

GIS 支撑下的长时段区域人口变动规律分析

——以 1776 至 1953 年陕甘地区人口为例

复旦大学历史地理研究中心教授 路伟东，上海，200433

该文首次发表于《历史地理》第 30 辑（上海：上海人民出版社，2014 年 12 月，第 314—324 页），本文在此基础上有较大修改。

摘要：历史人口是典型的具有时间序列特征的空间数据，本文根据宣统人口调查原始档案“地理调查表”甘肃分析户口数据，使用 GIS 人口重心模型，对 1776 年至 1953 年间西北区域人口空间变动进行了系统分析，认为：人口重心模型可以较好的模拟区域历史人口在这一个较长时段内的变动，曹树基先生对陕甘地区 1820 年和 1910 年人口数的修正存在问题。GIS 提供的空间分析的技术、理论与方法，对于满足人口史研究更深层次的需求来讲，具有重要的借鉴意义。

关键词：GIS、清代、陕甘地区、人口变动、人口重心、地理调查表

2001 年，复旦大学葛剑雄教授与哈佛大学包弼德教授（Peter K. Bol）共同推动的重大国际合作项目中国历史地理信息系统（CHGIS, China Historical GIS）正式启动。¹ 该项目基于中国传承有序的方志、总志等海量历史文献，以 GIS 技术为平台，核心目标是构建千年尺度具有完整时间序列的空间基础数据。² 最终完成的 CHGIS 可以为回溯的人文社科研究提供精确的时空定位，具有极其重要的学术价值。更重要的是，以 CHGIS 为起点，历史地理信息系统（HGIS, Historical

¹ 作为 CHGIS 项目的前期筹备，2000 年秋，我陪同葛先生一同前往美国印第安纳波利斯参加“电子文化地图集行动计划”（ECAI）的学术研讨会。2001 年 1 月 8 日 CHGIS 项目正式启动。该项目详细信息请查看复旦大学中国历史地理研究所官方网站禹贡网（<http://yugong.fudan.edu.cn>），本人有幸参与了其中一点工作。2009 年，以 CHGIS 为基础开发的中国历史地理信息系统展示平台，参加“辉煌 60 年——中华人民共和国成立 60 周年成就展”，成为教育部展厅仅有的整个教育部展厅仅有的三件实物展览之一，本人因此有幸分享第八届复旦校长奖。

² 路伟东：《CHGIS 数据模型与千年尺度完整时间序列空间基础数据——以 1912-1949 县级治所点数据为例》，《历史地理》第 33 辑，上海：上海人民出版社，2016 年年 7 月，第 269-279 页。

GIS)成为近三十年来,历史地理学界最重要的学术增长点之一。葛先生对于学科发展的推动泽被学林、功莫大焉!

HGIS 的核心是数据,但 HGIS 的灵魂则空间分析。对于研究者来讲,把时间和精力投入到基础数据的开发上当然极具工作价值,但如何能够把 GIS 的技术、理论与方法应用到传统人文社科研究中来,发现或者解决那些通过传统手段无法发现或者解决的问题,做出真正具有明确学术指向和学术关切的交叉研究个案更具吸引力,更能体现 HGIS 的魅力,也更有利于这种交叉研究的发展。本节在现有研究史料及研究成果的基础上,把 GIS 引入到中国人口史的研究中来,结合新发现的宣统人口调查地理调查表分村户口数据,对清以来近两百年间的区域人口发展过程进行系统研究。希望能在检验中国人口史现有成果的同时,给出研究长时段区域人口史的不同方法,为 GIS 在中国人口史研究领域的应用提供有益的尝试和探索。

一、清末民国初人口史研究的困境与历史 GIS 方法的引入

葛剑雄教授主编的六卷本《中国移民史》和《中国人口史》是中国人口史这一学科的重要奠基石,其重要性不言而喻。其中,曹树基著《中国人口史》第五卷《清时期》与侯杨方著《中国人口史》第六卷《1910-1953 年》对清末宣统人口调查的历史评价、数据可信度、修正方法以及修正结果都存在明显不同。这直接导致了同一套书中对于清以来整个中国人口发展进程的理解与表述出现了分歧,前后无法衔接。

作为一部由多位顶尖学界专家共同完成的多卷本中国人口史专著,不同作者之间因为学术见解不同而产生分歧与争议是很正常的。在第一卷导论开篇,葛剑雄先生就对此作了声明,称:“尽管我是本书的主编,但各卷都是作者独立撰写的,文责自负。导论部分是我撰写,也应由我负责。当然,我是在本书各卷的基础上撰写的,获益于各位作者的地方很多,但也有与他们的观点不尽相同之处,所以并不代表其他作者。”¹但是,历史人口数据是典型的具有时间序列特征的空间数据,人口的发展如同奔流的河水,前后是流动的连续的。因此,各卷册之间

¹ 葛剑雄主编、葛剑雄著:《中国人口史》第一卷《导论、先秦至南北朝时期》,上海:复旦大学出版社 2002 年,第 1 页。

在人口发展的重大问题上，应该尽可能作到前后衔接。

这一看似简单的问题，在实际撰写过程中面临巨大挑战。对于中国人口史这样一个宏大的题目来讲，内容所涉及的时空范围都极其广泛。限于时间、精力和研究能力，一个人很难完成，只能由多位作者断代分卷撰写。但这一工作模式也面临极大考验。首先，从时间尺度来讲，历史人口的发展历程与朝代的更替并不完全同步，传统断代人口史的划分方法往往人为割裂了历史人口发展的这种内在有机连续；其次，从空间尺度来讲，历史人口的空间分布及变化具有其独特的规律，传统历史学文字描述式和简单数学统计式的研究方法并不能完整表达其空间特性。

基于此种原因，《中国人口史》第五卷与第六卷间关于晚清宣统人口的不同见解更应该引起重视。但是，如何解决这一问题却困难重重。这是因为，第五卷作者曹树基与第六卷作者侯杨方都是学界最顶尖的专家，他们使用的史料、研究方法、论证逻辑、论证过程以及最终的结论，都已经处在很高的起点之上。如果使用相同的史料，遵循相同的工作模式，后辈学者很难有所突破和超越，也很难对他们之间的分歧进行有效地检验和判断。基于此种困境，笔者认为，在面对这一问题时，需要在一个相对完整的历史人口发展时段内，使用新的方法来重新进行研究，极有必要。

与自然科学研究者需要样本独立的基本研究前提不同，人文社科研究者往往是自己研究对象的一部分。因为只有这样，才可以更了解自己的研究对象。历史人口这类具有时间序列的空间数据通常具有非独立性，不符合经典统计学样本独立的基本假设。因此，专门的空间分析理论和技术在过去的二、三十年间迅速发展，日益成为社会科学研究中的重要研究方法和手段。¹GIS 是由遥感、全球定位系统、地理信息系统、数字传输网络等一系列现代信息技术高度集成的。作为以地理信息系统为核心的空间信息技术体系的总称，GIS 是操纵和处理空间数据的有力工具。任何人文、自然要素的发生、发生与演化，都处于特定的时间和空间中。而今天的 GIS 早已不仅局限于系统与技術，更多的强调是一种处理和分析空间数据的理论与方法。正因为如此，自 1990 年代初 GIS 从实验室走出来，进入普通研究者，尤其是人文社科研究者的视野后，其技术、理论与方法就得到了广

¹ Goodchild M F, Anselin L, Richardp A. *Toward Spatially Intergrated Social Science* [J]. *International Regional Science Review*, 2000, 23(2):139-159.

泛地应用。就国内研究现状而言，以 CHGIS¹为代表的 HGIS 研究工作，是近 30 年来历史地理学界最主要的学术增长点之一。

HGIS 的核心和 GIS 一样，都是数据，从人口史的角度讲，就是史料。人口史的研究对象发生在过去，研究数据来源于历史文献之中。在中国人口史研究的大部分时段内，尤其是民国以前，严重缺少建立在现代人口普查基础上的高质量人口数据，同时也缺少可以替代的相关数据或资料。因此，哪怕是对中国人口史最基本的认知和最粗略的分析，研究者往往都要，甚至是只能，依赖或借助于历史学文献学的方法和手段，通过对大量历史文献的搜集、整理、研读与分析，才能获取。

这一研究方法，人口史的学者非常熟悉，其过程大概包括从史书、史料，到数据、图表，最后到论文、专著等几个步骤。在这一典范式的工作过程中，研究者需要将大量的时间和精力聚焦在原始史料的收集、整理、分析、归纳与总结等方面上，人口数量的历史学考证和人口数据本身的统计学分析是研究者关注的重点，而获取数据的过程往往就是主要的，甚至是全部的研究过程。常常，某一特定时间截面的历史人口规模考证出来了，整个研究也就结束了。至于数据背后所隐含的那些地理空间上真正值得认真审视的现象与规律，反而常常被有意无意地忽略了。

通过使用 GIS 空间分析的理论与方法，我们可以对现有的、经过细致研究获取的大量研究数据，进行深入地挖掘和分析，探究数据背后的规律与特征。GIS 这一“文献爬梳 → 数据提取 → 空间模型建立 → 分析”，最后生成新结论的工作流程，是被既有研究证明的，行之有效的方法。²对于中国人口史研究中遇到的问题及面重的分歧，可以从 GIS 的视角，使用这些成功的研究范式，进行一定的尝试。

二、不同时间切面的人口数据与可视化

入清以来，西北地区的人口发展态势与全国人口相似，都经历了一段比较长的持续稳定发展时期。及至道咸年间，西北人口已经臻于极盛。虽然全国人口在

¹ 中国历史地理信息系统 (CHGIS)，复旦大学历史地理研究中心，2003 年 6 月。

² 潘威、孙涛、满志敏：《GIS 进入历史地理学研究 10 年回顾》，《中国历史地理论丛》2012 年第 1 期。

1850年代初（1851年太平天国运动开始，此后的十余年间，由于天灾人祸的影响，全国损失人口总数接近1个亿。¹）就已达到峰值，但就西北地区而言，人口峰值出现的时间稍晚一些，大约在1860年代初期。

1862年初，西北回民战争爆发，此后数年间，整个西北地区，尤其是人口稠密、富甲天下的关中、河西、宁夏平原等处，沦为双方厮杀的战场，大批汉人与回民因此而丧生。光绪初，还没有从战争废墟中完全恢复过来的陕西省，又遭到了光绪大旱灾的沉重打击。²在战争与灾荒的双重打击下，仅短短十余年间，³西北区域人口损失就在两千万左右。

同治中期以后，随着西北战事的逐渐西迁与结束，区域外人口开始陆续迁入，慢慢填补那些战乱灾荒造成的人口稀疏区或人口空白区。⁴抗战以来，又有大批关东人口西迁入关，躲避战乱。总之，这一人口入迁的趋势，一直持续至中华人民共和国成立，仍未停止。⁵在外来人口入迁导致的人口机械增长和区域人口自然增长的共同推动下，从光绪初至中华人民共和国成立的七十余年间，虽然西北地区仍然不时有战乱、灾荒发生，但区域人口一直处于缓慢地恢复和增长中。

对这样一个长时段内西北地区的人口变动进行研究，有4个比较重要的人口数据支点，即：乾隆四十一年（1776年）人口、嘉庆二十五年（1820年）人口、宣统二年（1910年）人口、1953年人口。1776年是清代在理论上废除人丁编审，将户口与赋税脱钩，并以全体人口为登记对象后，官方登记数据开始较为准确的起点。此时，中国的人口统计已经完成了从纳税的“人丁”向全体“人口”的转变。因此，至少在理论上，这一年的人口数较之前的人口数有了本质的区别。1820年人口数来源于《嘉庆重修大清一统志》，虽然像乾隆四十一年这样的全国规模的人口清查，直到1851年再也没有举行过，⁶但在严格、完善的登记制度的保障下，这一时期的户口统计数据基本还是可靠的。⁷曹树基认为，这

¹ 葛剑雄主编、曹树基著：《中国人口史》第5卷《清时期》，上海：复旦大学出版社2001年，第867页。

² 这次旱灾，因光绪丁丑(1877年)、戊寅(1878年)两年灾情最重，史称“丁戊奇荒”，又因晋、豫两省被灾最烈，亦称“晋豫奇荒”。在所有被灾省份中，陕西旱灾开始晚而结束早，灾情亦不如晋、豫两省严重。

³ 从同治元年（1862年）回民战争爆发到光绪六年旱灾完全结束，前后共约17年。

⁴ 钞晓鸿：《晚清时期陕西移民入迁与土客融合》，《中国社会经济史研究》1998年第1期。

⁵ 张颖：《抗战时期人口内迁对陕西民众社会生活的影响》，《西安社会科学（哲学社会科学版）》2008年第4期。

⁶ 姜涛：《中国近代人口史》，杭州：浙江人民出版社1997年，第48-50页。

⁷ 葛剑雄：《中国人口发展史》，福州：福建人民出版社1991年，第63页。

一数据被广泛引用，不只是因为其相对较为可靠，“而且因为其有一套以府为单位的完整数据。”¹1910年人口来源于宣统人口调查，宣统年间的人口调查以宪政为目地，其谱查办法及实施过程比较接近于现代人口普查，人口数据质量较高。1953年人口普查是中国有史以来第一次现代人口普查，数据非常可靠。是中国人口史研究的重要基石。

曹树基先生经过系统分析和研究，按照嘉庆二十五年（1820年）的行政区划，给出了以上6个时间切面上8个分府人口数据：分别是1776年考证人口数、1820年《一统志》人口数、1820年《一统志》人口修正数、1851年考证人口数、1880年考证人口数、1910年新政人口数、1910年新政人口修正数和1953年普查人口数。各时间切面人口数汇总见表1：

表1 1776-1953年西北陕甘两省人口数

人口单位：万

年份	陕西人口数	甘肃人口数	合计	环比增长
1776年考证人口数	796.5	1,579.9	2,376.4	0
1820年《一统志》人口数	1,197.4	1,713.9	2,911.3	22.5%
1820年《一统志》人口修正值	1,213.4	1,760.5	2,973.9	2.2%
1851年考证人口数	1,326.9	1,899.0	3,225.9	8.5%
1880年考证人口数	707.5	495.5	1,203.0	-62.7%
1910年新政人口数	807.5	470.4	1,277.9	6.2%
1910年新政人口修正值	954.5	716.1	1,670.6	30.7%
1953年普查人口数	1,583.4	1,411.0	2,994.4	79.2%

数据来源：曹树基《中国人口史》第5卷《清时期》第704页。

图1是以年份为X轴，以考证人口和修正值人口数为Y轴的柱状图，并添加了趋势线。陕甘地区入清以来的人口长时段稳定持续、同治及光绪初年西北战乱造成的巨大的人口损失、光绪年间人口的缓慢增长以及民国以来的人口急速恢复等人口变动在图上都清晰可见。

¹ 葛剑雄主编、曹树基著：《中国人口史》第五卷《清时期》，上海：复旦大学出版社2001年，第70页。

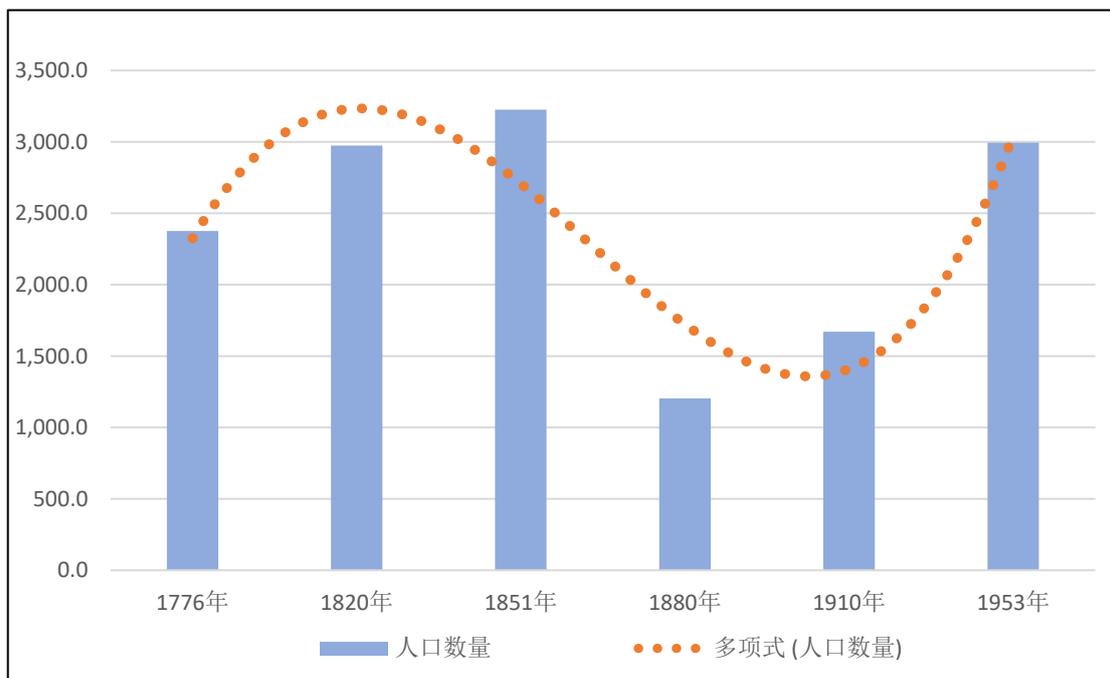


图 1 1776-1953 年西北陕甘两省人口规模变动

表 1 和图 1 比较直观地展示了 1776 年至 1953 年间，陕甘地区人口规模变动的大概趋势，从中可以看在这一个较长时段内，陕甘两省的人口从缓慢增长到急速减少，再到缓慢恢复的发展过程。表 2 是分府人口的汇总信息，从中可以看到研究区域内人口空间分布变动的更多细部细节。

表 2 1776-1953 年间 6 个时间切面上 8 组分府人口数据

人口单位：万人

府州	1776 年	1820 年	1820 年 R	1851 年	1880 年	1910 年	1910 年 R	1953 年
西安	242.3	296.2	294.4	325.1	103.9	206.2	206.2	463.8
同州	158.2	180.5	180.5	198.1	90.0	94.1	112.9	156.2
鄜州	24.8	31.4	27.2	30.8	14.1	10.4	17.8	21.8
兴安	10.0	121.4	121.4	131.3	67.3	95.3	95.3	164.9
商州	11.3	75.2	75.2	89.8	57.1	60.1	82.3	128.9
汉中	75.1	154.1	179.5	183.0	138.9	128.1	167.5	218.5
绥德	24.9	28.4	28.4	33.8	25.9	23.4	34.5	51.9
榆林	41.8	51.5	47.7	58.3	41.7	38.2	50.6	66.9

延安	56.0	63.8	63.8	68.9	30.0	15.3	15.3	65.5
凤翔	100.0	134.8	134.8	142.9	106.7	93.1	123.3	151.9
邠州	22.1	25.8	25.8	28.3	13.4	15.7	21.2	41.0
乾州	30.0	34.3	34.3	36.6	18.5	27.6	27.6	49.1
兰州	241.5	269.5	269.5	291.2	88.1	62.8	129.8	225.4
巩昌	340.1	379.5	379.5	410.0	75.9	60.7	118.6	225.1
平凉	230.0	234.0	253.9	274.3	32.1	37.6	57.6	133.0
庆阳	114.1	127.3	127.3	137.5	12.3	14.3	17.0	70.6
宁夏	135.3	139.3	151.0	162.9	17.0	28.9	22.4	89.0
甘州	81.0	81.4	90.4	97.6	18.8	24.5	28.5	51.5
凉州	134.8	150.4	150.4	162.5	45.8	41.0	71.6	135.9
西宁	63.8	70.9	70.9	78.8	26.4	41.3	41.3	125.3
泾州	80.0	83.8	89.8	92.1	23.5	18.3	24.5	61.3
秦州	77.9	86.9	86.9	93.9	93.3	83.7	116.4	160.1
阶州	34.0	37.9	37.9	41.0	47.1	36.4	64.9	84.1
肃州	40.5	45.2	45.2	48.8	11.6	16.8	19.0	36.6
安西	6.9	7.8	7.8	8.4	3.6	4.0	4.5	13.1
合计	2376.4	2911.3	2973.5	3225.9	1203.0	1277.8	1670.6	2991.4

（数据说明：1820年为《嘉庆重修大清一统志》人口数，1820年R是1820年修正人口数，1910年R是新政修正人口数。其他年份为曹树基考证人口数）

1776年至1953年间，陕甘地方政区变动较大，前后政区不同的人口数据要进行对比，需要统一在一个标准的政区切面上。曹树基先生在处理这一问题时统一使用了《嘉庆重修大清一统志》记载1820年的标准政区。分府人口数据的考证工作极为复杂，而把不同前后不同政区的人口标准化到1820年政区上来，处理过程更为繁琐。分府人口考证是曹树基先生《中国人口史》第五卷中的基础内容，尽管在局部细节方面存在诸多问题，并多被其研究者所诟病，但实际上，几乎所有这些个案数据，都可以在统计层面加以调整和消除。这一整套标准数据对相关的人口史研究提供了极大的便利，具有极其重要的学术意义。

人口统计数据本身就具有空间属性，使用 GIS 软件可以轻松把表格存储的统计数据，存储为以电子地图为背景的地理空间数据。进而根据人口规模大小，通过分级符号化的方式，实现数据可视化，从而更直观地显示人口数据在空间上的分布。图 2 显示了四个年份分级符号化后的陕甘两省人口空间分布态势。

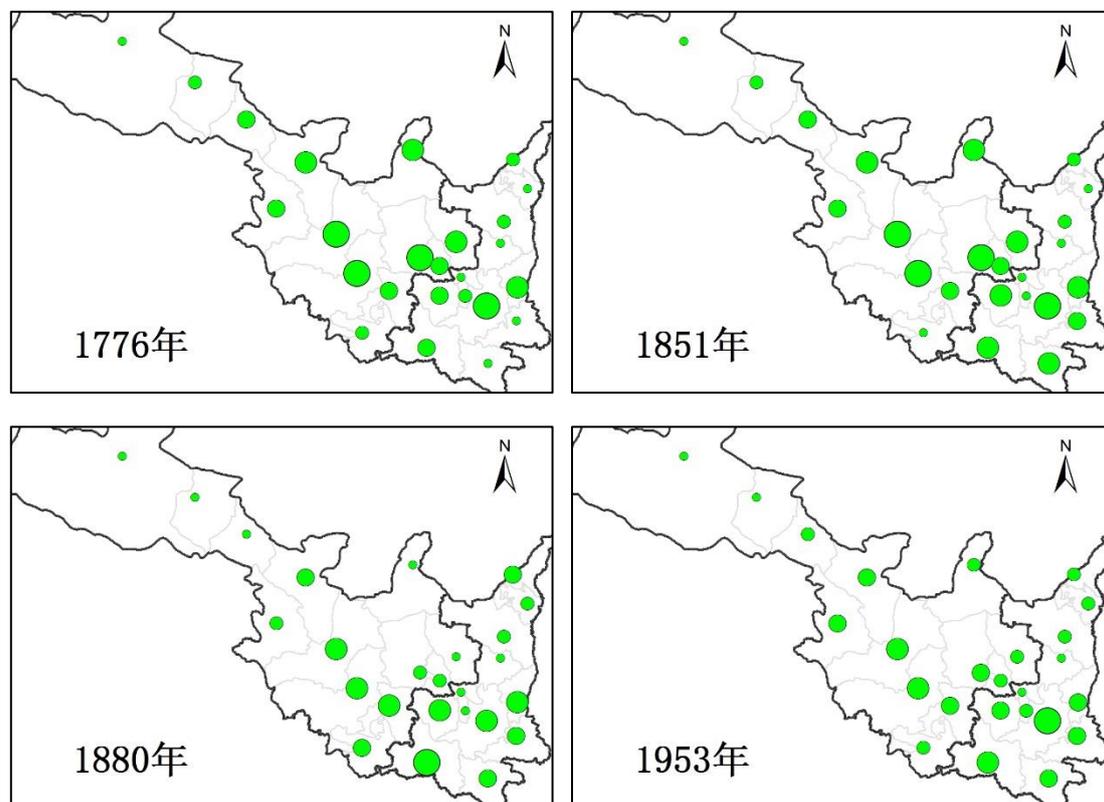


图 2 4 个不同时间切面上的陕甘分府人口空间分布

数据的分级符号化是指利用符号将连续的数据进行分类分级、概括化、抽象化的过程。这种方法可以同时直观地表达制图要素的数值差异和空间差异，增加对数据整体的认知。但是，数据的可视化只是数据表达的不同方式，本身没有任何学术意义。图 2 中任何一个时间切面的等级符号化数据都可以很清晰地展示当前数据空间分布状态。但是，当不同时间切面上分级符号化后的地图并列在一起时，反而使数据间的差异变得更加模糊。显然，单纯依靠分级符号，我们很难清楚地了解同一地理要素在不同时间切面上的变化。研究历史人口这样具有时间序列特征的空间数据，需要使用更为简洁的空间分析模，以便可以把复杂的数据简单化抽象化，从而可以更加清晰地看清数据之间的差异与变动趋势。基于此种考

虑，笔者使用人口重心模型来模拟 1776 年至 1953 年间的陕甘人口变动。

三、 人口重心模型的在西北地区的适用性检验

重心是物理学概念，指的是在重力场中，物体处于任何方位时所有各组成支点重力的合力都通过的那一点。假设人口所在区域为一均质的平面，而每一个人都是平面上的一个质点，具有相同的重量，则重心应为区域中距离的平和最小的一点，即一定空间平面上力矩达到平衡的一点。人口重心的概念由美国学者沃尔克 (F. Walker) 于 1874 年首先提出并使用，可提供某地区人口分布的简明、概括而又准确的印象，并可表明地区人口分布的总趋势或中心区位。

人口重心的计算是以某国家或地区中下属的若干子区域的人口数为权数，以子区域的代表地点的经纬度坐标为变量，分别计算经度和纬度的平均数，其交点位置即为人口重心。具体计算公式为：

$$X = \int x dp / p$$

$$Y = \int y dp / p$$

其中： X ， Y 分别为人口重心的横坐标和纵坐标， x ， y 分别为把这一地域划分为无限小的各区域的位置， dp 为相应的人口数。在实际应用中为便于计算，通常使用如下计算方法：

$$X = \sum p_i x_i / \sum p_i$$

$$Y = \sum p_i y_i / \sum p_i$$

其中 X ， Y 分别为人口重心的横坐标和纵坐标， X_i ， Y_i 分别为各小区域中心点的横坐标和纵坐标， P_i 为相应区域的人口数。显然，人口重心的横坐标是各子区域人口数与子区域中心点横坐标乘积的和与区域人口总数的除数。而人口重心的纵坐标则是各子区域人口数与子区域中心点纵坐标乘积的和与区域人口总数的除数。

子区域中心点的选择可以是各区域的几何重心，但人口空间分布受自然和人文环境直接影响，而现实中各区域自然与人文环境各不相同，且非均质。因此，人口从来就不是均匀的分布。一般情况下，计算人口重心时，会选择行政治所各子区域中心点。显然人口重心位置取决于人口的空间分布状态，如果人口的空间分布是均匀的，重心则应处于该区域的几何重心，偏移均衡状态的人口分布则

将导致人口重心旁移。在人口地理信息系统中运用人口重心分析的意义在于：从人口重心在一段历史时期的移动轨迹中,可以看出人口分布的变动的过程、方向、距离以及速度等。使用人口重心模型计算的陕甘地区 1776-1953 年间 6 个时间切面上的人口重心如下：

表 3 1776-1953 年间 6 个时间切面的陕甘人口重心

年份	经度值	纬度值	相对移动方向	移动距离
1776 年考证数	105.893859	35.786145	-	0
1820 年《一统志》数	106.191145	35.514627	往东偏南	40.41 公里
1851 年考证数	106.193818	35.525813	往北偏东	1.27 公里
1880 年考证数	106.801774	35.016425	往东偏南	79.04 公里
1910 年新政数	106.955449	34.958304	往东偏南	15.40 公里
1953 年普查数	106.565159	35.206570	往西偏北	45.99 公里

根据表 3 的经纬度数据，使用 ArcGIS 等 GIS 软件，我们可以很轻松的将以上 6 个年份的坐标值标注在地图上。图 3 显示了这 6 个时间切面上人口重心位置及转移的轨迹。

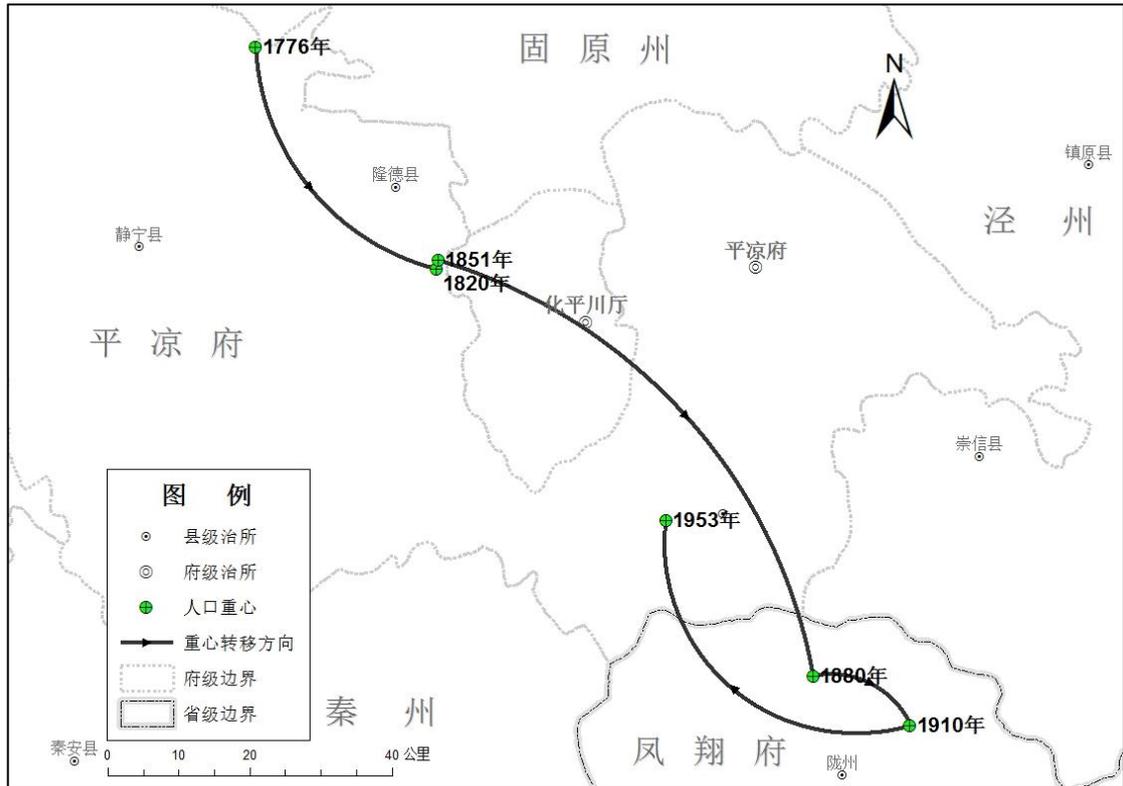


图 3 1776-1953 年陕甘区域人口重心

（数据说明：人口重心转移方向使用弧线并无计算意义，仅为避免直线链接重叠，增加制图美观之用，弧线表意与线段相同）

从表 2、图 3 可知，陕甘区域人口重心在 1776-1953 年的 177 年间，一直处于变动的状态之中，前后总共移动了 181.11 公里。从移动方向来看，6 个时间切面的 5 次相对移动，基本上可分为两个过程，即：其一、从 1776 年至 1910 年的 134 年间，总体呈现由西北向东南方向移动的趋势；其二、从 1910 年至 1953 年的 43 年间，移动方向发生逆转，改为从东南向西北方向移动。这 5 次移动的具体情况及其背后可能的历史原因如下：

1、从 1776 年至 1820 年的 44 年间，陕甘人口重心从西北向东南移动了 40.41 公里。这一时期人口重心向东南偏移的原因主要是陕南山区开发，外来移民大量迁入造成的。陕南山区北靠秦岭、南倚巴山，汉江自西向东穿流而过。群山怀抱之地，直至清初仍然覆盖大片原始森林，“虎迹狼蹄多于人迹，千里幅员，大半黄茅白苇。”¹其地分属商州直隶州、兴安府和汉中府三个行政区。陕南山区与川

¹ 康熙《西乡县志》卷九《风俗》。

北及鄂西北山地连为一片，是清代人口往山区迁徙运动的主要迁入地之一。康熙五十一年后，川陕总督通飭各地招民开垦，外来移民纷至沓来，“扶老携幼，千百为群，到处络绎不绝，……写地开垦，伐木支椽，上覆茅草，仅蔽风雨，借杂粮数石作种，数年有收。”¹随着移民迁入，仅乾隆年间，陕南就新设置了留坝（乾隆三十年（1765年）五月析汉中府凤县留坝一带地置）、孝义（乾隆四十七年（1782年）九月析西安府咸宁县孝义川地置）、宁陕（乾隆四十七年（1782年）九月析西安府同官县五郎关地置）三个厅。²现有研究表明，仅至乾隆末年，陕南山区接受的移民及其后裔总数就已高达 120 余万人。³这一人数，大约相当于同时期陕甘两省人口总数的 5%，陕西人口总数的 15%。外来移民入迁是这一时期陕西人口持续快速增长的一个极为重要的因素。⁴区域东南山地外来人口大量迁入，人口机械增长导致区域人口原来的空间分布的平衡状态被打破，重心往东南偏移。

2、从 1820 年至 1851 年的 31 年间，陕甘人口重心从西南向北偏东方向移动了 1.27 公里。相比于前一时段人口大幅度往东南移动，这一时期的人口重心基本保持了原地踏步状态，不再向东南方向偏移。其背后原因主要是 1820 年后，陕南山区外来移民入迁过程已基本完成，由大批外来人口迁入造成的区域人口增长不平衡的状态已经结束，人口增长方式改为内生性的自然增长为主。乾隆至道光初年，陕南山区新增政区较多，比如乾隆四十七年（1782年）九月兴安直隶州升为府，嘉庆七年（1802年）七月析汉中府西乡县渔渡路地方置定远厅，道光三年（1823年）四月，析兴安府安康县砖坪营地方置砖坪厅，道光四年（1824年）五月，析汉中府洋县与西安府盩厔县地方置佛坪厅。⁵县是历代地方行政区划的基本单位，集中建县，表明该地方的开发已臻成熟。道光四年后，陕南不再有新设政区，这表明，区域开发与人口入迁已基本停止。人口重心向东北方向略有偏移的情况表明，在区域人口完全依靠内生的自然增长的情况下，区域东北部宁夏平原的人口增长速度，要略高于其他地区。

¹（清）严如煜：《三省山内风土杂识》。

²周振鹤主编、傅林祥等著：《中国行政区划通史·清代卷》，上海：复旦大学出版社 2007 年，第 366-368 页。

³葛剑雄主编、曹树基著：《中国移民史》，第六卷《清、民国时期》，福州：福建人民出版社 1997 年，第 128 页。

⁴薛平栓：《陕西历史人口地理》，北京：人民出版社 2001 年，第 260 页。

⁵牛平汉主编：《清代政区沿革综表》，北京：中国地图出版社 1990 年，第 435-452 页；周振鹤主编、傅林祥等著：《中国行政区划通史·清代卷》，上海：复旦大学出版社 2007 年，第 368-369 页。

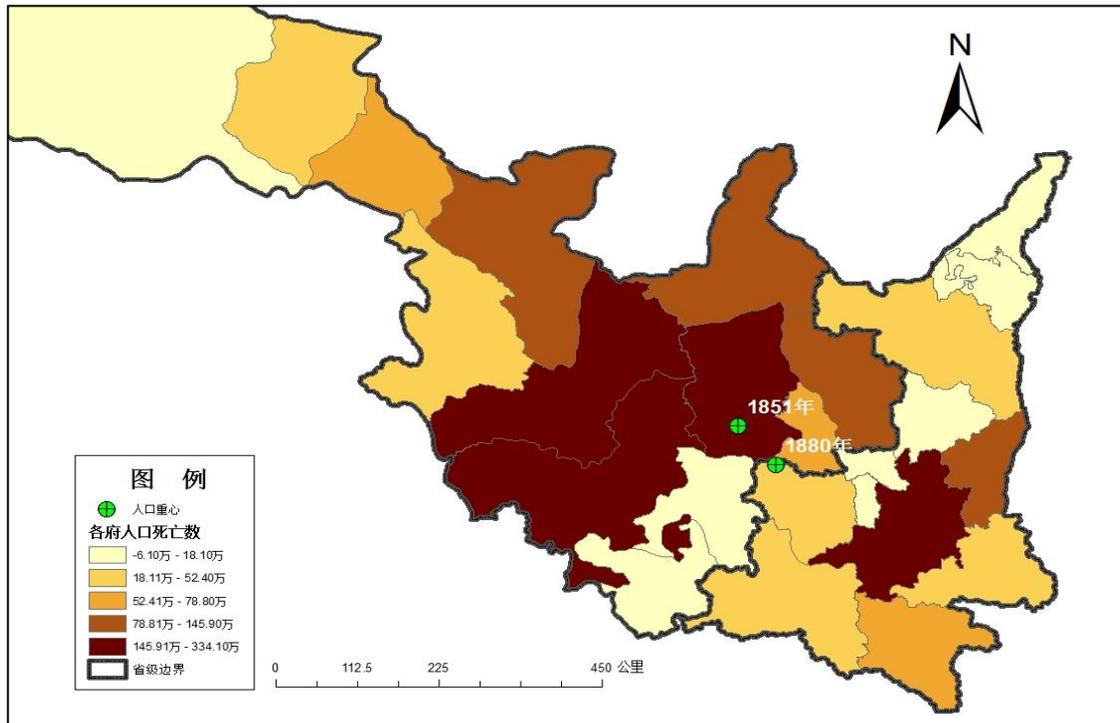


图 4 同治战争西北分府人口损失与人口重心

3、从 1851 年至 1880 年的 29 年间，陕甘人口重心从西北向东南大幅度移动了 79.04 公里。人口重心的这一移动方向与幅度，与同光年间西北地区的战争和灾荒有关。同治西北战争，源于关中，波及整个区域。十余年战乱造成了极其严重的人口损失，很多战前人口繁庶的沃野之区，战后都沦为焦土。其中，关中平原、河西走廊、宁夏平原、河湟谷等等，受影响最大，人口损失最严重。在区域人口普遍遭受严重损失的情况下，人口重心快速向东南大幅度向东南偏移，主要是两方面的原因：其一，虽然位于区域东南部的关中平原人口损失非常严重，但位于区域西北方向的兰州、巩昌、平凉等府人口损失更重，区域西北的河西走廊各府州以及区域北偏西的宁夏府等处，人口损失亦重；其二、位于区域东南部的阶（州）、秦（州）、安（康）、汉（中）等地区因战乱较少，大批邻域人口避难迁入，战时人口总数不但没有减少，反而所有增加。以秦州直隶州秦安县县城为例，清末人数高达三万，在整个甘肃省之中，其人口规模仅次于兰州省城，超过其他所有的府、州、厅、县。¹

4、从 1880 年至 1910 年的 30 年间，陕甘人口重心从西北向东南移动了 15.40

¹ 路伟东：《清代陕甘人口专题研究》，上海：上海书店出版社 2011 年，第 383 页。

公里。这一时期人口重心继续向东南方向偏移，是战乱之后区域东南部人口增长较快造成的。西北地区在灾祸之后，到处旷野，人稀地广，各地方官员都积极安置流民，招徕垦荒。但绝大多数地区收效不大，只有陕西关中地区，因为其特殊的地理位置，有大批客民涌入。这些移民大都来自临近的河南、湖北、四川、山东、山西、直隶等省。战后与外来移民迁出地距离较近的关中地区因为首先得到移民填补，人口增加较快，区域人口重心因此向东南方向移动。

5、从 1910 年至 1953 年的 43 年间，陕甘人口重心从东南向西北移动了 44.99 公里。这一时期人口重心的移动方向与 1776 年以来的陕甘人口重心移动方向相比，发生了逆转，由西北东南向转为东南西北向。造成这一现象的原因，主要是 1910 年后，位于区域东部的关中地区在外来移民大量迁入的推动下，人口增加较多，局部小区域人口已经趋于饱和。由此，外来人口开始越过陇山，继续向西进发，西北方向兰巩一带、河西走廊与宁夏平原等处的人口增长速度开始大于东南方向关中地区。但是，从人口重心移动的距离和速度看，直到 1953 年第一次人口普查时，整个区域人口不论从绝对数量还是空间分布状态来看，都远没有恢复到同治战前的水平。

综上所述，1776 年至 1953 年间陕甘区域人口重心移动轨迹与实际的区域人口空间分布变化吻合度较高，人口重心的转移方向与幅度完全可以找到合理的历史学解释。这表明，人口重心模型可以比较好的模拟较长的时段内陕甘区域人口空间变动趋势。

四、 地理调查表近 7,000 个分村数据支持的西北人口分析

传统典范式历史人口规模的研究，一般都是通过对大量相关史料的判读来进行概略的估算。曹树基先生认为，更多的样本数量和更高的样本精度可以有效提升最终的研究结论的可靠度。简单地说，就是估算一个省或更大区域的人口规模时，使用县一级的人口样本要优于使用府一级的人口样本，使用更多的县一级的人口样本要优于使用较少的县一级的人口样本。这一研究思路和方法，从统计的角度讲是正确的。在无法获取全部数据时，一般情况下，较多的样本容量（样本数）可以得到较高精度的估值。但是，这有两个重要的前提：其一，要有科学合理的抽样规则；其二，样本本身要足够准确。否则，更多的样本数，意味着更多

的样本误差，而更多样本误差累积，可能会使总误差超出可以接受的范围，最终导致错误的结论。实际上，历史人口数据很难做到这两点。因为所有的历史人口数据都是经过记录者和时间筛选之后的结果，研究者最终得到的样本数据都是不可控的。其二，几乎所有的历史人口数据，最初来源都是以收税为目的的保甲登记，这样的人口数据，一般都严重漏报未成年人、老人及女性等非纳税人口，并非实际的人口数。基于这些数据的研究，尤其需要谨慎处理。而过于主观随意的厅、县人口估计数，实际上缺乏真正实证的意义。

如何对曹先生的研究结论进行检验，是一个很有挑战性的工作。对于历史人口规模的估算，在大部分情况下，由于史料极其有限，研究的过程并不复杂。所使用基本方法也比较有限，就是历史文献学的方法，即史料的人工判读，和简单的数理统计。所以，如果没有新的研究手段或方法，仅使用原来的史料、按照原来的套路对现有研究数据进行重复检验，很难有新的发现，也很难证明其真伪。对于后来的研究者来讲，这种重复性的检验，既没有必要，也缺乏意义。在此，笔者使用人口重心模型，对曹树基先生给出的 1820 年《大清一统志》人口修正值和 1910 年新政人口修正值，分别进行检验。

（一）、1820 年人口数据的校验

《大清一统志》1820 年陕甘人口总数是 2,911.3 万，曹树基先生修正后的人口总数是 2,973.5，仅比原始人口数多了 62.2 万，约占《大清一统志》陕甘人口总数的 2.14%。

图 5 显示了两个不同的 1820 年陕甘区域人口重心，其中曹树基先生修正的 1820 年陕甘人口重心与《一统志》1820 年陕甘人口重心非常接近，测量数据显示，两者相距仅 1.95 公里。这一现象表明，曹先生修正的 1820 陕甘人口数与《一统志》的 1820 年陕甘人口数在空间分布上基本是一致的，两者不存在系统性的偏差。对于一个规模接近三千万的区域汇总人口数据而言，在样本数值极为有限，且大多出于个人估计的情况下，进行一个与原值变动幅度仅 2% 强的误差修正，本身的意义并不大。

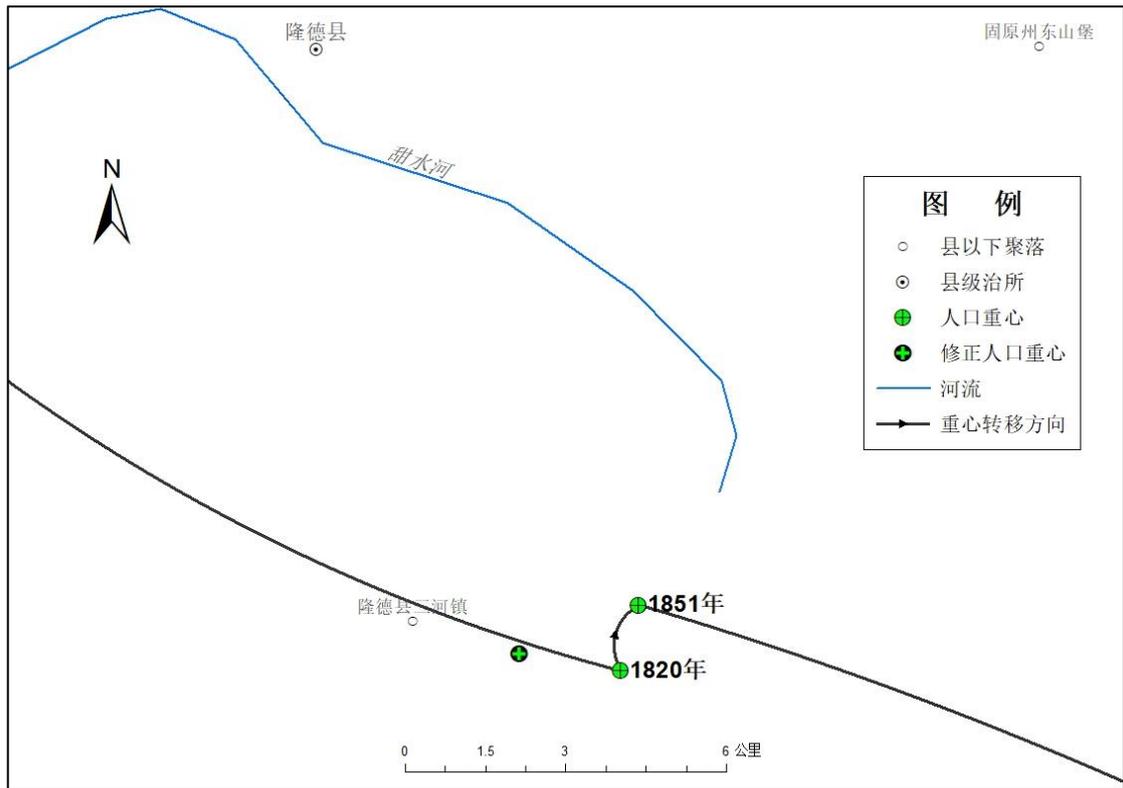


图 5 1820 年陕甘人口重心

人口信息对于一个国家的重要性不言而喻，除了征收赋税之外，它还是处理诸如确定学额、界定诸色人户法律身份以及赈济救灾等社会日常或重大突发事件的重要标尺。即便是现代国家，进行全面的人口普查，获取详实可靠的人口信息都是一件需要投入大量财力、人力及物力，且费时、费力的事情，人口普查体现了一个现代国家社会总动员的能力。对于任何一个传统农业社会的封建王朝来讲，获取人口信息的工作更是一项极为艰巨的任务。保甲体系是清代整个户籍管理体系的基础，除此之外，既没有可能，也没有必要，创建另一套户口管理制度。所以所有各层级的官方户口数据，都是由地方保甲统计之后逐层汇总上报的结果。方志中的户口数据与总志中的户口数据其实是同一来源。葛剑雄先生早就指出，“官方的总数正是各省上报后合计的结果，两者岂有不合之理？”¹因此，以各县方志的户口数来验证官方的省级户口总数，当然会得出了完全相同或相似的结论，这一做法，本身就缺乏逻辑意义。

¹ 葛剑雄：《中国人口发展史》，福州：福建人民出版社 1991 年，第 243 页。

（二）、1910 年人口数据的校验

1910 年陕甘地区新政人口数是 1,277.8 万，曹树基先生修正后的 1910 年陕甘人口总数是 1,670.0，比原始人口数多了 392.2 万，约占 1910 年陕甘新政人口总数的 30.7%。两个数值间的差距是极大的。曹树基先生对宣统普查数据进行了大幅度的修正，这表明了他对宣统人口调查数据的可靠性，基本持否定态度。侯杨方在其所著的《中国人口史》第六卷《1910-1953 年》中，则对这次人口调查给予了高度评价，他认为：“1953 年人口普查中的总人口性别比偏低（107.6）的原因，可能是由于此前的战争中男子死亡较多，但自 1949—1987 年间，中国人口的性别比始终徘徊在 105-108 之间，……1911 年时中国总人口性别比高达 121.6 几乎是不可能的，惟一的原因是女性人口的遗漏登记（特别是未成年女性），即便考虑到当时的女性人口死亡率较高的因素，1911 年中国总人口性别比也不会超过 110。另外，此次普查大约遗漏了占男性人口 5% 的男性未成年人口。”即便如此，总体而言，这次人口调查具有完善的制度保障、有力的人员支持，组织严密，实施标准，是中国历史上第一次现代人口普查。¹就陕甘具体数据来讲，甘肃性别比仅为 109.8，极为接近于正常水平，陕西性别比则比较高，约为 120.0。这表明，甘肃人口调查结果相当准确，陕甘人口调查中女性人口漏登者较多。

侯杨方所指 1911 年人口数即曹树基指的 1910 年人口数。宣统人口调查开始于 1909 年，结束于 1911 年。从甘肃地理调查表信息看，大部分地区上报的人口数实际上是 1910 年的。那么，何来评判两位学者对宣统人口调查数据两种相左的观点呢？GIS 人口重心空间分析模型，为我们提供了新的研究思路和方法。图 6 显示的是两个 1910 人口重心及变动趋势。

¹ 葛剑雄主编、侯杨方著：《中国人口史》第六卷《1910-1953 年》，上海：复旦大学出版社 2001 年，第 247-248、48-54 页。

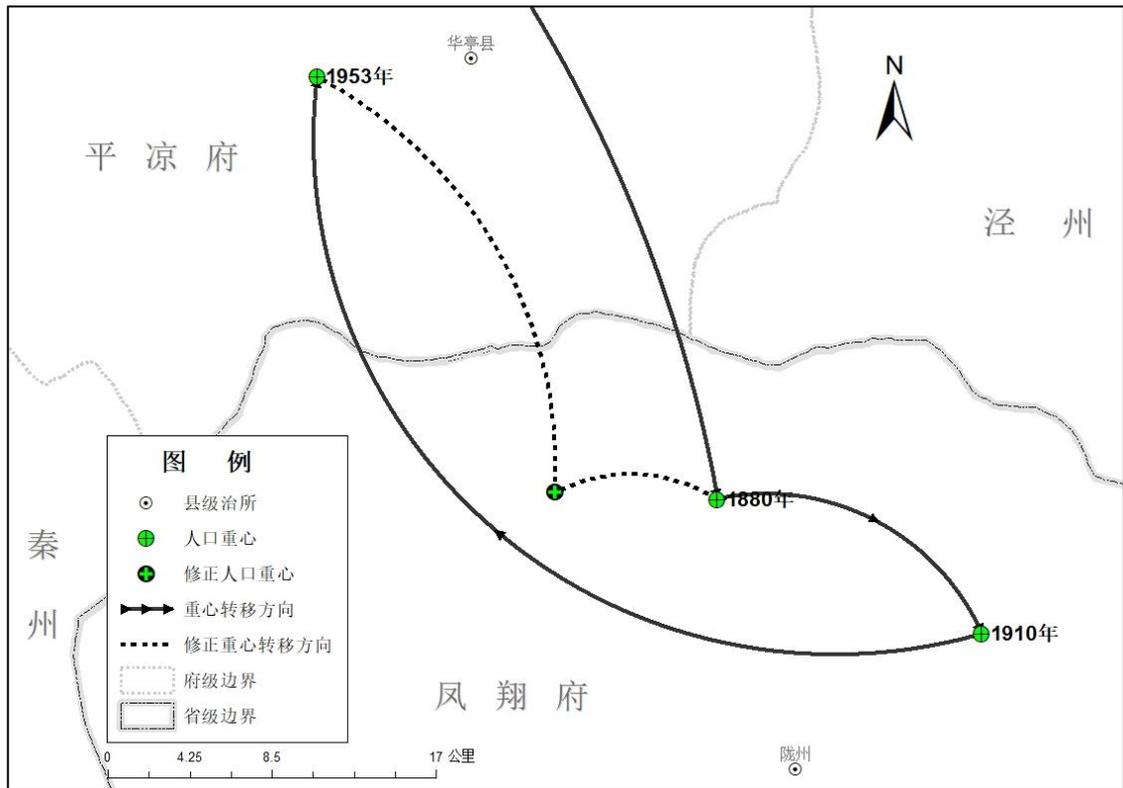


图 6 1910 年新政数和曹树基修正数人口重心移动趋势

从图 6 上可以看到，以 1880 年人口重心为参照点，新政 1910 年人口重心与曹树基先生修正后的 1910 年人口重心，出现了完全相反的走势。前者在 1880 年后继续往东南偏移了 14.5 公里，后者则相反，掉头往西北偏移了 8.3 公里。如果曹树基先生校正后的 1910 年人口数据是正确的，那表明以下两点推论至少有一个是肯定的：其一、1910 新政人口数在整个陕甘区域的西北部存在严重漏报现象，人口数严重偏低；其二、1880-1910 年间，整个陕甘区域的西北部人口增长率远大于东南部。然而，实际的情况是，以上两点都是错误的，具体理由分述如下：

1、关于陕甘区域西北地方宣统人口调查新政人口数的漏报问题。陕甘区域西北部主要是甘肃省，尤其是人口最集中的河西走廊地区。在这一省区，笔者发现了宣统人口调查最原始的普查档案地理调查表。研究表明，甘肃新政人口调查实施严格，人口调查数据质量较高，不存在严重漏报的情况。¹

¹ 路伟东：《宣统人口普查“地理调查表”甘肃分村户口数据分析》，《历史地理》第 25 辑，上海：上海人民出版社，2011 年 4 月，第 402—412 页；路伟东：《晚清西北人口五十年（1861—1911）》，上海：复旦大学出版社，2017 年，第 35—74 页。

从男女性别比上看，新政人口数甘肃性别比为 109.78，陕西为 120.00。¹两者相较，甘肃男女性别比更接近于正常值，陕西男女性别偏高，说明陕西新政人口数中女性人口漏报更多一些。仅就这一点来看，修正后人口重心也应该向东南方向，而不是西北偏移。

2、关于 1880-1910 年间，整个陕甘地区的西北部的人口增长率度远大于东南部。从前所述 1820 年至 1851 年的陕甘人口重心分布来看，外来人口入迁导致的人口机械增长因素消除后，区域内部的人口自然增长速度，相差不大。如果 1880 年后，陕甘区域的西北部人口增长率度远大于东南部，那就说明，1880-1910 年间，西北地区的外来人口迁入数量和速度，要远大于东南地区。

然而，实际的情况是，由于战后移民迁出地主要是陕甘区域东部及东南部的河南、湖北、四川、山东、山西、直隶等省，移民西迁入潼关后，首先补充的是靠近迁出地，且人口稀少，自然条件优越的关中地区，陕南山地。而这些区域人口基本饱和，外来移民开始大规模的越过陇山，继续涌入更西部的地区，至少是 1910 年之后的事。

综上所述，显然，曹树基先生对于陕甘地区 1910 年新政人口数的判断出现了错误，他对新政人口数的修正是不可靠的。GIS 提供的人口重心模型，是我们实现对这一研究数据进行校验，并证伪的重要手段。

宣统人口调查是中国现代人口普查的萌芽，其普查目地、登记方法及人口数据等，较之前的户口数有本质区别。在此之前的户口数据，不论出于总志、方志，还是清册、通典，抑或是其他官私文献，实际上都属于同一个系统，来源都是一样的，那就是以收税为目的的、依托保甲体系逐层汇总的、最终记录于户部清册中的户口数。在这样一个数据体系里面，用方志的数据去校验总志的数据，用汇总的地方数据，去检验更高层级的区域或全国数据，不论在什么样的精度上进行讨论和分析，最后的结论都应该是相同或相近的。但当用同样的方法去检验宣统普查人口这样一个系统外的数据时，问题就出现了。

五、 余论

从学科分类上讲，历史人口学是现代人口学一个分枝，但从研究方法与研究

¹ 王士达：《民政部户口调查及各家估计》表十一，见《社会科学杂志》第 6 卷第 2 期。

侧重点来看，两者不存在很大的差别，对于历史人口学家来讲，研究工作首先需要解决的是历史人口数据的研究考证，而不是现代人口学的统计分析。历史人口的研究过程，实际上主要就是人口数据的获取过程。历史人口数据往往不符合现代人口学的规范，考证一个看似简单的人口规模已经穷尽所能，至于人口年龄结构、性别结构、职业结构等这些详细的人口学指标，则更是难上加难。因此，现代人口学的统计分析手段对于人口史研究更深层次的需求，往往无能为力。

历史人口数据是具有时间序列的空间数据。历史人口的时间序列、属性信息与空间信息脱节，将严重制约着各种信息，如环境数据、自然资源数据、经济统计数据等的集成。解决这一问题的重要方法就是进行人口数据空间化。以管理、分析空间数据见长的 GIS 的研究方法和技术手段，是我们可以选择的重要依赖。

GIS 的核心是数据，而优质的历史人口数据正是历史人口研究所长。使用人口分布的数据模型，结合传统的人口专题地图表示方法，可以更好地反映人口分布的特征。GIS 支持下的 1776-1953 年陕甘区域人口长时期变动规律分析研究个案表明，GIS 与历史人口学可以很好地结合起来，进行交叉研究。