beijing: National Defence Industry Press, 2005.
(张玉柱. 调整型抽样检验系统理论与应用. 北京: 国防工业出版社, 2005.)
- [11] YU Shanqi. Sampling inspection and quality control [M].Beijing: Peking University Press, 1991.
(于善奇. 抽样检验与质量控制. 北京:北京大学出版社, 1991.)
- [12] Yu Zhenfan. The latest statistical sampling inspection and process control of practical tutorial [M] .Beijing: Standards Press of China,2004.
(于振凡. 最新统计抽样检验与过程控制.中国标准出版社, 2004.)
- [13] Xiao Hui, Zhang Yuzhu, Ma Yilin, et al. GB/ T 2828. 1 2012 Sampling procedures for inspection by attributes: part 1 sampling schemes indexed by acceptance quality limit (AQL) for lot by lot inspection. general administration of quality supervision [M] .Beijing: Standards Press of China, 2012.
(肖惠, 张玉柱, 马毅林, 等. GB/ T 2828. 1 2012 计数抽样检验程序第 1 部分:按接收质量限(AQL) 检索的逐批检验抽样计划 理解与实践[M]. 北京: 中国标准出版社, 2012.)
- [14] Goodchild, M.F., Jeansoulin, R., 1998. Data Quality in Geographic Information[C]: From Error to Uncertainty, Hermes, Paris, 192pp.
- [15] Huang Dongmei, Du Yanling, He Qi, Migration Algorithm for Big Marine Data in Hybrid Cloud Storage[J], Journal of Computer Research and Development, 2014, (01): 199-205.
(黄冬梅, 杜艳玲, 贺琪. 混合云存储中海洋大数据迁移算法的研究[J]. 计算机研究与发展, 2014, (01): 199-205.)
- [16] Wang Zhenhua. Principle, methods and application of sampling inspection for quality control of geospatial data[D]. Tongji University, 2011.
(王振华, 空间数据质量抽样与控制的理论、方法和应用.2011, 同济大学博士论文).
- [17] Liu Dajie, Liu Chun. Study on sampling inspection schemes to digital products in GIS [J] . Journal of Wuhan Technical University of Surveying and Mapping, 2000, 4(25): 348.
(刘大杰, 刘春. GIS 数字产品质量抽样检验方案探讨[J] . 武汉测绘科技大学学报, 2000, 4(25) : 348.)

黄冬梅, 女, 1964 年生, 博士生导师, 教授, 主要研究方向为 WebGIS 智能信息处理和辅助决策系统。

周雪楠, 女, 1991 年生, 在读研究生, 主要研究方向为空间数据质量控制研究。

王振华, 女, 1982 年生, 博士, 讲师, 主要研究方向为空间数据的质量控制理论研究。