

尺度分布的 Getis 统计对遥感图像特征参量 空间自相关性的研究

颜锋华 金亚秋

(复旦大学波散射与遥感信息国家教育部重点实验室, 上海 200433)

摘要 分析了 Getis 统计中空间尺度设定对于标准 Z 值的影响, 提出用变化的空间尺度设定来进行 Getis 空间相关性的计算, 直观反映遥感特征参量的空间分布, 并用被动微波遥感 SSM/I (special sensor microwave/imager) 的辐射亮度温度数据进行了验证分析。进一步地, 针对 Getis 标准 Z 值分布不能评估中值聚类的缺点, 提出了多层尺度分布的 Getis 空间自相关性统计方法, 用模拟的图像和主动微波遥感 ERS-2 SAR (European Remote Sensing-2 Synthetic Aperture Radar) 数据进行了验证分析, 结果证明多层尺度分布的 Getis 能够比较全面地反映地表特征参量分布的空间信息。

关键词 空间自相关性 Getis 统计 多层与多尺度分布

中图分类号: TP751.1 P407.8 O212.1 **文献标识码**: A **文章编号**: 1006-8961(2006)02-0191-06

Spatial Auto-correlation of Characteristic Parameters in Remotely Sensing Image Using Optimal Multi-Scale Getis Statistic

YAN Feng-hua, JIN Ya-qiu

(Key Laboratory of Wave Scattering and Remote Sensing Information (Ministry of Education) Fudan University, Shanghai 200433)

Abstract The effect of varying spatial scale of the Getis statistic on the standardized Z value is studied. It is proposed to employ different spatial scales for evaluation of the spatial autocorrelation in the Getis statistic. Multi-scale Getis is analyzed and validated by using passive remote sensing SSM/I data. Furthermore, a novel multi-layer and multi-scale Getis is studied to evaluate the spatial autocorrelation of the pixels with the medium values. An image of active remote sensing ERS-2 SAR gives an example for analysis and validation of this novel multi-layer and multi-scale Getis.

Keywords spatial autocorrelation, Getis statistic, multi-layer and multi-scale

1 引言

空间遥感与对地观测中, 地表特征参量的空间分布与变化, 地表形态(指建筑、草地等)的空间相关和自相关性以复杂的方式反映在测量的遥感数据上。对于形态相同或相似的区域, 测量到的遥感数据应具有相似性。对于某一区域测量获得的遥感数据图像中, 各像素并不是完全独立的, 而是在空间结

构上存在与相邻位置的相关性。对于同一通道的数据图像, 这种相关性主要表现为相邻点数据值大小的相似, 表征了地表形态的空间结构相似相关的特点。地表特征参数空间自相关性在时间尺度上的变化也表征了它们的空间分布所发生的变化。由于空间自相关性是建立在相邻数据的比较之上, 能够比较好地减小由于其他因素变化对遥感观测数据带来的整体影响。空间自相关性统计对测量的偶然性具有很强的抗干扰能力, 成为研究地表特征分布的一

基金项目: 国家重点基础研究发展计划项目(2001CB309401); 国家自然科学基金项目(60571050)

收稿日期: 2005-01-14; 改回日期: 2005-06-19

第一作者简介: 颜锋华(1978 -), 男, 软件工程师。2001年与2004年先后于复旦大学获信息与工程学士学位与硕士学位, 研究方向为电磁场与微波技术。已在国内外发表论文6篇。

联系人: 金亚秋。E-mail: yqjin@fudan.ac.cn

种重要的遥感理论和方法。

几十年来,对空间相关性和自相关性已开展了许多工作,其中常见的方法有 Moran'I 和 Geary 统计^[1]。但是,这两种方法一般只对区域的像素点相关性做出单一的评估,不能够评估整个区域内空间相关性的变化程度。Getis 方法是一种局部空间自相关性方法,通过设定一定的空间尺度大小,可以来计算极值变量的局部聚集程度^[2,3]。但是,由于需要预先设定空间尺度,Getis 方法不能准确地评估观测区域中实际存在的不同空间尺度的自相关性。同时,虽然 Getis 统计能够评估高值和低值的聚类和分布,但是缺乏对中值点聚类的敏感。

采用不同的空间尺度设定计算 Getis 自相关性的大小与区别,提出了直观反映变量空间分布特点的尺度分布,用被动微波遥感 SSM/I (special sensor microwave/inager) 的辐射亮度温度遥感数据的分析验证,证明了尺度分布能够比较好地适用于不同空间范围的自相关性大小的计算。同时,针对 Getis 标准 Z 值分布不能评估中值聚类的缺点,提出了多层尺度分布的 Getis 空间自相关性统计,用模拟的图像和 ERS-2 主动遥感 SAR 的数据,证明了多层尺度分布的 Getis 统计能够比较全面地反映地表特征参量在不同大小尺度上的空间自相关性的分布信息。

2 Getis 自相关性统计

Getis 统计用来研究局部位置上变量 x 值的范围和集中程度。考察一张 n 个像素点的图像,以第 i 个像素点为中心,距离该点不超过 d 的区域中的变量(含第 i 个点)和与整张图中 n 个像素点变量之和的比值定义为

$$G_i^*(d) = \frac{\sum_{j=1}^n W_{ij}(d)x_j}{\sum_{j=1}^n x_j} \quad (1)$$

其中, W_{ij} 为以第 i 点为中心的权重因子。以第 i 点为中心,权阵 W_{ij} 不为 0 的点组成了一个窗口,窗口的形状可以根据研究目的和图像特点定义。一般地,若窗口定义为正方形,则当第 i 点和第 j 点相距不超过 d 时取 1,其余为 0^[2,3]。

计算归一化的 Z 值为

$$Z_i(d) = \frac{[G_i^*(d) - E(G_i^*(d))]}{\sqrt{Var(G_i^*(d))}} \quad (2)$$

其中, $E()$ 为期望值, $Var()$ 为方差。式(2)描述图像区域中变量 x 的集中程度。由式(2)计算得到的正 Z

值表示高值 x 的集中,负 Z 值表示低值 x 的集中。

图 1(a)是 2001 年 3~8 月间在中国巴丹吉林沙漠与腾格里沙漠及其附近地区 ($E96^\circ \sim 106^\circ$, $N36^\circ \sim 43^\circ$) 的被动微波遥感 SSM/I 卫星 F13 得到平均沙漠化指数 $DI (= T_{B37v} - T_{B37h}, 37GHz$ 垂直与水平极化辐射亮温之差) 分布^[4] (图中采用了双三次插值)。DI 作为沙漠化判据的一个指数。

图 1(b)~(d)分别是上述地区当 $d=1$ (即 3×3 窗口)、 $d=2$ (5×5 窗口)和 $d=3$ (7×7 窗口)对应的标准 Z 值分布。暗黑色表示负相关性较强,是低值点的聚集区,而白色则表示高值的聚集,是沙漠所在的地区。从图 1(b)~(d)可以知道,标准 Z 值的计算需要预先设定正方形窗口的大小 d 。计算结果表明了两块沙漠之间某些点值虽然不是最高,但是由于周围是高值聚集而使 Getis Z 值比较大。窗口越大,则计算了更大范围内的相关性,使得相邻点的 Z 值更接近,高值和低值聚集区的边界更圆滑。

3 空间自相关性的尺度分布

上述分析表明,Getis 的标准 Z 值分布需要设定一个空间尺度 d ,再进行空间统计分析。空间尺度的预定一般用 3×3 或 5×5 的窗口。

假定一个矩阵分布为

m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
m	A	A	A	A	A	m	m	B	B	B	B	B	B	m
m	A	A	A	A	A	m	m	B	B	B	B	B	B	m
m	A	A	A	A	A	m	m	B	B	B	B	B	B	m
m	A	A	A	A	A	m	m	B	B	B	B	B	B	m
m	A	A	A	A	A	m	m	B	B	B	B	B	B	m
m	m	m	m	m	m	m	m	m	m	m	m	m	m	m

这里一共有 $7 \times 14 = 98$ 个点,设 $A = m(1+c)$, $B = m(1-c)$, $0 < c < 1$,以 A 区中心点为考察点,则对于 $d=1$ 和 $d=2$ 的情况,可以计算 $Z_i(d)$ 如表 1 所示。

从表中可以看出, d 的扩大使得 $Z(G_i^*)$ 增大, $d=2$ 的 Z 值差不多是 $d=1$ 时的两倍。A 区中心点所处位置的空间尺度应该是 5×5 的窗口,而不是 3×3 的窗口。一般地,对于具有空间自相关性的点 i ,存在一个最佳空间尺度

$$d_{op} = \arg \max(Z_i(d)), |Z(d_{op})| > T_i \quad (3)$$

其中, T_i 为一控制参数,以确保在不同的空间尺度上具有不同的相关性阈值。不满足式(3)则假定 $d_{op} = 0$ 。若认定变量在范围 (\min, x_a) 和 (x_b, \max) 的

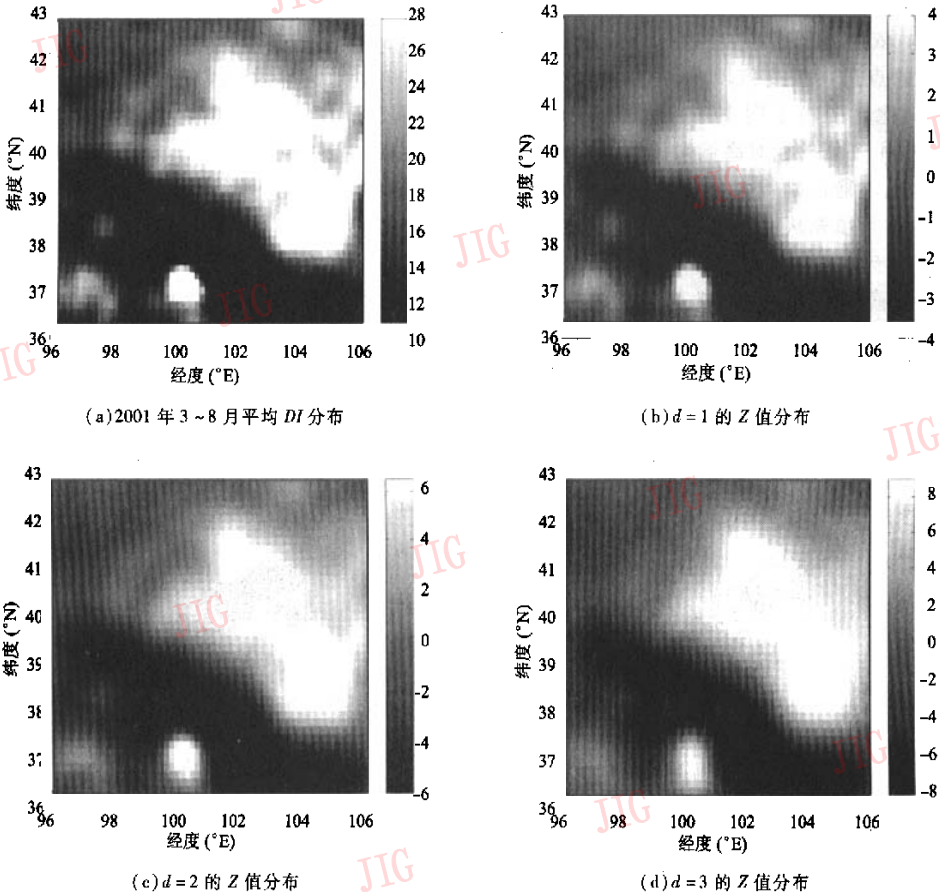


图 1 DI 与 Getis 分布

Fig. 1 The DI and Getis distributions

表 1 不同窗口标准 Z 值的计算

Tab.1 Calculation of the standard Z value at different windows

	$d = 1$	$d = 2$
G_i^*	$\frac{9(1+c)}{98}$	$\frac{25(1+c)}{98}$
$E(G_i^*)$	$\frac{9}{98}$	$\frac{25}{98}$
$Var(G_i^*)$	$4.39 \times 10^{-4} c^2$	$1.00 \times 10^{-3} c^2$
$Z(G_i^*)$	4.39	8.07

值区间聚集才具有相关性,则有系数

$$\alpha = \int_{\min}^{\tau_a} p(x) dx = \int_{\tau_b}^{\max} p(x) dx \quad (4)$$

这里的 $p(x)$ 为变量 x 值分布的概率密度函数, α 表示多大的像素值的聚合才被判定具有空间自相关

性。引入控制参数 α , 可以从图像像素值分布上直接决定空间自相关性限制条件的严格程度, 以决定 T_i 。因而 T_i 可以记为

$$\begin{aligned} T_{ia} &= Z(d_i) \Big|_{x=x_a} \\ T_{ib} &= Z(d_i) \Big|_{x=x_b} \end{aligned} \quad (5)$$

式(4)和式(5)可以用来决定阈值对的大小。对于 d_{op} 不为 0 的点, 表示在一定的限制条件下, 该点存在一个大小为最佳的空间尺度 d_{op} , 表示该点在该限制条件下与多远处的像素值存在相关性。

图版 1 图 1(a) 是符合正态分布的随机产生的模拟图像(均值 $\bar{x} = 0$, 方差 $\sigma = 10$), 大小为 100×100 像素, 其中分别嵌入了 3 大块数据, 其空间尺度分别为 $d = 1$ 、 $d = 2$ 和 $d = 3$, 数据块内像素值的大小是全图最高值和最低值。

图版 1 图 1(b) 是 $d = 1$ 时的 Getis 标准 Z 值分布。计算($\alpha = 0.1$)后得到的空间尺度分布如图版

I 图 1(c), 可以发现每一小块数据的中心部分都被分辨出来了。依据 d 的大小, 可以任意对中心点尺度 d 范围内的点进行尺度值填充, 得到图版 I 图 1(d) 的结果。比较图版 I 图 1(b) 与图 1(c)、(d) 知, 图 1(b) 中用 Z 值的相对高低来评估某点自相关性大小的方法, 不能区别像素值与尺度大小的分别贡献, 也无法进行填充从而获取最佳尺度信息。

为了进一步检验空间尺度分布的可用性, 采用了被动微波遥感 SSM/I 卫星 F13 的数据进行分析。

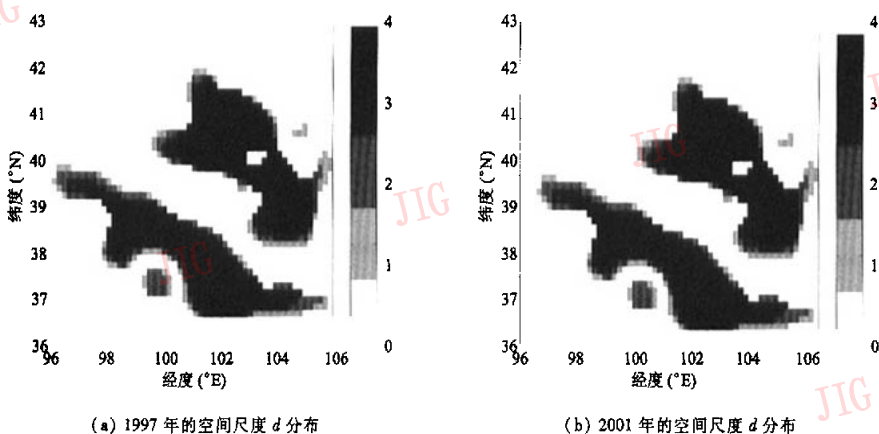


图 2 巴丹吉林沙漠与腾格里沙漠地区 1997 年和 2001 年的 Getis 空间尺度 d 的变化

Fig. 2 Distribution of spatial scale d in 1997 and 2001 at Badanjilin desert and Tenggelt desert

4 Getis 多层尺度分布

在图版 I 图 1 中采用了以 α 为限制的 Getis 空间自相关性的尺度分布。空间自相关性有两个要素, 一是空间尺度, 二是在空间尺度窗口内的值的大小。由于不同的尺度有不同的相关性阈值 T_i , Getis 方法往往对高值和低值有很高的敏感性, 但是对于平均值不在高值和低值范围的窗口, 上述简单的尺度分布方法并不能很好地加以辨别, 容易被误判为无空间相关性的区域。所以, 如何平衡空间尺度和窗口内均值对结果的影响是空间自相关性研究需要解决的一个重要问题。

假定在参数 α 的限制下已经计算了空间自相关性的 d_{op} 值分布, 规定为第 1 层; 对于余下的点再进行 d_{op} 值的计算, 这时式(1)调整为

$$G_i''(d) = \frac{\sum_{j=1}^n W_{ij}'(d) x_j}{\sum_{j=1}^{n'} x_j} \quad (6)$$

其中, n' 为余下的在上层判断为不具有空间相关性

与图 1 一样, 图 2(a)、(b) 分别对应 1997 年和 2001 年的 3~8 月间在中国巴丹吉林沙漠与腾格里沙漠及其附近地区平均沙漠化指数 DI 的空间尺度分布。空间尺度分布可以看出两大沙漠有会聚的趋势^[4]。与文献[4]中 Getis Z 值分布的分析相比, 空间尺度分布没有圆滑化的效果, 更在两沙漠接近的趋势上符合实际。所以, 空间尺度分布可以用来很好地分析区域地表不同时的变化趋势。

的像素点数目, $W_{ij}'(d)$ 是窗口内这些点的权值, 满足条件 $\sum_j W_{ij}'(d) > N$ 。 N 是在窗口内有相关性聚集时的最小像素数目。低于 N , 即使满足式(6)也是几个孤立的点, 不足以构成大的相关性。式(6)的结果纪录在第 2 层, 以此类推, 可以得到 l 层的空间尺度分布。经过 l 层计算, 若仍旧有 $d_{op} = 0$, 则判断该点没有明显的空间自相关性。

在计算完成各点的 d_{op} 之后, 需要在有相关性的点的周围进行填充, 使其具有自相关性 d_{op} 。虽然这些开始时 d_{op} 为 0 的被填充点与直接用式(3)计算得到的 d_{op} 的点具有不同的地位, 但是仍旧是空间尺度分布的一部分。此时的尺度分布不再完全表示是该点周围多少范围内存在 Z 的极值, 而是表示该点属于哪一个空间尺度的覆盖。同时, 必须及时地去除一些值在本层但不具有空间相关性的像素。

图版 I 图 2(a) 是符合正态分布的模拟图像(均值 $\bar{x} = 0$, 方差 $\sigma = 1$), 大小为 200×200 , 其中随机地嵌入了一些 $d = 1$ 、 $d = 2$ 和 $d = 3$ 的数据块。相同数

据块内的像素值是一样的,不同数据块内的像素值是随机设定的。图版 I 图 2(a)有些数据块中的像素值并不是在最大值或最小值附近,有的数据块甚至接近于平均值。图版 I 图 2(b)是 $d=1$ 时计算的 Getis 标准 Z 值分布,从中难以区分像素值不在高值和低值范围内的数据块。图版 I 图 2(c)是多层的空间尺度计算结果(10 层),与第 3 节不同,这里引入了负的空间尺度值, d 为负值表示负相关性。可以发现,图版 I 图 2(a)中的数据块基本上被辨别出来,并且赋予了相应的 d_{0i} 值(进行了填充)。图版 I 图 2(d)则是层号分布,可以清楚地看到有些数据块不是在第 1 层被辨别出来的。层号与层阈值 T_i 相对应,表示窗口内值的大小标度,也是空间信息的一部分,可以作为进一步处理的输入。

图版 II 图 3 是 2002 年 4 月 9 日的位于中国上海虹桥机场附近的 ERS-2 C 波段 SAR 遥感功率 ($10 \times \log(\text{功率})$) 图像,大小为 250×450 像素,图像的分辨率是 $12.5\text{m} \times 12.5\text{m}$ 。为了消除高值噪声,对数据的高值部分(20%)进行了“+”中值滤波和对数化。图像左下角为虹桥机场,图像的北部主要是居民区,有着相对密集的建筑,南部则有公园、动物园和网球中心等。

假定最大的自相关性为 7×7 的窗口(这个窗

口包含了大于此相关性的情况), $\alpha = 0.22$, 经过 20 层计算的空间尺度 d 分布如图版 II 图 4 所示,分布图代表了各处空间延续的大小。为了使得多层运算中相关性的阈值不断向平均值靠近,同时也为了减少像素值概率运算的误差,在进行每一新层的自相关性计算前,需进行直方图调整,使之符合正态分布。从图上可以看到,机场的自相关性为 -3 ,属于低值的聚集,空间尺度很大。北部区域空间尺度多为正,南部则 $d = -1$ 区域比较多。与实际地貌相比,几块公园类的区域是很符合的。

图版 II 图 4 和对应于图版 II 图 4 的层号分布可以作为获取图像中特征参量空间分布的重要信息输入。图版 II 图 4 中存在大量没有相关性的像素 ($d=0$),简单地,可以用空间聚类的方法将这类点归并到要进行的分类中,比如,进行正负相关性的空间尺度分类。对于 ERS-2 的图像而言,正值的空间尺度分布表明各类建筑比较集中,负值的空间尺度分布则表明草地、宽大公路或河流等后向散射系数比较小。

图 3 是利用最近归属的方法(即周围哪类多属哪类)得到的正负空间尺度分布图。可以比图版 II 图 4 更清晰地展示虹桥地区 2002 年地域特征性的空间分布。

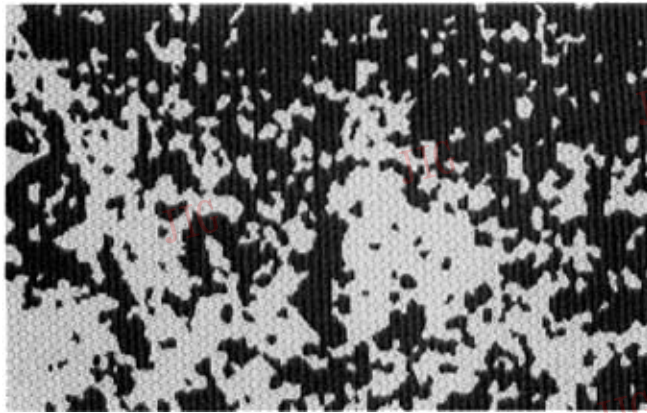


图 3 正负相关性的空间尺度分类结果(黑色表示 $d > 0$, 灰白色表示 $d < 0$)

Fig. 3 Classification of positive and negative scale (black $d > 0$, grey $d < 0$)

5 结 论

本文提出了单层和多层的空间尺度分布的

Getis 自相关性统计方法,用于对遥感图像特征参量相关与变化程度的评估。应用这种方法,对模拟的图像、SSM/I 数据和 ERS-2 SAR 图像的空间尺度进行了实例分析。结果表明,空间尺度分布能够比较

好地获得地区的空间变化相关性及其特征参数,特别是多层的空间尺度分布,可以辨别非极值的聚集。结合层号分布,本文提出的空间尺度分布作为特征参量空间分布结果获取方法的输入,可以获得符合实际的空间自相关性信息。

参考文献 (References)

- 1 Goodchild M. Spatial autocorrelation, Concepts and Techniques in Modern Geography[M], Norwich, UK: Geo Books, 1986, 47:3-6.
- 2 Wulder M, Boots B. Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic [J]. International Journal of Remote Sensing, 1998, 19 (11): 2223-2231.
- 3 Getis A, Ord J. The analysis of spatial association by use of distance statistics[J]. Geographical Analysis, 1992, 24(2): 189-206.
- 4 Jin Y Q, Yan F. Monitoring sandstorms and desertification in northern China using SSM/I data and Getis statistics [J]. International Journal of Remote sensing, 2004, 25 (11): 2053-2060.

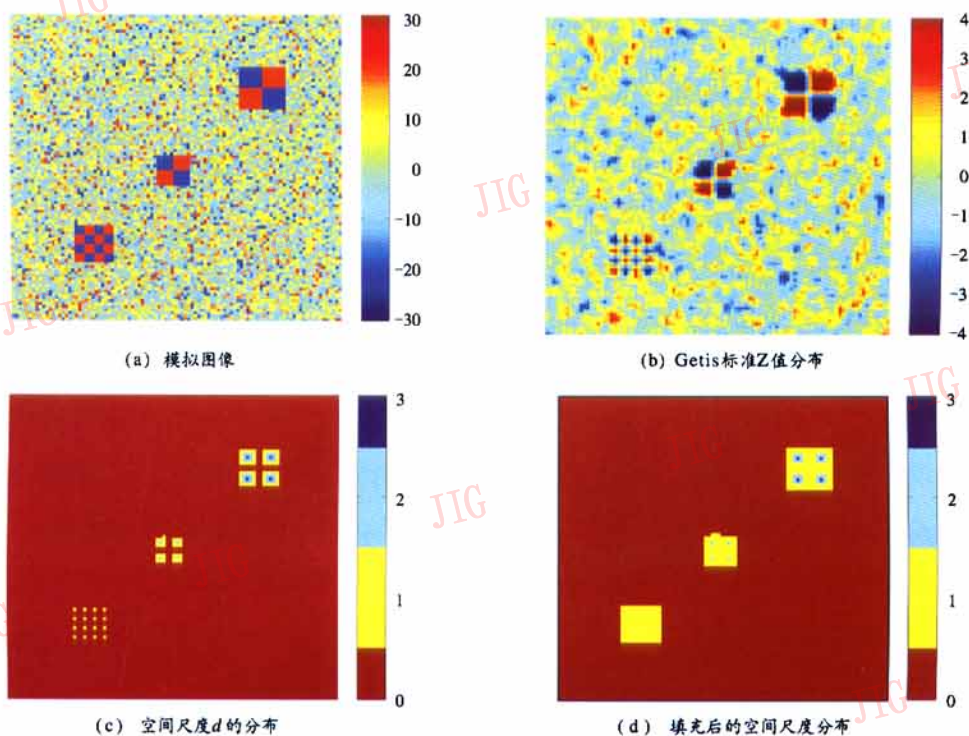


图1 模拟图像的空间尺度分布
Fig.1 Distribution of spatial scale for the simulated image

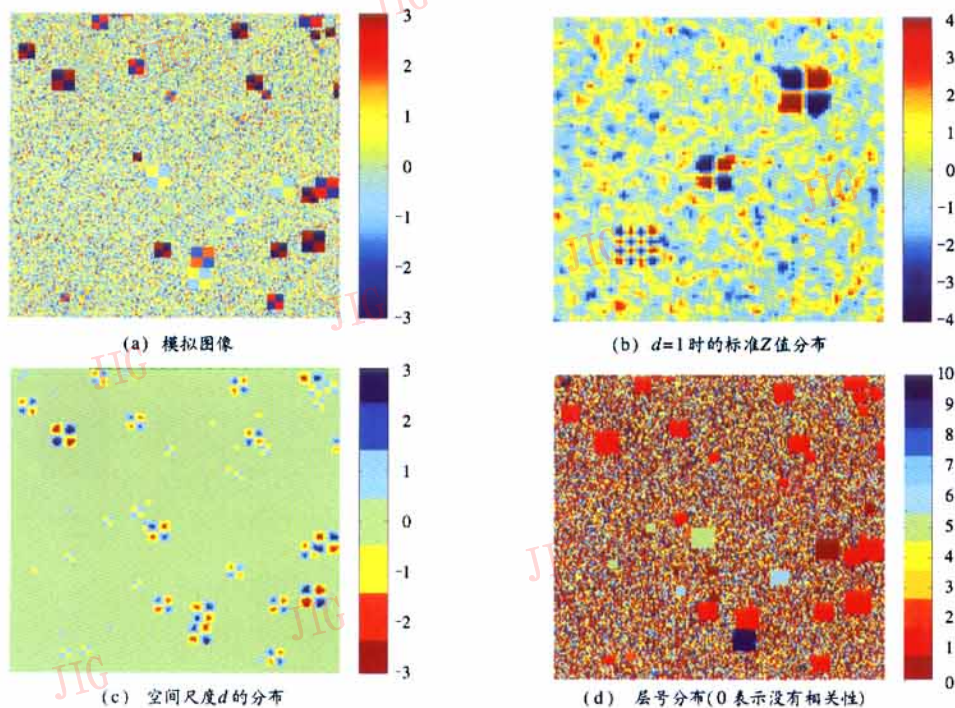


图2 模拟图像的多层空间尺度分布
Fig.2 Multi-layering spatial scale of the simulated image

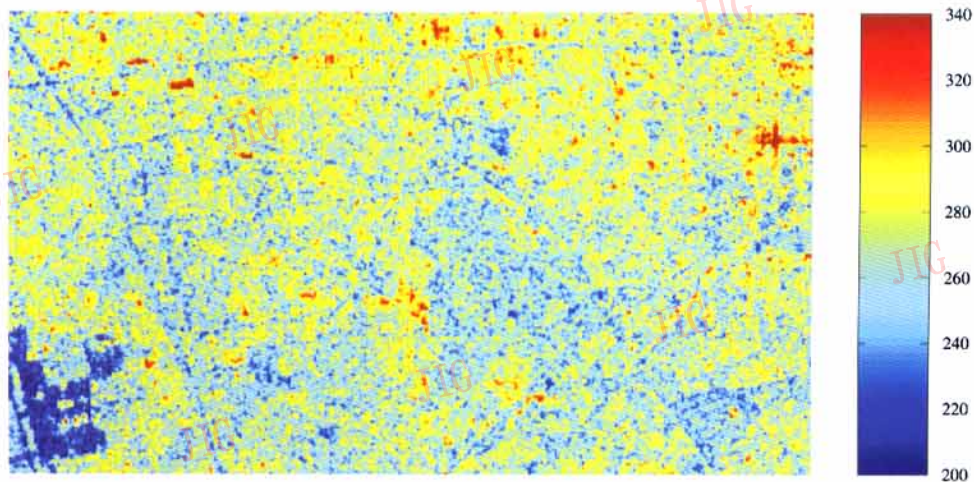


图3 2002年4月9日虹桥机场附近的ERS-2 SAR图像遥感功率($=10 \times \log(\text{功率})$)
Fig.3 EAR-2 SAR image over the Hongqiao airport in April 9, 2002($=10 \times \log(\text{total power})$)

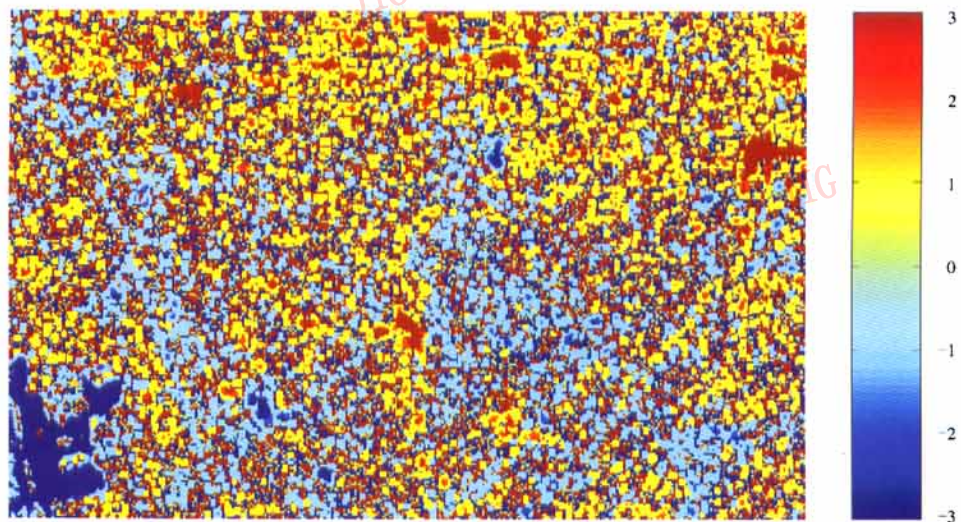


图4 多层空间尺度 d 的大小分布
Fig.4 Distribution of multi-layering spatial scale d