

基于队列变化的小区域人口预测及注意的问题

Small-Area Population Projections Using Cohort Methods and Application Issues

复旦大学 王桂新

Prof. WANG Guixin, Fudan University



城市大数据及城市仿真论坛2018



北京 2018年9月20日

讨论的主要问题 Content

引言 Introduction

1、队列人口预测原理与方法

Principles and Methods of Cohort Population Projections

2、小区域人口预测的问题及其解决方法

Problems and Solutions to Small-Area Population Projections

3、小区域人口预测应用举例

Application of Small-Area Population Projections

4、人口预测及预测结果的特点与注意的问题

Features and Application Issues of Population Projections



引言 Introduction

(1) 进行小区域人口预测的意义 Importance of Small-Area Population Projections

基本公共服务设施布局 如便利店的配置

气候变化应对 如碳排放

城市仿真 如PM2.5影响的人口

城市精准治理 灾害防治

(2) 人口预测方法的发展 Development of Population Projection Methods

人口学的发展

人口预测方法的发展

数据挖掘、计算方法的发展——

人口预测——人口学与数据、计算机科学的结合



1、队列人口预测原理与方法

Principles and Methods of Cohort Population Projections

1.1 人口变动规律与地区人口平衡方程式

Rules of Population Change and Demographic Balancing Equation

(1) 人口变动规律 Rules of Population Change

- 方向性，即人口变动一般都是沿一定方向进行的。
- 惯性，即人口变动一般都表现出由此前变动所决定、与原来方向相同的变化趋势。
- 阶段性和连续性。这是由人的生命周期所决定的，它使每个人在一生中的生命过程中都具有一定的阶段性和连续性。
- 相关性，即各人口要素及其发展变动都具有一定的内在联系。

人口变动的这些规律，是我们能够进行人口预测的客观基础和基本前提。



(2) 地区人口平衡方程式 Demographic Balancing Equation

人口变动的内涵如图1所示。一个地区的人口变动可用以下方程式表示：

$$P_t - P_o = (B - D) + (I - E) = \Delta N + \Delta M$$

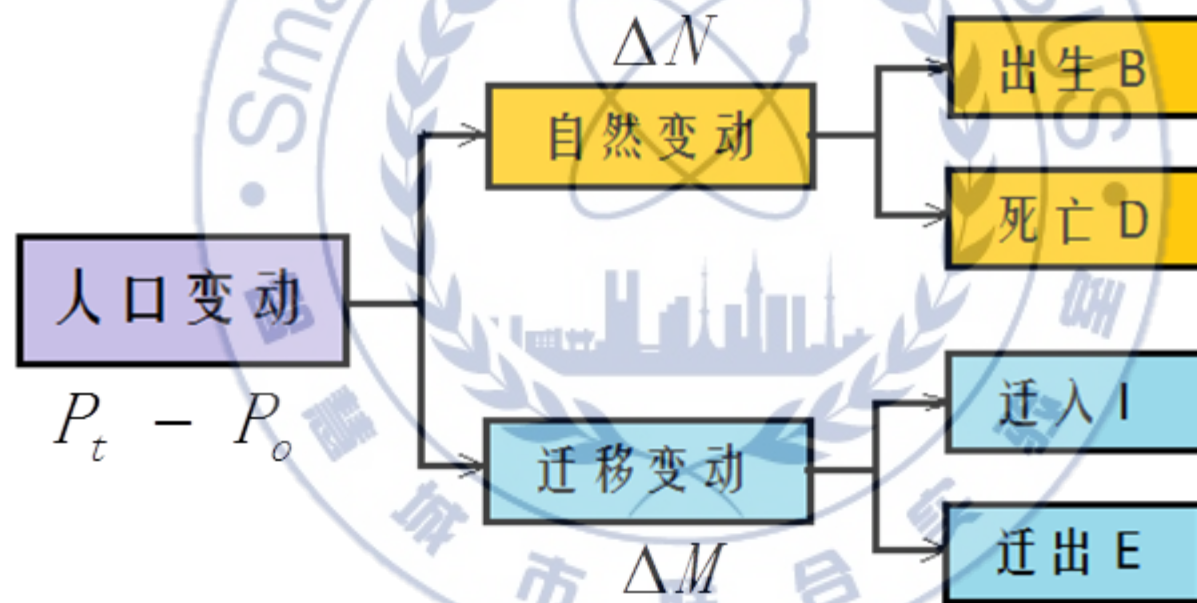


图1 人口变动的内涵

■ 人口预测方法简介 **Methods of Population Projection**

参考国际人口学会（IUSSP）和日本人口学家的分类，我把人口预测方法概括为以下几种（王桂新，2000）：

- （1）数学方法
- （2）队列要素法（又称同批人生存法、年龄移算法等）
- （3）多区域矩阵法（转移矩阵法）
- （4）社会经模型法
- （5）其他方法

各种预测方法的长短、优劣。

另外，还可以划分为：单区域人口预测与多区域人口预测。



2.1 队列人口预测思路与方法

Ideas and Methods of Population Projections

根据以上人口变动规律和地区人口平衡方程式，人口年龄增长（岁）与时间（年）推移具有一致性，即在封闭人口条件下，某年 x 岁的人口数与从 x 岁活到 $(x+1)$ 岁的生存概率之乘积就是下一年 $(x+1)$ 岁人口数。再考虑迁移因素，以此就可以推算出下一年1岁以上各年龄组人口。同时，根据妇女生育率和育龄妇女人数可以推算出出生的0岁人口数；进一步根据出生性别比，即可推算出0岁男、女人口数。





图2 队列预测法基本结构框图

(1) 队列要素法 Cohort Component Method

◎ 即根据人口变动3要素（出生、死亡、迁移）的假定，预测将来每一队列人口的方法。

队列，即同一年份出生的人（出生队列），又称为同批人。

人口变动3要素：

直接改变未来人口的，是出生、死亡、迁移3要素。其他因素，如结婚、离婚、患病率、劳动力需求等，都将通过3要素间接影响未来人口。

队列要素法

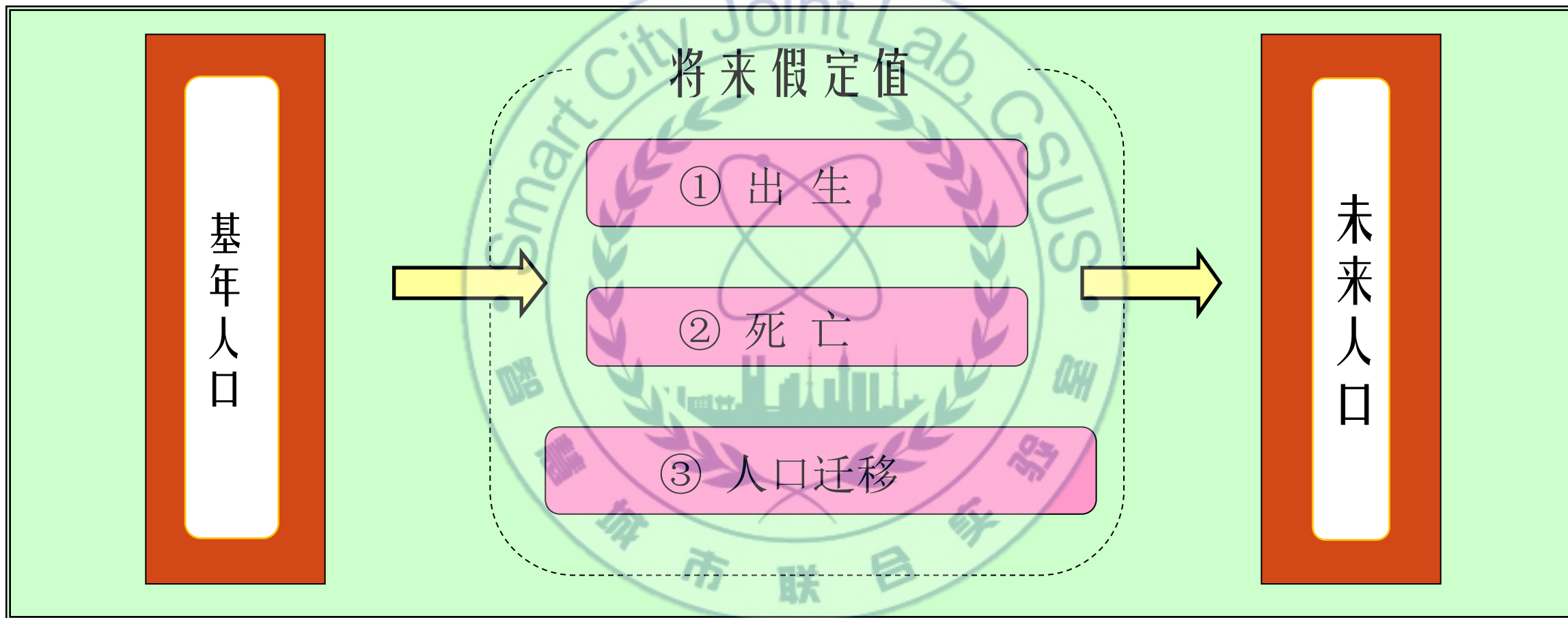


图3 队列要素预测法框图

队列要素法的人口预测计算公式 Formulas for population projections

具体计算公式如下：

$$M_x^{t+1} = M_{x-1}^t \times (1 - {}_m q_{x-1}^t) + M_x^t \times {}_m m_x^t$$

$$F_x^{t+1} = F_{x-1}^t \times (1 - {}_f q_{x-1}^t) + F_x^t \times {}_f m_x^t$$

$$M_0^{t+1} = {}_m b_0^{t+1} \times \sum_{x=15}^{49} F_x^t \times f_x^t + M_0^t \times {}_m m_0^t$$

$$F_0^{t+1} = {}_f b_0^{t+1} \times \sum_{x=15}^{49} F_x^t \times f_x^t + F_0^t \times {}_f m_0^t$$

$${}_m b_0^{t+1} = \frac{r_0^{t+1} / 100}{1 + r_0^{t+1} / 100}$$

$${}_f b_0^{t+1} = 1 - {}_m b_0^{t+1}$$

其中， M_x^t ：t年x岁年龄组男性人口数；

F_x^t ：t年x岁年龄组女性人口数；

${}_m q_{x-1}^t$ ：t年x-1岁年龄组男性死亡概率；

${}_f q_{x-1}^t$ ：t年x-1岁年龄组女性死亡概率；

${}_m m_x^t$ ：t年x-1岁年龄组男性净迁移率；

${}_f m_x^t$ ：t年x岁年龄组女性净迁移率；

f_x^t ：t年x岁年龄组妇女年龄别生育率。

${}_m b_0^{t+1}$ ：(t+1)年0岁人口中男性人口比重；

${}_f b_0^{t+1}$ ：(t+1)年0岁人口中女性人口比重；

r_0^{t+1} ：(t+1)年0岁人口性别比。

(2) 队列变化率法 Cohort Change Rate Method

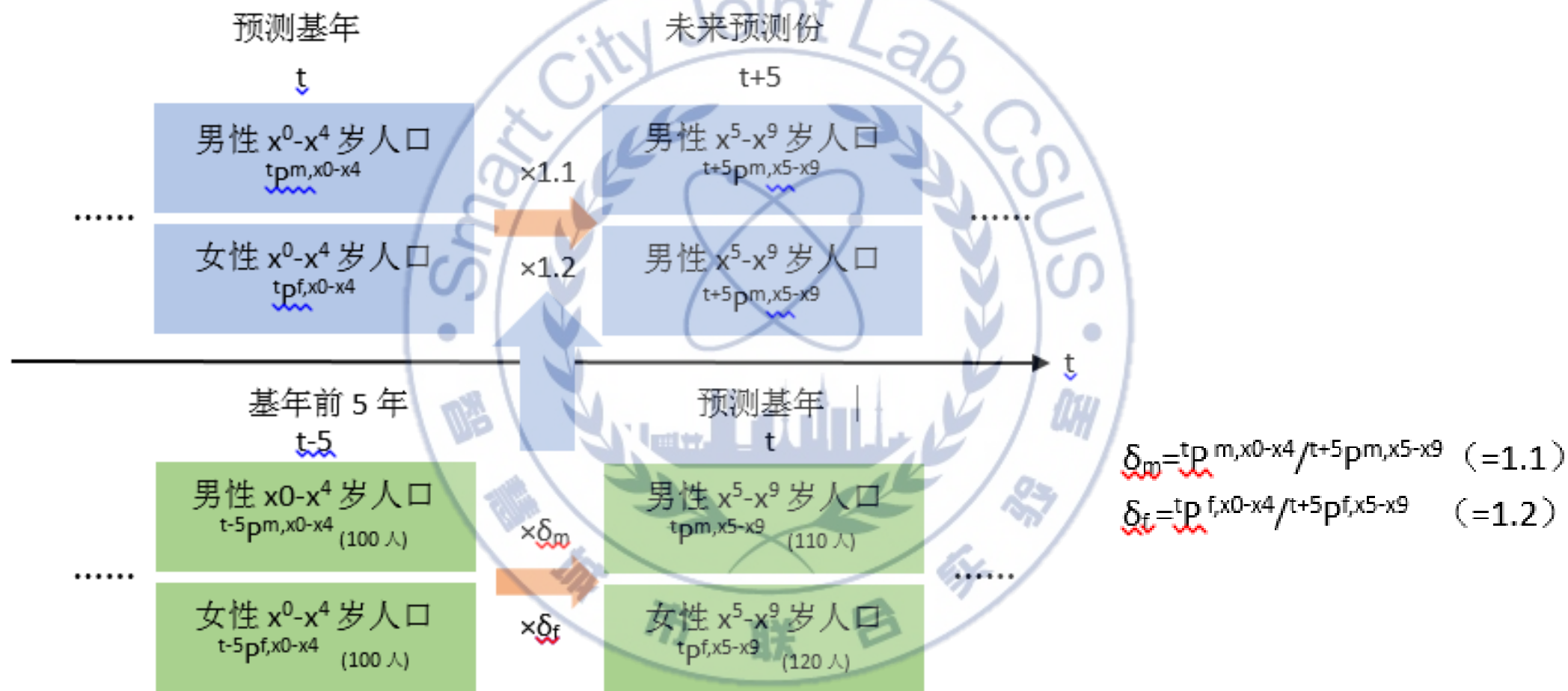


图4 队列变化率法概念图

队列变化率法的人口预测计算公式 Formulas for population projections

5~84岁人口:

$${}^{t+5}P_i^{g,x+5\sim x+9} = {}^tP_i^{g,x\sim x+4} \cdot \frac{{}^tP_i^{g,x+5\sim x+9}}{{}^{t-5}P_i^{g,x\sim x+4}} = {}^tP_i^{g,x\sim x+4} \cdot R_i^{g,x\sim x+4} \quad (1)$$

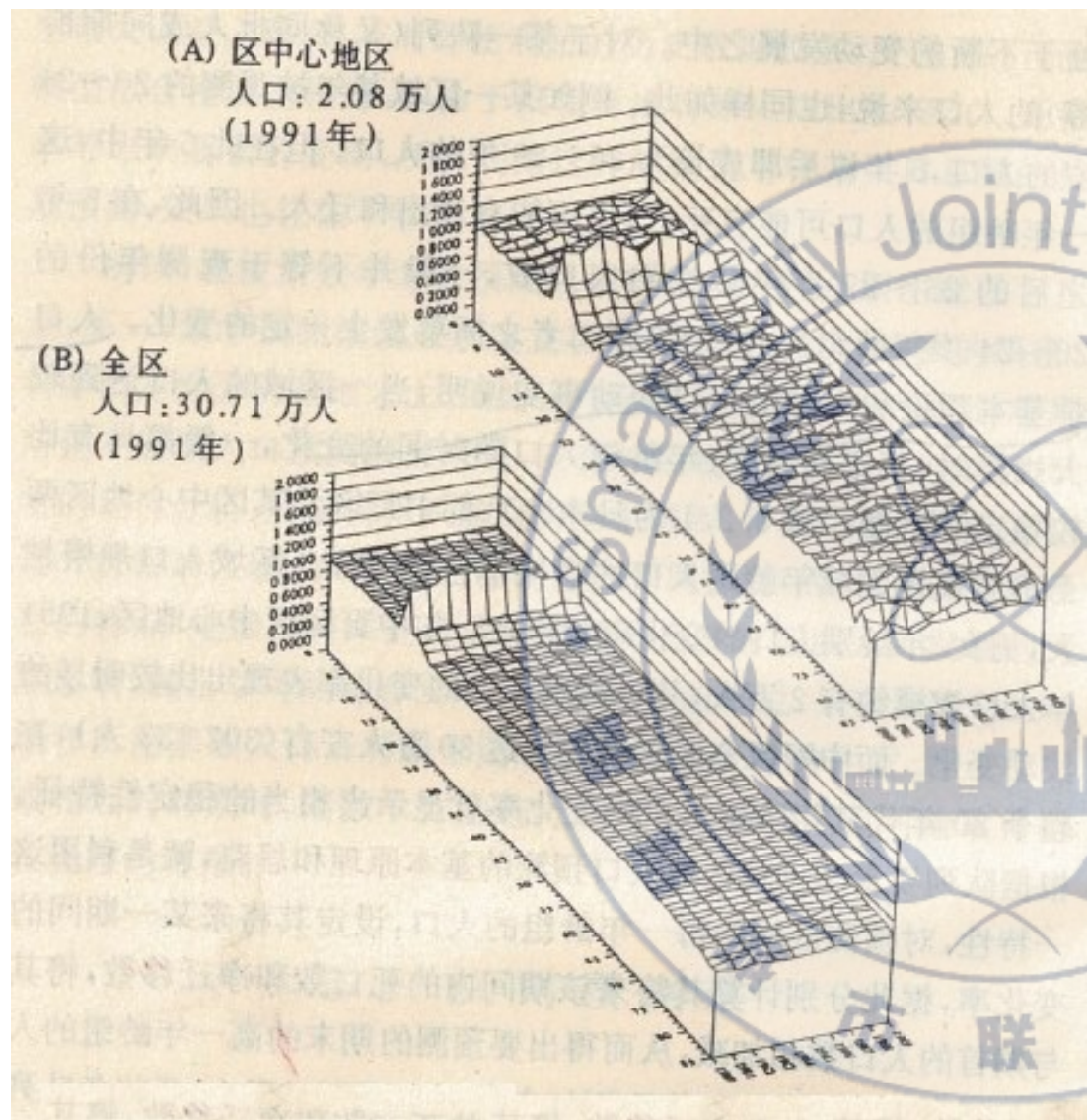
85岁以上人口:

$${}^{t+5}P_i^{g,85\sim} = {}^tP_i^{g,80\sim} \cdot \frac{{}^tP_i^{g,85\sim}}{{}^{t-5}P_i^{g,80\sim}} = {}^tP_i^{g,80\sim} \cdot R_i^{g,80\sim} \quad (2)$$

0~4岁人口:

$${}^{t+5}P_i^{f+m,0\sim4} = {}^{t+5}P_i^{f,25\sim39} \cdot \frac{{}^tP_i^{f,0\sim4}}{{}^tP_i^{f,25\sim39}} = {}^{t+5}P_i^{f,25\sim39} \cdot r_i^{f,25\sim39} \quad (3)$$

${}^tP_i^{g,x\sim x+4}$ 表示t年i小区域 $x^0\sim x^4$ 岁男女别 $g = \{m, f\}$ 人口。



小区域人口规模越大，其变化也比较稳定，应用变化率法预测人口效果也就越好。所以，使用变化率法进行人口预测时要适当考虑人口规模问题。

The larger population size



The better result

Population size to consider when using the method

图5 人口规模与小区域人口变化的稳定性

(3) 将来人口预测所需要的基础数据

Basic data required for future population projections

■ 队列要素测法所需要的基础数据 For Cohort Component Method

- 基年人口 · · 男、女年龄别人口 (人口普查或大规模人口抽样调查)
- 3个人口变动要素的假定值

出生

将来预测年份的女性年龄别人口出生率

将来预测年份的出生性别比

死亡

将来预测年份的男女年龄别生存概率 (将来生命表) 或男女预期寿命

人口迁移

将来预测年份的男女年龄别净迁移率

人口变动要素将来值的假定方法

Methods to Predict Future Values of Demographic Components

■ 传统预测法：依赖专家，假定高、中、低三方案（即中间值与上、下限值）。

优点——操作方便，可获得男女年龄别人口以及出生人口与迁移增长人口等。

缺点——不知道各个方案未来的发生概率，因而无法知道各个方案未来发生可能性的大小。即无法估算未来人口发展变化的不确定性（Alho 等，2005）。

■ 概率预测法：事后评估法

随机情景法（专家法）

贝叶斯概率法

时间序列模型法，又分为全概率预测方法和随机误差标度模型。

优点——一是对数据要求比较低；二是可对预测的未来人口变化做出相对合理的解释。如报告某方案未来发生的概率或可能性是多少。

缺点——对预测人员科学知识的要求比较高，所用的方法比较复杂。特别是其中的专家随机方案法，在本质上与高中低方案法是一致的，所以存在的问题也是也是相同的。



■ 队列变化率预测法所需要的基础数据 For Cohort Change Rate Method

- 基年人口 · · 男、女年龄别人口（人口普查或大规模人口抽样调查）
- 出生 · · 将来预测基年前5年的育龄妇女人数（特别是25-39岁妇女人数）
将来预测基年前5年的出生性别比

利用队列变化率法进行人口预测，所需要的基础数据比较少，在数据不足时多被采用，但输出的预测结果也相应比较少。

2、基于队列变化率法的小区域人口预测的问题及其解决方法

Problems and Solutions to Small-Area Population Projections

2.1 小区域人口预测的主要问题

Major Problems of Small-Area Population Projections

小区域的相对性。也就是说小区域的“小”是相对的。

The relativity of
small area

(1) 小区域人口规模小 Small Population Size

因为小区域人口规模小，就容易发生变化；容易发生变化，就容易出现过大偏小的异常变化值，甚至出现计算变化率的分母为0的情况。

(2) 小区域人口数据多是短期数据 Short-Term Demographic Data

因为小区域人口易变化，所以基本都是短期数据。如容易发生行政区划变动，导致人口变化快，多短期人口数据。



(3) 很难把握人口迁移状况 Availability of Migration Data

由于小区域过小，不容易把握其人口迁移数据。而且在人口普查或人口抽样调查中，小区域（如街道）之间的人口居住地变化一般都不被定义和统计为迁移。

(4) 涉及个人隐私 Personal Privacy

这是一个大问题。特别是近几年随着GIS的快速发展及手机数据的广泛应用，总或多或少地使用个人信息数据，这就难免涉及到个人隐私。作为个人毫无疑问要保护自己的私密信息，即使是所在区域或单位，也都负有保护居民或职工个人信息的信息的责任。如上海统计部门一般就不对外提供小区居民的详细信息以及居委会一级的电子地图。

2.2 小区域人口预测难题的解决思路和方法

Solutions to Small-Area Population Projections

(1) 开发适用于小区域人口预测的方法 Development of Appropriate Methods of Small-Area Population Projections

现在的解决思路只是根据小区域人口预测的问题和局限，简化预测方法，如把队列要素预测法简化为队列变化率预测法。

开发适用于小区域人口预测的方法有无可能？

(2) 评价、补正小区域人口数据及其方法 Evaluation and Mending Methods

以上说明，即使用最简单的队列变化率法预测小区域人口，仍面临很多需要解决的数据问题。如

1) 评价数据质量，补正数据问题 Importance of evaluation and mending
重要性



2) 数据补补 Data Mending

■ 无数据

空间插值

— 空间插值方法

■ 数据异常（如过大或过小）

空间平滑

— 空间数据平滑方法（根据空间统计学及流行病学方法）

二者实质上是一致的，所以可以采用同样的方法进行空间插值或空间平滑修正值。



3) 主要空间插值或平滑方法 Methods of Spatial Interpolation and Data Smoothing

■ 趋势面方法

趋势面分析即利用数学曲面模拟地理要素在空间上的分布及变化趋势的一种方法。

设某人口要素的实际观测值为 $z_i(x_i, y_i)$ ($i = 1, 2, \dots, n$)，趋势面拟合值 $\hat{z}_i(x_i, y_i)$ 为

$$z_i(x_i, y_i) = \hat{z}_i(x_i, y_i) + \varepsilon_i$$

式中， ε_i 为残差值。

采用趋势面分析方法插值或平滑异常值，一般采用二次或三次趋势面。



■ 根据平均值或中间值的平滑方法

以 R_i 作为小区域 i 的根据平均值或中间值计算的平滑推断值，其他小区域的观测值赋以权重的平均值可分别以下式表示：

$$\tilde{R}_i = \sum_j \tilde{w}_{ij} R_j,$$

$$\tilde{w}_{ij} = \frac{w_{ij}}{\sum_j w_{ij}},$$

使用这种方法，其权重的确定也有多种方法，在此不再赘述。

■ 非参数方法

非参数平滑方法有多种，一般多采用空间比率平滑方法。其思路是对计算队列变化率的分母和分子分别给出权重，再求出其平均值。和预先计算比率的平均值平滑方法相比，这更是一种有效平滑异常值的方法。使用这种方法，是用距离的函数确定权重。

■ 区域化平滑方法

区域化平滑方法就是使几个小区合计的结果趋向稳定的方法。与以上两种方法不同，这种方法不利用空间相关关系，而是根据聚类再合计的方法。即首先进行聚类分析分类，然后再对每一类计算平均值的方法。



■ 贝叶斯平滑方法

小区域由于人口少，其死亡数/人口之比率是不稳定的。对这种情况，一般是把人口规模较的小区域的值，作为事前信息纳入贝叶斯模型处理。由于不知道比率数据 R_i 是否服从普阿松分布，所以可考虑用以下方法解决。

小区域推定的代表性模型是Fay 和 Herriot1979年提出的Fay-Herriot模型：

$$R_i = \theta_i + e_i, e_i \sim N(0, \hat{\sigma}_e^2 / P_i)$$

$$\theta_i = x_i' \beta + v_i, v_i \sim N(0, \sigma_u^2)$$

上式中的 $\hat{\sigma}_e^2$ 是已知常数。数据 R_i 的方差，由于人口少的小区域大、人口多的小区域小，所以可以认为Fay Herriot模型是一个能较好表示人口数据特征的模型。 x_i 为说明变量矢量， β 为对应的参数矢量。

4) 数据补正的依据 Basis of Data Mending

■ 主要依据

地理学基本规律

人口变化规律

常识

■ 补正数据要探索适用方法

要根据数据存在的具体问题，有针对性地选用补正方法。

要注意选用方法的科学性和适用性。

3、小区域人口预测应用举例——以上海为例

Application of Small-Area Population Projections—Taking Shanghai as an Example

3.1 基本情况 Background

2010年 全市共214个街道乡镇单位，常住人口约2302万人。

市统计部门只提供分街道乡镇的分性别、年龄常住人口数；有些乡镇街道缺乏相关数据。

3.2 人口预测方法及数据处理

The Method and Data Processing of Population Projections

根据获得的数据情况，拟以队列变化率方法为基础进行预测。
数据评估，进行数据异常情况的补正处理。



3.3 预测结果 Results

根据队列变化率方法预测得到上海市各街道乡镇的分性别、年龄常住人口数，与另2种方法预测的全市常住人口数对比，可以发现用队列变化率方法预测的各街道乡镇常住人口合计起来的全市总人口数最大，但与最小的概率方法的预测结果比较，其绝对误差和相对误差2020年分别小于30万人和1%，2030年为60万人左右和2%以内，说明队列变化率方法对预测上海市分街道乡镇人口是有效的，预测结果也是合理、可用的。

表1 3种方法预测的上海市常住人口总数的比较

	2020年	2030年	说 明
队列变化率方法	2832	3133	各街道乡镇预测合计 高中低3方案 80%的概率，中间值
3方案预测方法	2828	3128	
3方案概率方法	2804	3072	

预测所得各街道乡镇人口的分布状况如下图所示。



2010年上海市常住人口密度特征：外环以内基本为每平方公里25000人的高人口密度区，大于35000人的超高密度区大致呈头上尾下、左右展翅的飞鸟型。

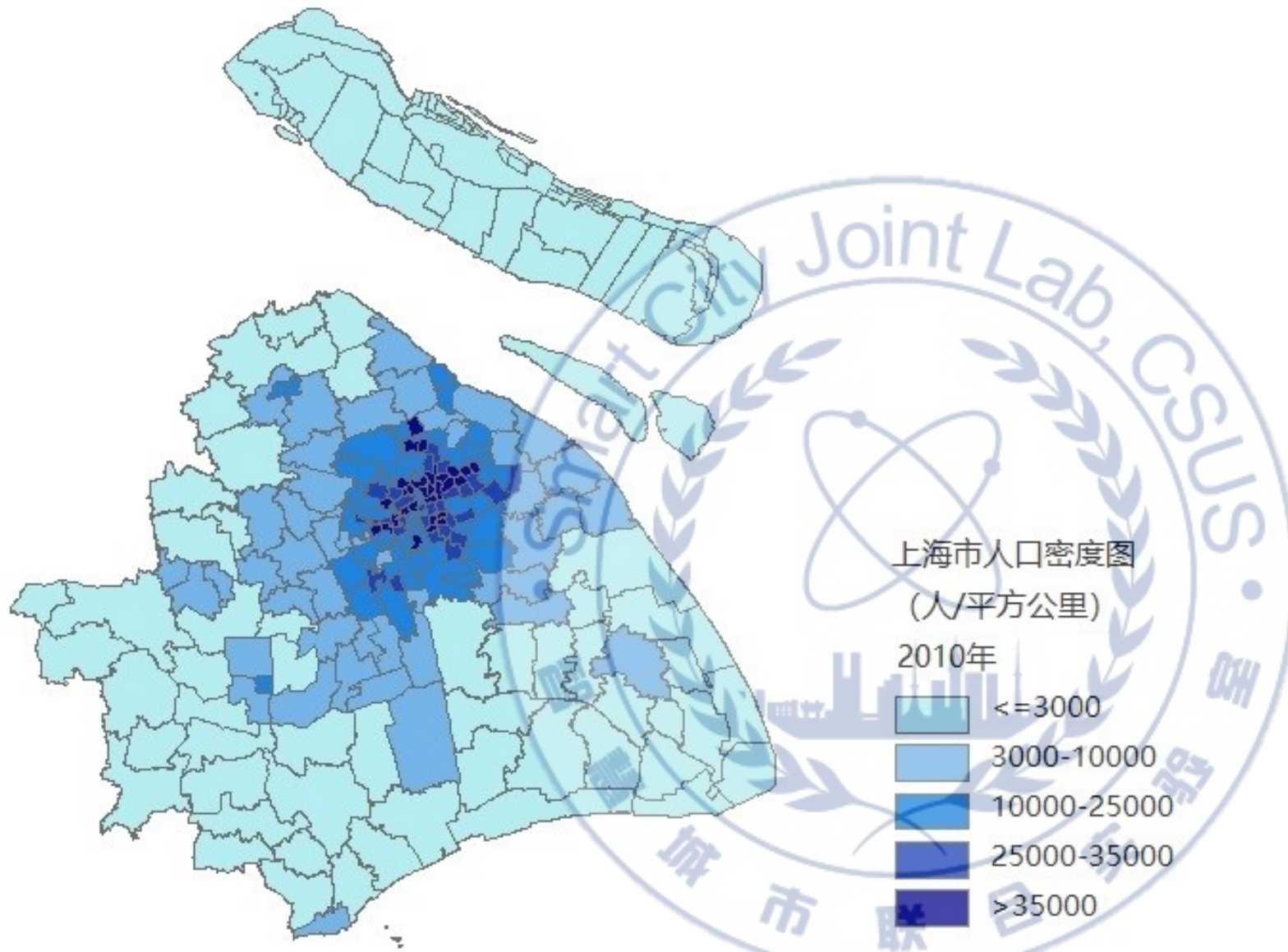
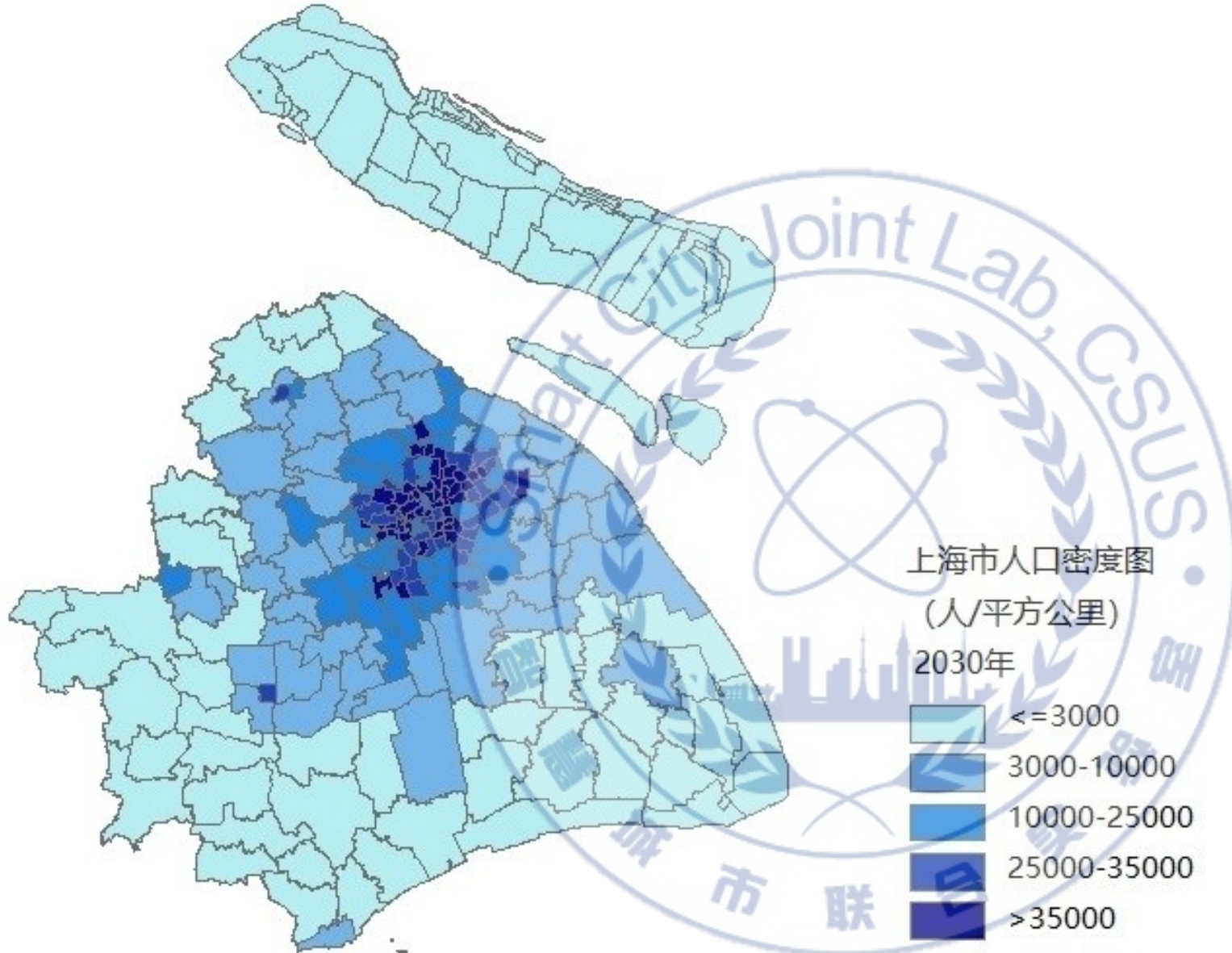


图6 2010年上海市常住人口密度 (人/平方千米)

Population Density of Shanghai: 2010

资料：第六次人口普查数据。



2030年上海市常住人口密度变化：每平方公里25000人的高人口密度区变为以东北西南为长轴的椭圆形；大于35000人的超高密度区也有明显变化。

图7 2030年上海市常住人口密度（人/平方千米）

Population Density of Shanghai: 2030

资料：作者预测。

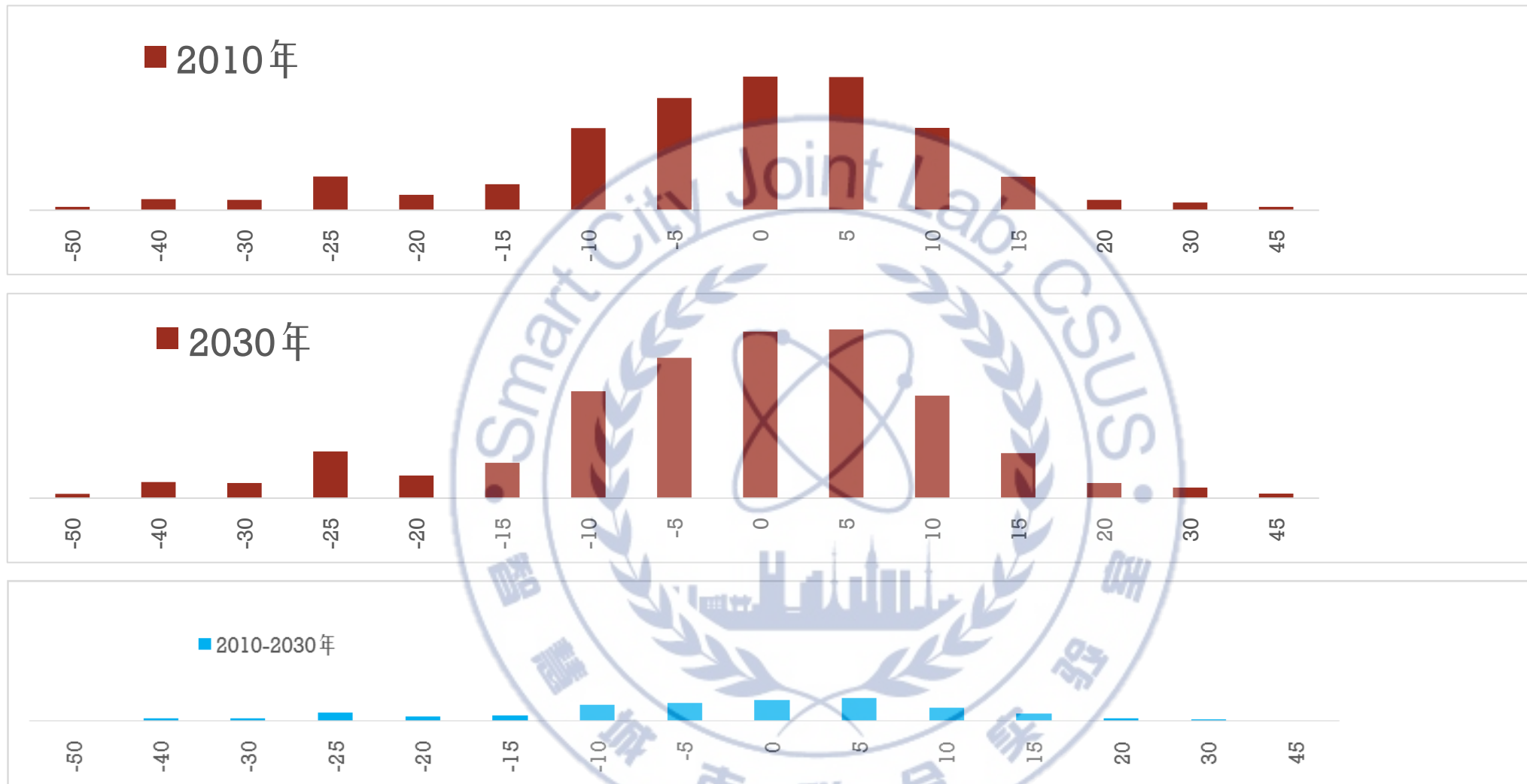


图8 2010-2030年上海市东西方向常住人口密度及其变化 (人/平方千米)

Distribution of Population Density in the East-West Direction

资料：作者预测。

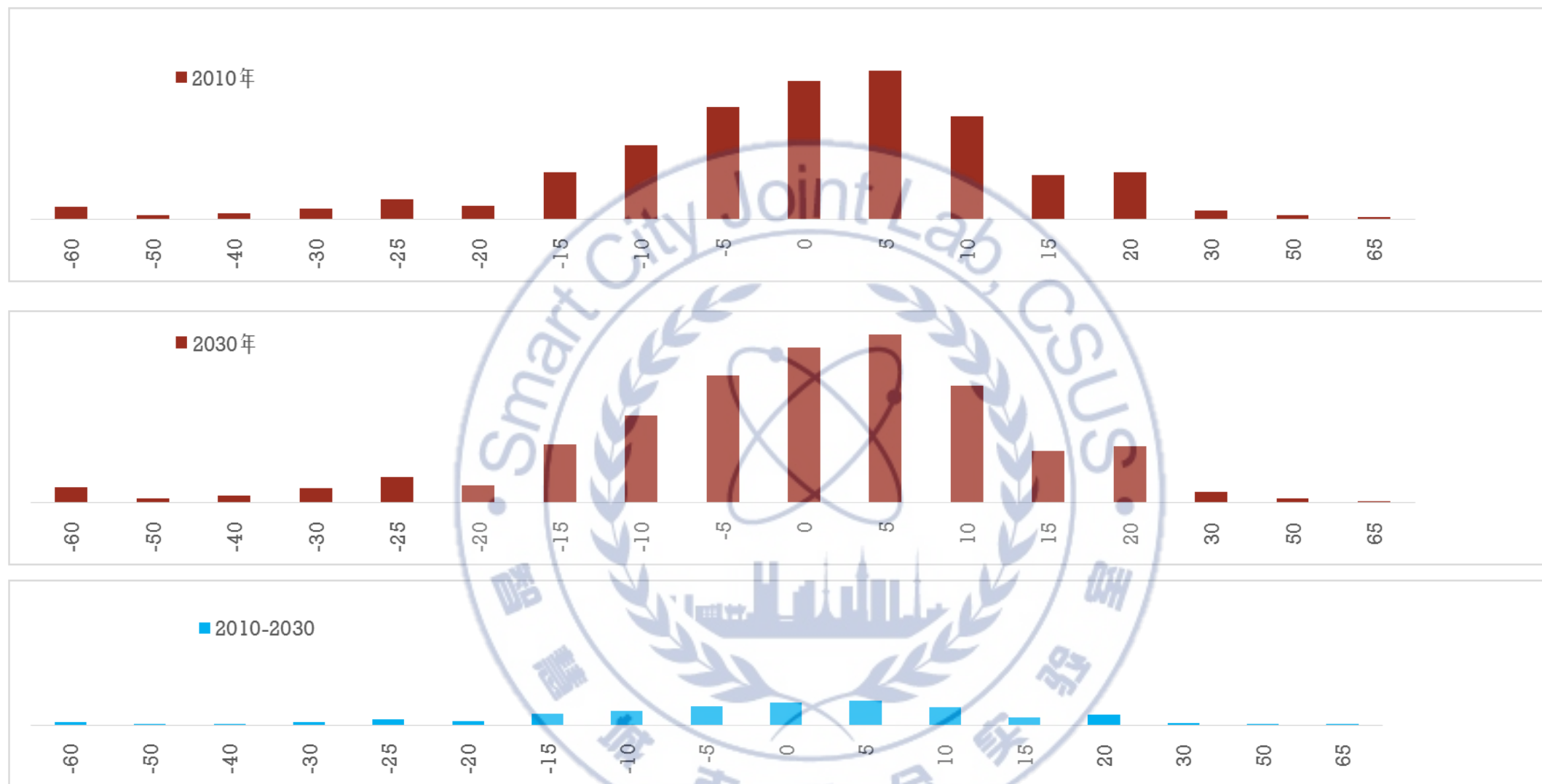
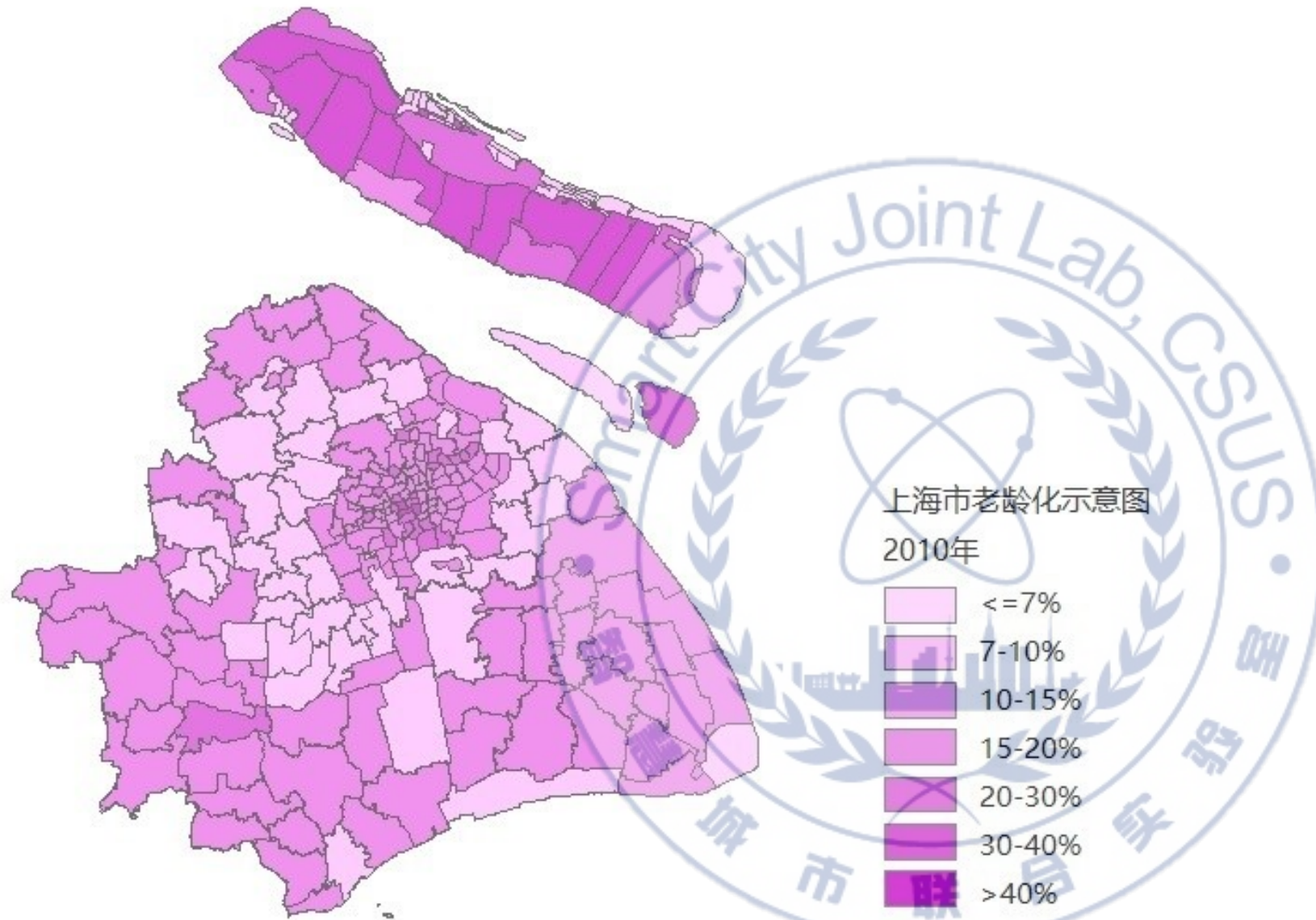


图9 2010-2030年上海市南北方向常住人口密度及其变化 (人/平方千米)

Distribution of Population Density in the North-South Direction

资料：作者预测。



2010年上海市常住人口老龄化率的分布特征：中心城（外环以内）和远郊区比较高，近郊区相对比较低。

图10 2010年上海市常住人口老龄化率的分布

Population Aging in Shanghai: 2010

资料：根据第六次人口普查数据。

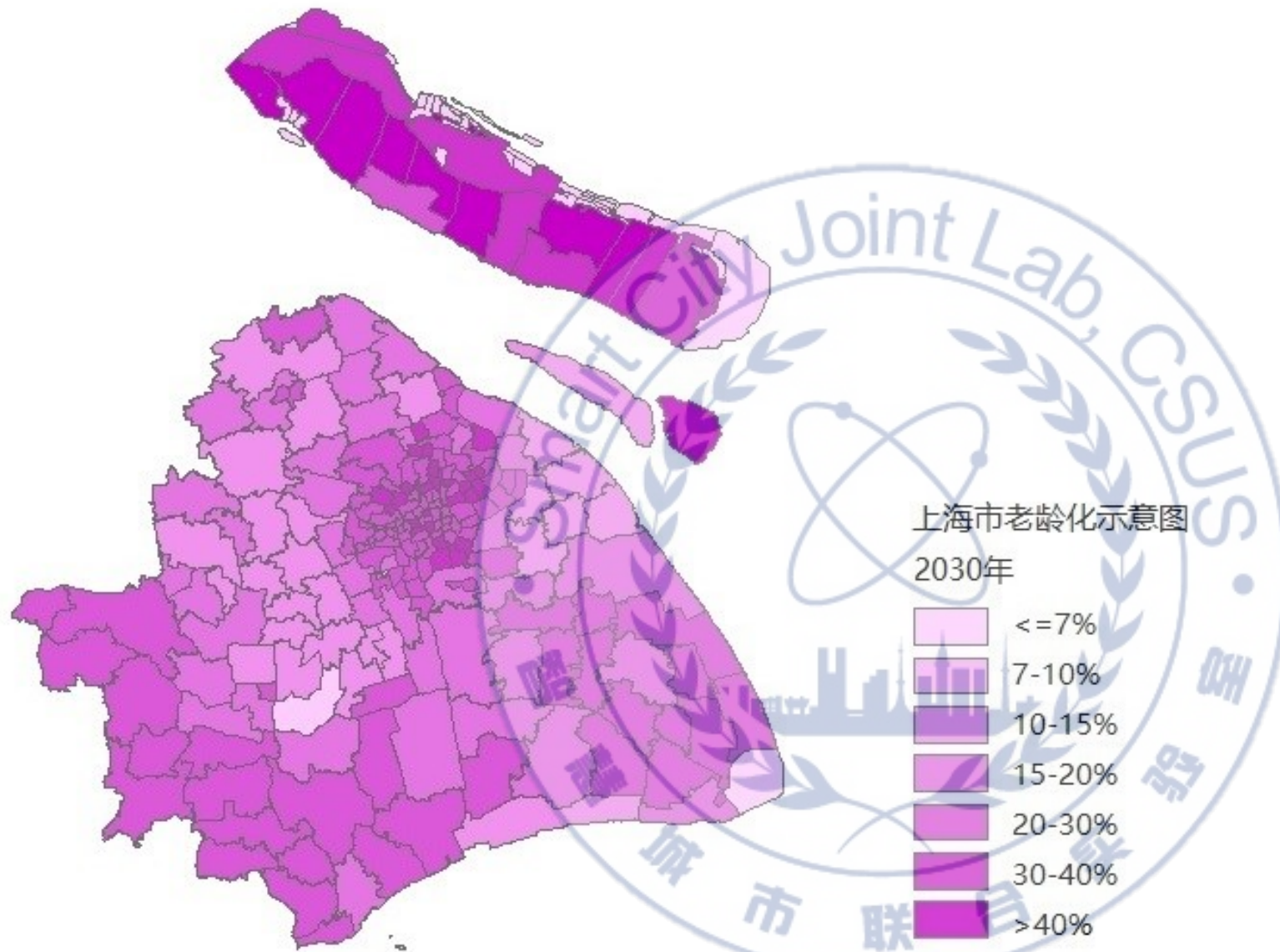


图11 2030年上海市常住人口老龄化率的分布

Population Aging in Shanghai: 2030

资料：作者预测。

2030年上海市常住人口老龄化率的分布特征：各地区老龄化水平普遍提高，以远郊区上升最明显。这提示我们未来养老服务设施的配置应该向郊区特别是远郊区倾斜。

4、人口预测及预测结果的特点与注意的问题

Features and Application Issues of Population Projections

4.1 人的特性与社会科学研究其特征

Characteristics of Human Beings and Social Science Research

■ 人的特性

高级动物、有思想、受价值观影响

■ 人类社会与自然界的不同

人的思维和价值观，决定了人的变化无常

人的变化无常又决定了人类社会的极端复杂性

人的变化无常使人口预测或模拟仿真比自然现象更加复杂、更难实现

■ 社会科学研究的特征

社会科学研究的是人类社会，需要面对人类社会的极端复杂性

宏观与微观的结合、个体与群体的差异，统计规律的发现



4.2 人口预测及预测结果的特点和注意的问题

Features and Application Issues of Population Projections

(1) 人口预测及预测结果的特点 Features of Population Projections

(特别是小区域人口预测)

■ 局限性 Limitations

采用以上任何一种方法的人口预测都是以一定的假设为前提进行的。在一定意义上，能否做出合理的假设，甚至比预测方法更重要，对预测结果也具有更大的影响。

假设的局限性决定了人口预测及预测结果的局限性。

■ 不确定性 Uncertainties

人的变化无常、人类活动的极端复杂性，以及人口预测假设的局限性，使人口预测很难得到十分精确的预测结果。

短期预测、预测大的趋势。



(2) 对人口预测结果应注意的问题 Application Issues of Population Projections

■ 预测者应注意的问题 For researchers

- 基础数据的收集、评估和补正
- 选择适用的预测方法（没有十全十美的方法，只有对某种情况相对适用的方法；但对某种情况，可以探索相对更佳更适用的方法）
- 保证假设的合理性
- 说明预测的假定条件和预测结果的局限性

■ 使用者应注意的问题 For users

- 注意预测的假定条件和预测结果的局限性
- 短期可作政策制定依据、长期只能作参考依据
- 考虑预测后社会经济环境实际变化的影响



谢谢聆听

THANK YOU FOR LISTENING!

