

几种气象台站网设计方法简介

气象台站网的布局是否合理直接影响着天气预报、气候研究和有关工程计划的制订,各国对台站网的设计都非常重视。自苏联把德罗兹多夫-塞普列夫斯基的结构函数作为台站网设计的一种依据以来,信息理论、因子分析、经验模式等新的方法已开始应用于气象台站网设计,目的是寻求适合于不同地区、不同观测要求的站网设计。本文将简略介绍沙特阿拉伯、英国、我国台湾的站网设计方法。

1. 信息论法

气象台站网的规划是一项涉及到时、空范围内多变量现象的研究课题。为了提高现有气象台站网的效益,研究现有台站所有可能的多变量组合是十分必要的。在一个拥有大量台站的研究区域里,考虑多变量所有可能的组合在计算上是不可能的。为了解决这种复杂问题,可以应用 Shannon 的信息理论。该理论使某一子区域内的一组台站所传送的有关另一子区域的信息减至最少。将某一测点的总信息量分解为该点所在子区域共有的内部信息和其它台站提供的该测点的信息。为了更好理解分区结果的意义,提高这一方法的空间分辨率,下面着重介绍三站为一组的分类法。

先考虑平均信息量与信息的关系。我们考虑这样的情况:在某气象站实际气象输入量“Z”具有可能的离散值 $z_i, i=1, 2, \dots,$

N , 它在气象站以变量“X”被测出, 其值为 $x_j, j=1, 2, \dots, N$ 。分别用 $P(z_i)$ 和 $P(x_j)$ 表示事件 z_i 和 x_j 的概率分布, 它们的联合概率分布用 $P(z_i, x_j)$ 表示。

用 $I(z_i, x_j)$ 表示 x_j 所传送的有关 z_i 的信息(以比特为单位), 它与 $P(z_i)$ 和 $P(x_j)$ 有关。Shannon(1949)将其定义为:

$$I(z_i, x_j) = \log_2 P(z_i/x_j) - \log_2 P(z_i) \quad (1)$$

全部输入/输出事件的组合所传送的平均信息是每对可能的输入/输出事件所传送的信息乘以它们出现的概率之和, 即:

$$I(Z; X) = \sum_i \sum_j P(z_i, x_j) \cdot [\log P(z_i/x_j) - \log P(z_i)] \quad (2)$$

其中 $P(z_i/x_j)$ 为对于给定的 x_j, z_i 的条件概率, 它等于 $P(z_i, x_j)/P(x_j)$ 。由此可得

$$\begin{aligned} I(Z; X) = & - \sum_i P(z_i) \cdot \log P(z_i) \\ & - \sum_j P(x_j) \cdot \log P(x_j) \\ & + \sum_i \sum_j P(z_i, x_j) \cdot \log P(z_i, x_j) \end{aligned} \quad (3)$$

根据平均信息量的定义, 有

$$H(X) = - \sum_j P(x_j) \cdot \log P(x_j)$$

$$H(Z) = - \sum_i P(z_i) \cdot \log P(z_i)$$

及

$$H(Z, X) = - \sum_i \sum_j P(z_i, x_j) \times \log P(z_i, x_j)$$

可以将整个区域分成若干个子区域,然后计算诸子区中各种可能组合的交互信息量,以便评价每个子区内的台站网密度,但要计算所有这些组合是不可能的。因此,有必要探索简化的、三站为一组的分区方法。

三站为一组的分区法是很简单的,对于地区性气象研究来说在计算上是可行的。图1表示三站一组所构成的三角空间单元,假设三个站的待测变量为 X_1 、 X_2 和 X_3 。经过推导,组间的信息可简化为:

$$\begin{aligned} I(X_1; X_2; X_3) &= H(X_1) + H(X_2) \\ &\quad - H(X_1, X_2, X_3) \end{aligned} \quad (4)$$

但是,由 X_3 传输的 X_2 信息为:

$$I(X_2; X_3) = H(X_2) + H(X_3) - H(X_2, X_3) \quad (5)$$

因此,内部信息的增加乃是 X_1 与 X_2 和 X_3 联合的结果,即

$$\begin{aligned}\Delta Int(X_1) &= I(X_1; X_2; X_3) - I(X_2; X_3) \\ &= H(X_1) + H(X_2, X_3) \\ &\quad - H(X_1, X_2, X_3) \\ &= I(X_1; X_2, X_3) \quad (6)\end{aligned}$$

我们知道,由一站网传输的信息等于该站网
内所有站观测到的变量的联合平均信息量

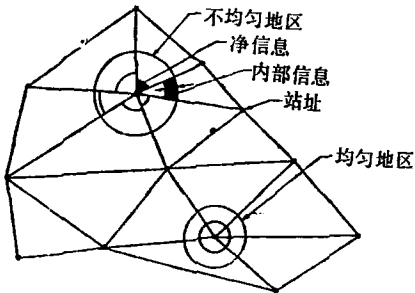


图 1 三站为一组的分区图

(Husain, 1979)。所以,由 (X_1, X_2, X_3) 输送的信息为 $H(X_1, X_2, X_3)$,而在 X_1 与 (X_2, X_3) 联合之前输送的信息为 $H(X_2, X_3)$ 。由于“ X_1 ”与 (X_2, X_3) 联合而增加的净信息量用 $\Delta Net(X_1)$ 表示,即为:

$$\Delta Net(X_1) = H(X_1, X_2, X_3) - H(X_2, X_3) \quad (7)$$

由(6), (7)两式可得:

$$\Delta Int(X_1) = H(X_1) - \Delta Net(X_1)$$

或 $H(X_1) = \Delta Int(X_1) + \Delta Net(X_1)$ (8)

这里, $H(X_1)$ 是 X_1 的总信息量; $\Delta Int(X_1)$ 是由 X_1 、 X_2 和 X_3 构成的三角区中 X_1 的内部信息。

(8)式表明,一个站提供的总信息量等于它的资料平均信息量,这一总信息量包括两部分: a) 本小组内部信息量的贡献, b) 对该小组网点提供的信息增量的贡献。在一个具有一定密度及空间分布的站网里,可以预料所有站都会对其邻近小组提供 $\Delta N_{et}(X_1)$ 的相似数值(如图 1)。这并不意味着所形成的各三角单元应该具有规则的形状或相等的面积。这些三角单元形状与面积的差异应该反映出被监视的气象场的空间不均匀性。

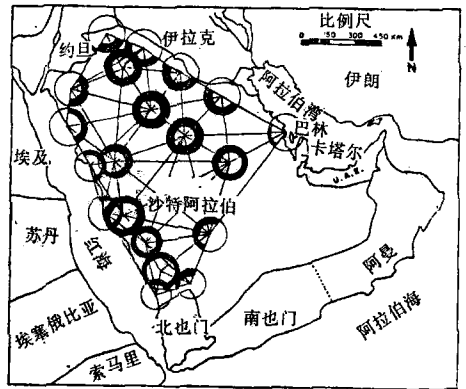


图 2 信息组成的示意图(平均相对湿度)

将整个研究区域分为若干个可能的三角区(图2)。利用一致的观测资料和(6)、(7)式计算出每个站的平均信息量,三站一组形成的子区的平均信息量,以及每三站为一组的净信息量和内部信息量。为了计算公式中单变量的平均信息量和联合平均信息量,把经过处理的每个变量的一致性观测资料,从最小值到最大值分为8个档次,并计算单变量和多变量个例的相对频率。经过试验,发

现超过 8 个档次净信息量和内部信息量的变化并不明显。另一方面,三站一组的组合量大,超过 8 个档次时计算平均信息量的时间就会大大增加。因此,最佳划分间隔确定为 8 个。图 2 为沙特阿拉伯区域平均相对湿度的平均信息量和净信息量组成示意图,每个三角区域中的每个站的信息分量用各站点上的两个同心弧表示,外弧代表总信息量,内弧代表三角区域的内部信息量。

从图 2 可以看出该国平均相对湿度站网分布,东北部站净信息值特别低,而内部信息值相当高。这表明该地区站网较密,所以站与站间有较强的相关。内部信息值高,还表明该子区域内测站传输多余的信息。在达兰(Dhahran)净信息值高表明达兰站和该区其它站之间交互作用差。西海岸的阿-焦夫、延布、吉达、基赞及哈米斯木夏特等站具有非常高的净信息值,表明它们与西海岸其它站的交互作用很弱。这些站的内部信息也很低,从相对湿度考虑,应在这个地区增加一些气象站。

2. 因子分析法

此法将对试图确定气候均匀区的可能性进行探讨,并提出一种方法,据此可确定某区域在气候上复杂与否,进而对台站的最佳数量作出估计。

气候区域的划分基本有三种方法:1)根据人为的主观方法,它假定区域上气候均匀。2)根据农田的利用来划定农业气候区。3)根据基本的地形特点来划分区域,例如,在地图上沿海平原、内陆山区、内陆凹地、沼泽地、高原以及大城市的周围画上线条。

为站网设计需要而划分气候区的上述三种方法,其可行性也许相同,也许不同,但它们都留下一个难以解决的问题,即如何在更复杂的地区决定台站的数量。

以下介绍确定气候均匀区的因子分析。Spackman 和 Singlebon(1982)已描述过因子分析在区域质量控制上的应用。区域质量

控制过程的基础是以下式来表示某天 i 某站 j 的气候要素 X :

$$X_{ij} = a_{i1}f_{1j} + a_{i2}f_{2j} + \cdots + a_{in}f_{nj} + r_{ij}$$

其中 $f_{1j} \cdots f_{nj}$ 是 j 站的因子, $a_{i1} \cdots a_{in}$ 是 i 日的因子权重, r_{ij} 是这个特定日的误差或残差。已经发现, X_{ij} 总方差的 85% 可用前 15 个因子 ($f_{1j} \cdots f_{15j}$) 及其相应的权重 ($a_{i1} \cdots a_{in}$) 来描述。

这些因子由各站的物理特征(或物理特征的组合)所决定,而因子权重则取决于天气形势,因此在西风带形势下,考虑温度时,纬度因子显得很重要,其相应的权重亦很高。反之在反气旋形势下,当考虑最低温度时,代表夜间辐射因子的权重将很高。那些在相似天气形势下受到相似影响的台站将具有相似的因子。因此,通过选择具有相似因子值的台站归组的方法来确定气候均匀区的努力似乎是合乎逻辑的。这种选择或分组可用很多方法来实现,这里选择了“聚类法”。

聚类过程是通过确定 15 维因子空间的台站间距来进行的,即

$$d_{ij}^2 = \sum_{k=1}^{15} (f_{ik} - f_{jk})^2$$

其中 d_{ij} 是台站 i 和 j 之间的“距离”, f_{ik} 和 f_{jk} 是这两个站上的第 k 次因子。这种聚类法先产生一定数量的组数,同组内台站之间在因子空间上比其他组的台站更为接近。

然后,用主观或客观方法可将归为一组的台站用线条围起来。图 3 表示对最低温度聚类成 40 组的分析结果(图 3 原分 a、b、c d 4 图,分别为最低温度、最高温度、日照和日降水量的聚类分组结果,这里仅给出 3 a)。在进行聚类分析时,对气温和降水使用了 700 个以上的台站,但对日照仅使用 460 个左右的台站。所使用的台站是气候站而非纯粹的雨量站。其中三张分析图上画有 40 个以上的区域,因为有些站可能在因子空间上较为接近而在地理上相隔甚远。例如,当边界处在风口或霜谷就会发生这种情况。对于某些

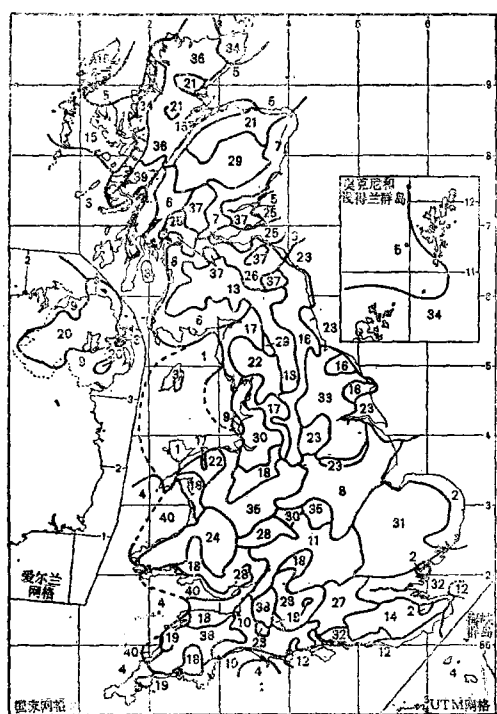


图 3 采用聚类分析法将最低气温划为 40 组的气候均匀区, 编号一样的区域属于同组

特定的要素而言, 这样的地点可能在气候上很相似, 尽管彼此相距甚远并分属不同的气候区。对不同要素进行分析的复杂程度是不同的, 这一方面取决于具有本研究所需资料台站的数量, 另一方面取决于地形对所考虑要素的影响。从各要素聚类分组图(图 3 只给出最低温度的聚类分组结果)上可以看出, 虽然有若干相当大的区域对于一两个要素来说是气候均匀的, 但毕竟还存在着其他结构复杂的地区, 在这些地区, 地形起着很重要的作用。图 3 还显示出对于一个要素来说是结构简单的地区, 对另一个要素就可能显得结构复杂了。因此, 对于一个要素来说气候均匀的地区, 对于另一个要素就不一定是气候均匀的。

所以, 试图应用因子分析法来划定气候均匀区是行不通的, 主要是因为这种方法不可能对全部要素确定唯一的一组区域。然而应当指出, 对单个要素而言, 图 3 中每个区域的站网要求可望在每 1000 km² 设 1 站的基

础上作出规定。有时需适当增加密度, 以便使每个区域不论如何小至少也有 1 个站; 面积在 1000 km² 至 2000 km² 之间的区域至少有 2 个站; 面积在 2000 km² 至 3000 km² 之间的区域至少有 3 个站, 如此等等。当然, 首先要假定有足够的资料去确定这些区域。从英国某些地区的资料覆盖情况来看显然是不符合这一要求的。因此, 根据单一要素来确定站网要求的分析仅适用于拥有足够数量的台站来进行聚类分组的地区。

3. 经验公式估计法

气象观测网中最复杂的问题是如何确定雨量站网的密度。由于降水不仅有量的差异, 而且降水形式有很大差异, 因而, 必须考虑如何减少因降水形式不同而引起的误差, 雨量站网的规划密度是各界关心的问题。以站网推求暴雨面积雨量尚有误差, 涉及点降水的站网密度更难以考虑。

在同一精度要求下, 测站控制面积愈大, 所得的标准误差愈大, 如要求推估的降水精度更高, 则必须设置较密的站网。此外, 雨量站的设置, 应参考雨量的分布情形(尤其是暴雨中心), 故山区雨量站的密度应较平地为密(台湾正好相反)。从水文预报角度看, 面积雨量比点雨量重要, 因此, 一地区雨量站网的数目要能精确求得区域降水分布。虽然决定一适当站网密度必须考虑的因素颇多, 诸如气候、地形、区域发展情况、用途、预算以及观测人员等, 但是真正能符合此要求者不多。因而, WMO 订有最佳站网及最少站网的设置标准表(表 1), 可供参考。

表 1 降水站网的最低密度

地 区	一般标准 (每站面积, km ²)	困难情形 (每站面积, km ²)
气候温和的平原、内陆及 热带地区	600~900	900~30,000
上述的山区	100~250	250~2,000
雨量不均匀的山丘、小岛	25	
干燥区及极区	1,500~10,000	

ks 曾建议,台湾雨量站网的设计可依下式加以估算:

$$N_p = \frac{K' \times A}{1000} \times P_d^{0.615}$$

式中 N_p : 雨量站数; A : 陆地面积; P_d : 每平方公里人口数; K' : 常数, 与资源开发利用情况有关, 对台湾省可用 $K'=0.645$ 。

台湾省约 2/3 的面积是山区及盆地, 全部河川都很短, 大小河川有 152 条, 地形的效应加上降雨的不均, 常造成洪汛。其余 1/3 的平原地区面积约 72000 km², 大约每 1000 km² 可设 6~25 个站 (实际采用 15 站/1000 km²)。Bocks 曾依上式预测到 1984 年应为 997 站, 1988 年应为 1059 站。依此计算, 目前台湾省的测站增长数还是合理的。

对于有些高山峻岭地区, 设站不易, 可以利用间接的遥测系统, 通过雷达和卫星图片, 来间接获取雨量资料 (借助于温度、亮度估计降水量)。Smith 及 Dixon (1976) 曾有如下的设计构想: 即利用降水的回波特性与雨量计可测出雨量出现的频率, 设计一种“三角形”的雨量计密度空间 (图 4), 计算公式为:

$$a = r \sqrt{\pi / (n \sin 60^\circ)}$$

式中: r 为圆圈半径, n 为在圈内设有雨量站的数目, a 为两邻近雨量计的距离。

上式的依据是以雷达站较大的探测范围去配合雨量站网的设计。由图 4 可见, 在圆

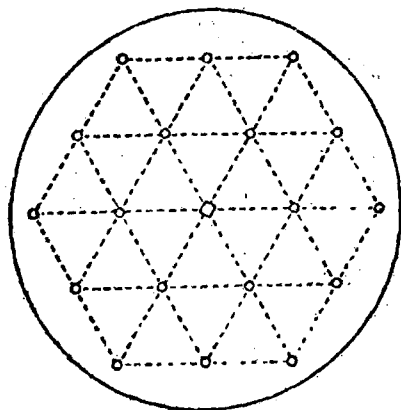


图 4 模拟的测站网, □表示雷达站, ○表示雨量站

圈内有 18 个雨量站, 故其两站的距离为 $0.45 r$, 这表示两相邻测站的空间距离大小是依雷达的探测范围而定的, 若可测范围大, 则密度低, 反之密度较高。

张强、杨贤为、何素兰、耿全震编译

参 考 文 献

- [1] T. Husain, M. A. Ukayli and H. U. Khan, 1984: Evaluation of Meteorological Network of Saudi Arabia Using Information Theory, J. of the Meteorological Society of Japan, Vol. 62, No. 5.
- [2] F. Singlebon, E. A. Spackman, 1984: Climatological Network Design, The Meteorological Magazine, Vol. 113, No. 1341.
- [3] 刘复诚, 从水文预报观点探讨雨量资料之搜集与解析, 台湾气象学报, 第二十九卷, 第三期, 1983年。