

用动态模糊聚类(ISODATA)作丹东大暴雨区划*

谢世俊 丛安佐 张心有

(辽宁省丹东市气象科学研究所)

丹东是我国北方降水最多的地区之一,大暴雨频繁,大小洪灾无年不有。在研究暴雨预报方法时,分区建立预报工具更为有效。在大暴雨气候分析、气象服务以及抗灾中,也需要对各县区大暴雨的频繁严重程度作出划分,以便确定工作重点。为此,本文应用1956—1980年共25年37个测点的资料来进行区划。这些测点包括国家和军队气象台站11个,其余26个为水文站,是从近百个站点中挑选出来的。测点距离一般在10—20km之间。各点资料年代相同,在选择过程中对资料的均一性进行过审查,统计方法相同。

1. 区划所依据的指标

x_1 , 24 h 最大降水量,取各点 25 年中的极值,分布情况如图 1。

x_2 , 大暴雨日数,取各点 25 年出现日雨量大于 100mm 的天数,分布情况如图 2。

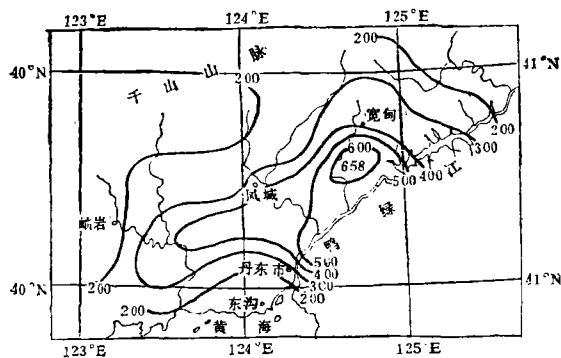


图 1 丹东地区 24h 最大雨量
(单位: mm)

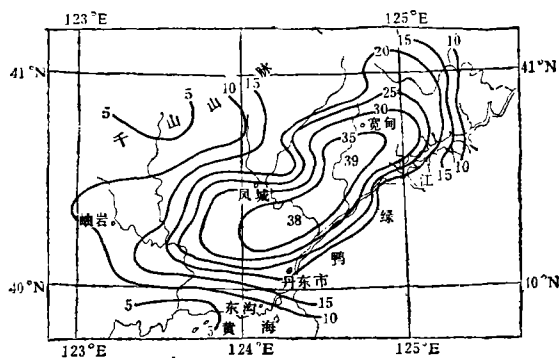


图 2 丹东地区 1956—1980 年大暴雨日数
(单位: d)

x_3 , 地形地势和地理位置。 $x_3 = f(\varphi, \lambda, h)$ 。指标 x_3 是纬度 φ , 经度 λ 和高度 h 的函

* 本文于 1984 年 1 月 30 日收到, 1984 年 12 月 11 日收到修改稿。

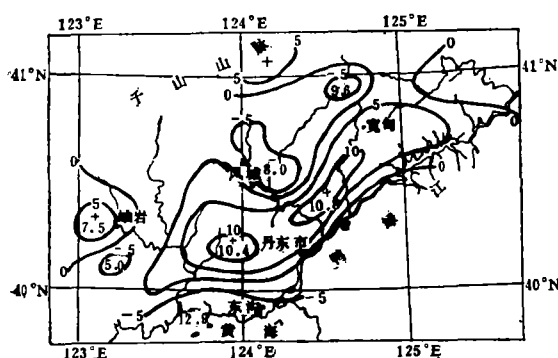


图 3 地理条件对丹东大暴雨日数的影响
(单位: d)

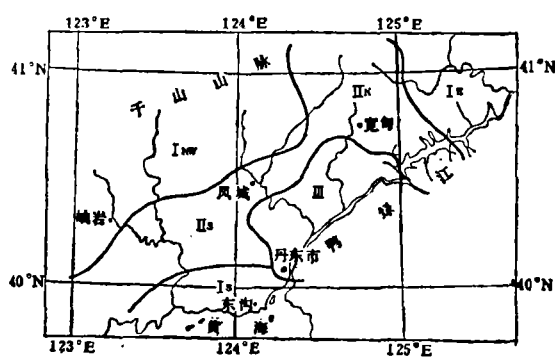


图 4 丹东大暴雨区划

数。本文直接引用了局地因子影响大暴雨日数的趋势面分析(TSA)的结果¹⁾, 其意义是, 地形在各地引起大暴雨日数增加和减少的情况, 见图 3。

这三项区划指标中, x_1, x_2 是预报和防汛所关心的基本问题; x_3 是气象工作中所必须注意的地理条件影响。

2. 区划所用的方法及原理

我们采用动态模糊聚类(ISODATA 算法)²⁾来进行区划, 其原理是: 分类结果使泛函(目标函数) $J_m(u, v)$ 达到最小。

目标函数的计算

$$J_m = \sum_{i=1}^n \sum_{j=1}^k (u_{i,j})^m \|x_i - v_j\|^2 \quad (1)$$

其中 $u_{i,j}$ 权重数值, x_i 因子值, v_j 第 j 类中心值, $\|$ 范数, 这里为绝对距离。 n 为样本数, k 为分类数, m 为待选定参数。 Bezdek 证明了当 $m > 1$, $x_i \neq v_j$ 时, 可以对 $u_{i,j}$ 和 v_j 进行迭代运算, 并且运算过程是收敛的^[1]。

类中心值的计算

$$v_j = \frac{\sum_{i=1}^n (u_{i,j})^m \cdot x_i}{\sum_{i=1}^n (u_{i,j})^m} \quad (2)$$

权重数的计算

$$u_{i,j} = \left[\sum_{l=1}^k \left(\frac{\|x_i - v_l\|}{\|x_i - v_j\|} \right)^{2/(m-1)} \right]^{-1} \quad (3)$$

$l=1, 2, \dots, s$

v_l 类中心值

1) 丹东气象科学研究所, 谢世俊、丛安佐、张心有, 用趋势面分析(TSA)计算局地因子对暴雨分布的影响, 丹东暴雨成因及规律研究报告 18, 复印本, 1983 年。

2) ISODATA 是 Iterative Self-Organizing Data Analysis 的缩写。

l 为指标数。本文指标为 3 个, $s=3$ 。以简单形式计算, 取 $m=2, k=3$, 限定误差值 $e=0.01$, 迭代到误差小于 0.01 时, 即达到分类目的。

3. 计 算 步 骤

1) 给出区划指标的原始数据矩阵 X

$$X = \{x_{i,j}\} = \begin{Bmatrix} x_{1,j} \\ x_{2,j} \\ x_{3,j} \end{Bmatrix} \\ (j=1, 2, \dots, 37)$$

j 为地点号。

2) 给定初始分类矩阵 u_0 , 使满足

- (1) 每个元素的权重 $u_{i,j}$ 都在 $[0, 1]$ 间取值;
- (2) 每列元数之和为 1;
- (3) 每行元数之和大于 0。

如:

$$u_0 = \begin{Bmatrix} 0.9 & 0.8 & 0.7 & 0.1 & 0 & \dots \\ 0.1 & 0.1 & 0.3 & 0.7 & 0.7 & \dots \\ 0 & 0.1 & 0 & 0.2 & 0.3 & \dots \end{Bmatrix}$$

3) 按公式(2)计算类中心值 $v_{i,k}$ 。因 $i=3, k=3$, 故 $v_{i,k}$ 值是 9 个。

4) 按公式(3)计算权重数值 $u'_{i,j}$, t 为迭代序号, 第一次计算 $t=1$ 。

通过计算得到矩阵:

$$u^t = \{u'_{i,j}\} \quad i=1, 2, 3, \quad j=1, 2, \dots, 37$$

判断

$$\alpha = \max |u'_{i,j} - u_{i,j}^{t-1}| < e = 0.01?$$

如果 $\alpha > e$, 即未达到精度要求, 则以 u^t 为初始矩阵, 重复迭代, 直到达精度要求为止。

5) 达精度要求后, 便将 u^t 矩阵中每一列 $u_{i,j}$ 的最大者变为 1, 其余变为 0, 即获得分类矩阵, 形如:

$$(u^t)' = \begin{Bmatrix} 1 & 1 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 1 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \end{Bmatrix}$$

4. 区 划 结 果

经过九次迭代($t=9$), 达到精度要求, 误差 $e=0.009$, 得出分类矩阵。区划结果如图 4, 共分为三类六区。

由于丹东地区大暴雨次数和雨量都多于北方各地, 因而对三类大暴雨区定名为: I, 大暴雨较频繁严重区; II, 大暴雨频繁严重区; III, 大暴雨最频繁严重区。各区情况如下:

III 区: 为鸭绿江中下游右岸河谷浅山区, 包括宽甸县南部和丹东市郊一部分。这里是东北的暴雨中心, 年雨量在 1200mm 以上。参加区划的 25 年中, 丹东地区发生 500mm

以上特大暴雨过程 7 次, 有 5 次发生在 III 区。本区暖河、蒲石河汇入鸭绿江处, 是历年防洪的重点。历史上几场特大的台风暴雨, 中心都在本区的黑沟。

II 区, 年雨量在 1000—1200mm 之间, 是气旋暴雨最多之处, 大暴雨也很频繁, 25 年中发生过两次 500mm 以上的特大暴雨过程。根据本区地理特点, 分为两个亚类区, II_s 为岫凤南部丘陵区, 包括岫岩县和凤城县南部和东沟县西北部, 位于大洋河中下游, 常有洪涝及内涝发生。II_N 为凤宽北部山区, 包括凤城县东北部及宽甸县北部。本区几乎每年都发生“啸山”(泥石流)事件, 冲毁铁路、公路和农田

I 区, 年雨量在 800—1000mm 之间, 多于北方广大地区。大暴雨次数较 II、III 区少, 而多于丹东以外的地区。分为三亚类区: I_s 为南部沿海平原区, 包括东沟县的大部, 副高西部切变线引起的暴雨, 多分布于本区。当暴雨与风暴潮同时发生时, 在鸭绿江口和大洋河口便遭受严重损失。I_{NW} 为岫凤西北部山区, 包括岫岩县、凤城县西北部, 是靠近辽东半岛分水岭的山脊地区。本区暴雨多为冷涡和偏北冷锋引起的局部性暴雨, 飑线等强对流天气多由本区进入。I_e 为宽甸东部山区, 处于鸭绿江支流浑江南侧, 属深山区, 也是易于发生“啸山”和洪涝的地区。

以上区分, 与防汛实践经验基本相符, 而且更明确地判明了工作重点。从地理环境看, 丹东处在辽东半岛向东南的斜坡上, 有利于夏季东南气流的抬升。雨量线、暴雨次数等值线等, 略与地形等高线平行。把丹东地形划为三级阶梯, 与区划结果的界线基本一致。

5. 小 结

此方法的优点在于能根据分类要求的数目和精度, 获得明确的分类结果。但分类数及精度含有经验性。对此, 我们先用点聚图方法, 对 x_1, x_2 作了分析, 发现点子聚为三个群落, 这样取 $k=3$ 使其不失客观性。此法如用手算, 颇感烦琐, 但用微型计算机, 程序并不复杂, 很快能打印出结果。

参 考 文 献

[1] 楼世博、陈化成, Fuzzy 聚类与天气预报, 模糊数学, 1981 年创刊号。

ON THE REGIONALIZATION OF HEAVY RAIN AREAS IN DANDONG PREFECTURE

Xie Shijun Cong Anzuo Zhang Xinyou

(Dandong Meteorological Research Institute, Liaoning Province)