通过正则化方法进行空间泊松点过程的变量选择

Andrew L. Thurman, Jun Zhu

摘要

使用回归分析研究事件在空间中的发生与空间索引协变量之间的关系通常很有意义。泊松点过程是此类回归分析的一种模型。在本文中，我们开发了一种方法，通过正则化方法同时进行协变量的选择和模型参数的估计。我们通过模拟研究评估了我们方法的有限样本性能。此外，我们提出了一种我们方法的变体，允许在多个像素分辨率上进行协变量选择。为了说明，我们考虑巴罗科罗拉多岛上的一个树种Beilschmiedia pendula在中美洲巴拿马的研究区域。我们发现Beilschmiedia pendula在海拔较高、坡度较陡的地方出现频率更高。此外，我们识别出三个与Beilschmiedia pendula有关联的树种：两个它倾向于被吸引的树种，一个它似乎被排斥的树种，以及一个没有明显关系的树种。

关键词

自适应Lasso、强度函数、最大似然估计、模型选择、空间统计

1. 引言

空间点过程模拟空间中事件的随机位置，已被用作分析空间点模式数据的统计模型（例如，[14]）。空间点过程模型和方法在生态学、地理学、林业和流行病学等多种领域得到应用。这些研究的一个常见目标是将空间辅助信息与空间中感兴趣事件的发生联系起来。空间点过程与回归相结合非常适合实现这一目标，也是本文的重点。

例如，在巴罗科罗拉多岛（BCI）的一个长期生态监测项目中，自1980年代以来已经普查了近400,000棵单独的树木。特别是，已经识别、标记和绘制了胸高直径至少为10毫米的独立木本茎，形成了超过300个物种的树干图。使用回归分析空间点过程来分析这些数据，可以研究树种之间以及与环境因素相关的关系。此外，通过这里将开发的正则化方法，我们可以同时研究这些关系在不同空间尺度上的变化。我们将在第5节和第6节中说明这一点。

泊松点过程因回归分析目的而受欢迎，因为它们在理论上可追踪且在计算上易于实现。点过程的矩和似然函数可以解析推导。基于似然的方法已经开发用于拟合这些模型到数据并获得参数估计。似然的近似可以很容易地实现，适用于加权广义线性模型[2]。此外，这些模型参数的最大似然估计量的渐近性质，即一致性和渐近正态性，已经建立。例如，Rathbun和Cressie[16]采用了一个增长的空间域渐近框架，其中点过程的空间域允许增长，并使用局部渐近正态性理论建立了渐近结果[13]。

任何回归过程中的一个具有挑战性的组成部分是决定哪些协变量应该被选择用于模型。协变量的选择（即变量选择）在统计学中一直是一个活跃的研究领域，因为有效推断严重依赖于适当的模型规范，以减轻过拟合导致的方差膨胀和欠拟合导致的偏差。传统方法包括逐步程序（参见例如[5]）。从初始协变量集开始，逐步程序考虑在每一步添加或删除一个协变量，使用诸如F统计量之类的标准。在某些情况下，这样的程序可能会有偏差。例如，对应于F统计量的F检验是在假设两个候选模型是预先固定的情况下开发的，因此逐步程序与这一标准倾向于有偏差，因为候选模型是自适应选择的，因此不是固定的[9]。

最近，Tibshirani[17]通过惩罚（或正则化）方法引入了变量选择。特别是，最小绝对收缩和选择算子（Lasso）产生了回归系数的零和非零估计，同时执行选择和估计。这种正则化方法已被修改以实现改进的选择性能（例如，[7,22,24]）。大多数变量选择研究都集中在独立分布的数据上，但方法正在出现用于时间序列数据[18]、空间格点数据[21,10,20]和空间点过程数据[19]。

本文的目的是开发一种正则化方法用于泊松点过程与回归的变量选择。尽管变量选择有不同的可能方法，我们发现正则化方法是可取的。Oracle属性提供了渐近地选择正确协变量子集的保证，并且惩罚估计量具有与已知正确协变量子集相同的渐近方差，就好像正确的协变量子集是先验已知的一样。自适应Lasso是一种已被证明在线性回归

中具有这些oracle属性的方法，适用于独立数据[22]和依赖数据[18,20,3]。在这里，我们通过模拟提供经验证据，表明oracle属性也适用于带回归的泊松点过程。

本文的其余部分组织如下：第2节给出了泊松点过程模型的介绍。第3节描述了我们用于泊松点过程模型中变量选择和参数估计的方法。第4节展示了我们方法的模拟研究结果，随后在第5节中应用于BCI森林普查数据集。第6节提出了我们方法的一个变体，允许在多个像素分辨率上进行变量选择和参数估计。第7节给出一些结论性评论。

2. 泊松点过程

2.1 泊松点过程模型

令(Ω, A, P)表示概率空间，$D \subset \mathbb{R}^d$表示d维空间中的感兴趣空间域。在这里，我们考虑二维情况$d = 2$。令Y表示从(Ω, A, P)到X_D的映射，其中X_D表示局部有限配置的集合。局部有限配置是那些实现$y \subset D$，使得$y \cap A$对于每个有界Borel集$A \subset D$是有限的。换句话说，y是D中的有限空间坐标集。令$N(A) = N_Y(A) = N(Y \cap A)$表示Y在A中的事件随机数量。然后映射Y被称为D上的点过程[8]。

令$s_1, \ldots, s_n$表示观察到的空间点模式数据，包括n个事件在D中的位置。空间点模式数据的统计模型将事件的位置视为空间点过程Y在D上的实现。一阶矩测度，或强度测度，μ定义为

$$\mu(B) = E[N(B)] = E\left[\sum_{s \in Y} I(s \in B)\right],$$

其中B是D中的有界Borel集。通常假设存在强度函数λ(s)，使得μ(B)是λ相对于Lebesgue测度在集合B上的积分

$$\mu(B) = \int_B \lambda(s)ds.$$

空间点模式数据的基本统计模型是具有强度函数λ的泊松点过程，由以下两个条件定义：

1. 任何有界Borel集B在D中有$\mu(B) \in (0, \infty)$且$N(B) \sim \text{Poisson}(\mu(B))$。

2. 事件位置$s_1, \ldots, s_n$的联合密度f与N(B)条件独立：$f(s_1, \ldots, s_n|N(B) = n) = \prod_{i=1}^{n} \lambda(s_i)$。

条件1表示B中事件的数量遵循具有平均事件数μ(B)的泊松分布。条件2表示，在事件数量的条件下，事件位置是独立的，并且事件在小区域周围点s的概率与面积ds成正比，概率为λ(s)ds。因此，事件更可能出现在D中强度值λ较高的区域。见图1作为示例。对于这个模型，主要兴趣在于估计λ。

令x(s)表示位置s处$p \times 1$的协变量向量，其中$s \in D$。强度函数λ(s)可用于模拟事件位置与协变量之间的关系。一个常用的模型是对数线性规范

$$\lambda(s; \beta) = \exp[\beta_0 + x(s)^T \beta],$$

其中β_0是截距，β是$p \times 1$的回归系数向量，且$\beta = (\beta_0, \beta^T)^T$。

2.2 最大似然估计

从上面给出的泊松点过程定义，泊松点过程的对数似然函数可以推导如下。令$g(s_1, \ldots, s_n, n) = f(s_1, \ldots, s_n, N = n)$表示事件位置和事件数量的联合密度函数。令N = N(D)和P(N = n)表示事件数量的概率质量函数。同时，定义$\theta = \int_D \lambda(s; \beta)ds$。由此可得

$$g(s_1, \ldots, s_n, n) = f(s_1, \ldots, s_n|N = n)P(N = n) = [\prod_{i=1}^{n} \theta^{-1}\lambda(s_i; \beta)] \cdot [\exp(-\theta)\theta^n/n!].$$

在简化这个表达式后，β的对数似然函数，除了一个加性常数外，定义为

$$\ell(\beta) = \sum_{i=1}^{n} \log \lambda(s_i; \beta) - \int_D \lambda(s; \beta)\, ds$$

[8]。最大化ℓ(β)给出β的最大似然估计量(MLE)，记为β_{MLE}。

3. 变量选择和参数估计的正则化方法

3.1 惩罚最大似然

一般来说，正则化方法试图最小化形式为$-\ell(\beta)+n\rho_\gamma(\beta)$的目标函数，其中$\ell(\beta)$是某个感兴趣模型的对数似然函数，n是观测数量，$\rho_\gamma(\beta)$是由γ参数化的非负惩罚函数。Lasso惩罚是

$$\rho_\gamma(\beta) = \gamma \Sigma_{j=1}^{p}|\beta_j|。$$

在这里，这个惩罚函数在β = 0处不可微，并产生包含一些恰好为0的估计的解。因此，变量选择和参数估计可以同时获得。自适应Lasso惩罚函数$\rho_\gamma(\beta) = \Sigma_{j=1}^{p}\gamma_j|\beta_j|$通过允许每个回归系数有不同的调优参数，改进了Lasso的选择性能，从而提供更大的灵活性。

令$\ell(\beta)$为(1)中的对数似然函数，γ_j为自适应Lasso惩罚中的非负调优参数。j = 1, . . . , p。对于泊松点过程模型，我们定义β的惩罚对数似然函数为

$$\ell_p(\beta) = -\ell(\beta) + n \Sigma_{j=1}^{p}\gamma_j|\beta_j|。$$

最小化$\ell_p(\beta)$的值是β的惩罚最大似然估计。

3.2 计算算法

为了最小化(3)中的$\ell_p(\beta)$，我们采用迭代算法。在步骤0，β的MLE被用作初始值，$\beta^{(0)} = \beta_{MLE}$。在步骤m，我们使用$\ell(\beta)$的拉普拉斯近似，

$$\ell^*(\beta) = (\beta - \beta^{(m-1)})^T \partial \ell(\beta^{(m-1)})/\partial \beta + (1/2)(\beta - \beta^{(m-1)})^T \partial^2 \ell(\beta^{(m-1)})/\partial \beta \partial \beta^T (\beta - \beta^{(m-1)}),$$

其中$\beta^{(m-1)}$是在步骤m-1获得的，m = 1, 2, . . .

接下来，我们重新排列$\ell^*(\beta)$的项。定义$y^* = (A^T)^{-1}\{\partial \ell(\beta^{(m-1)})/\partial \beta - \partial^2 \ell(\beta^{(m-1)})/\partial \beta \partial \beta^T \beta^{(m-1)}\}$, $X^* = A\text{diag}[\gamma_j]_{j=1}^{p}$，和$\beta^* = \text{diag}[\gamma_j]_{j=1}^{p}\beta$，其中A是$\partial^2 \ell(\beta^{(m-1)})/\partial \beta \partial \beta^T$的Cholesky因子；即，-

$\partial^2 \ell(\beta^{(m-1)})/\partial\beta\partial\beta^T = A^TA$。然后(4)中的近似$\ell^*(\beta)$可以重写为二次形式

$$\ell^*(\beta) = -(1/2)(y^* - X^*\beta)^T(y^* - X^*\beta)。$$

截距β_0不受惩罚，我们使用每个m的轮廓截距估计$\beta_0^{(m)} = \beta_0^{(m-1)}$。此外，我们设置$\gamma_j = \gamma \log(n)|\beta^{(0)}_j|^{-1}, j = 1, \ldots, p$，使得γ是一个共同的调优参数，其中回想$\beta^{(0)} = \beta_{MLE}$[22]。令$\beta(\gamma)$表示$\ell_p(\beta) = -\ell^*(\beta) + n\Sigma_{j=1}^{p}\gamma_j|\beta_j|$的最小化器，可以通过众所周知的最小角回归(LARS)算法获得[5][6]。

为了选择调优参数γ，在这种情况下定义的贝叶斯信息准则(BIC)为

$$BIC(\gamma) = -2(\ell(\beta(\gamma)); \gamma) + e(\gamma) \log(n),$$

其中$e(\gamma) = \Sigma_{j=1}^{p} I(\beta_j(\gamma) \neq 0)$是非零回归系数估计的数量。我们固定γ的路径$\gamma \geqslant 0$并选择调优参数γ和估计$\beta^{(m)}$，使其最小化BIC(γ)。然后用$\beta^{(m)}$替换$\beta^{(m-1)}, m = 1, 2, \ldots$，在(4)中的拉普拉斯近似中，并迭代这个程序，直到满足某个收敛标准。将这个最终估计表示为β。拉普拉斯近似$\ell(\beta)$和通过LARS算法的迭代更新之前已经研究过，用于空间格点数据[20]。

3.3 标准误差

我们使用[16]中给出的渐近协方差矩阵的插件估计来近似回归系数估计的标准误差。在不失一般性的情况下，假设协变量被重新排列，使得我们可以写$\beta = ((\beta^1)^T, (\beta^2)^T)^T$，其中$\beta^1$是非零回归系数估计的向量，$\beta^2 = 0$。令$x^1(s)$表示对应于$\beta^1$在位置s的协变量列向量，令W是一个$(n+M)\times(n+M)$对角矩阵，对角线元素为$w_i \exp\{x(s_i)^T\beta^1\}$，其中$w_i$是求积权重。令$X = [x^1(s_1), \ldots, x^1(s_{n+M})]^T$。假设真实模型仅依赖于$x^1(\cdot)$，观察到的Fisher信息是$-\partial^2\ell(\beta)/\partial\beta^1\partial(\beta^1)^T$。我们代入$\beta^1$并计算$-\partial^2\ell(\beta^1)/\partial\beta^1\partial(\beta^1)^T = (X^TWX)^{-1}$以获得$\beta^1$的协方差矩阵估计。

4. 模拟研究

使用增长的空间域渐近框架来评估我们方法的性能。空间域为$[0, 10]^2$、$[0, 20]^2$和$[0, 30]^2$。生成了七个独立的高斯随机场，协变量函数$C(h) = \sigma^2 \exp(-\phi||h||)$，作为协变量$z(\cdot)$，具有1 × 1单位像素分辨率，其中h是D中两个位置之间的欧几里得距离，σ^2和ϕ是控制位置之间空间相关方差和范围的参数。参数设置为$\sigma^2 = 0.05$和$\phi = 0.5$。我们使用高斯随机场构建空间相关进入协变量，模拟可能的生态因素，如温度或土壤营养含量。为了在协变量之间引入相关性，我们定义Σ，使得$(\Sigma)_{ij} = 0.5^{|i-j|}$，并将独立协变量向量$z(s)$转换为相关协变量向量$x(s) = B^T z(s)$，其中$\Sigma = B^T B$。我们模拟来自具有对数线性强度函数规范$\lambda(s; \beta) = \exp[\beta_0 + x(s)^T \beta]$的泊松点过程的空间点模式，其中$\beta_0 = 0$且$\beta = (4, 3, 2, 1, 0, 0, 0)^T$。

比较了四种拟合程序。$LARS_\infty$更新估计直到收敛。在大数据集的情况下，$LARS_\infty$可能在计算上过于昂贵。我们还研究了$LARS_1$，它在一次迭代后将参数估计作为参数估计，以及$LARS_2$，它在两次迭代后取$\beta^{(2)}$。此外，作为比较这些LARS程序的基准，oracle拟合真实模型，其中前四个协变量。每个程序在100个模拟的空间点模式上运行。

表1显示了我们提出方法的选择性能。对于每个模拟空间域，给出了100个模拟空间点模式的平均样本大小。然后，我们计算每个回归系数和每个拟合程序的非零估计的相对频率。在oracle情况下，如果没有错误地选择协变量，前四列应该是1，最后三列应该是0。

在最小的$[0, 10]^2$空间域上，每个模式的平均事件数约为132。$LARS_1$程序选择最多的非零回归系数，但$LARS_2$和$LARS_\infty$程序正确识别更多的零回归系数。在$[0, 20]^2$空间域上，平均有约1120个事件。所有三个LARS程序似乎选择正确的协变量，因为它们的值接近1，对于前四个协变量，对于最后三个协变量为0。在最大的$[0, 30]^2$空间域上，每个模式有约1884个点，选择最准确。所有三个LARS程序正确选择前四个非零协变量。最后三个协变量在不超过100个模拟模式中的5个中被错误选择。表1还表明，更多的LARS程序迭代倾向于选择更多的零回归系数。此外，LARS程序倾向于导致具有较大幅度的非零回归系数的误差较小。例如，在$[0, 10]^2$空间域上，第一协变量比第二协变量更频繁

地被正确选择，第二协变量比第三协变量更频繁地被正确选择，第三协变量比第四协变量更频繁地被正确选择。这些比较对应于真实回归系数的幅度排序。

表2显示了通过提供回归系数β_1、β_2、β_3和β_4的偏差、方差和均方误差(MSE)，展示了非零回归系数估计的性能，这些系数在三个空间域$[0, 10]^2$、$[0, 20]^2$和$[0, 30]^2$上，使用四种拟合程序。

在最小的$[0, 10]^2$空间域上，LARS程序的估计量相比oracle估计量可能有相当大的负偏差。此外，$LARS_1$估计似乎比$LARS_2$和$LARS_\infty$估计更有偏差，而$LARS_2$和$LARS_\infty$估计似乎具有相似的偏差值。LARS程序的方差和MSE值大于oracle估计量的值，在LARS程序中，$LARS_1$估计似乎具有较低的方差和MSE值，而$LARS_2$和$LARS_\infty$程序的值相似。在$[0, 20]^2$空间域上，尽管LARS估计量仍然比oracle估计量有更多的偏差，但LARS和oracle之间的偏差差异比在$[0, 10]^2$空间域上要小得多。再次，$LARS_1$程序似乎产生比$LARS_2$或$LARS_\infty$程序更有偏差的估计。LARS程序的方差和MSE值大于oracle程序的值，但它们比在$[0, 10]^2$空间域上更相似于oracle程序。在LARS程序中，似乎方差和MSE值是相似的。在最大的$[0, 30]^2$空间域上，偏差、方差和MSE值对于所有程序都是相似的，并且很小。

总的来说，模拟研究的结果表明，这种正则化方法在合理条件下适用于变量选择和估计。在$[0, 10]^2$空间域上，每个空间点模式中的平均事件数很小，该方法低估了前四个协变量，高估了最后三个协变量，其估计具有比oracle程序更大的偏差和方差值。在两个较大的空间域上，平均事件数大于1000，正确的协变量被更频繁地选择，偏差和方差值与oracle程序相当。此外，随着回归系数幅度的增加，选择和估计性能也有所改善。

5. 应用

在这里，我们将分析4026棵Beilschmiedia pendula (B. pendula)树在一个50公顷(ha)(500 m × 1000 m)的研究区域中的位置。除了各种树种的位置信息外，还有关于以5 m × 5 m像素分辨率提供的海拔和坡度的信息。我们模拟B. pendula树的强度$\lambda(s; \beta)$，作为海拔、坡度和其他六种树种出现的对数线性函数

$$\lambda(s; \beta) = \exp[\beta_0 + \Sigma_{j=1}^{8} \beta_j x_j(s) + \beta_9 \text{elevation}(s) + \beta_{10} \text{slope}(s)],$$

其中$x_j(s) = I$ (树种j在位置s) 对于$j = 1, \ldots, 6$。这六种树种是Eugenia nesiotica (E. nesiotica), Eugenia oerstediana (E. oerstediana), Piper cordulatum (P. cordulatum), Protium panamense (P. panamense), Sorocea affinis (S. affinis), 和 Tabebuia rosea (T. rosea)。

图2包含海拔和坡度的图，图3包含B. pendula树的位置图。作为示例，我们在图3中显示了P. panamense在5 m × 5 m、10 m × 10 m和20 m × 20 m像素中的数量。

比较图2中的协变量与图3中绘制的B. pendula树，这些树似乎在具有更温和海拔和较陡坡度的区域更为丰富。图3似乎表明B. pendula和P. panamense树倾向于在研究区域的相似区域生长。

我们在第3节中实现了我们的正则化方法，以在5 m × 5 m、10 m × 10 m和20 m × 20 m像素分辨率下选择协变量。10 m × 10 m和20 m × 20 m像素分辨率上的海拔和坡度值是像素平均值。这些实现的回归系数估计及其标准误差在表3中给出。

结果显示，无论像素分辨率如何，海拔、坡度、P. cordulatum和P. panamense都被一致选择，而E. nesiotica在任何像素分辨率下都不被选择。B. pendula树似乎被P. cordulatum、P. panamense、S. affinis以及海拔和坡度增加的区域所吸引，但被T. rosea和E. oerstediana排斥。

6. 像素分辨率的选择

在上述方法中，计算协变量的像素分辨率是预先选择的，每次考虑单一像素分辨率。所选协变量的集合在不同像素分辨率之间可能不一致。例如，在10 m × 10 m和20 m × 20 m像素分辨率下选择的三种树种在5 m × 5 m像素分辨率下没有被选择，如表3所示。在这里，我们提出我们方法的一个变体，以同时计算多个像素分辨率的协变量值并同时在不同像素分辨率下选择协变量。

假设空间域D被划分为K个不同网格$G_1, G_2, \ldots, G_K$上的像素。令M_k表示像素分辨率k中的像素数量，$k = 1, 2, \ldots, K$。我们设置$M_1 = d_1 d_2$，

其中d_1和d_2是整数，使得$G_1 = d_1 \times d_2$。然后我们设置$M_k = 2^{2(k-1)}M_1 d_k, k = 1, 2, \ldots, K$，使得$G_k = 2^{k-1}d_1 \times 2^{k-1}d_2$。为了在每个像素分辨率获取协变量值，我们从最细的网格G_1开始，并确定如在第2节中的协变量。然后每个像素在较粗的网格G_k对应于$2^{2(k-1)}$个像素在G_1上。我们通过平均适当的$2^{2(k-1)} \times 2^{2(k-1)}$像素子集在$G_k$上计算协变量。回想一下，为了计算泊松点过程对数似然函数，所有协变量都在相同数量的像素M上计算。为了强制这个条件，我们创建新的网格G_k^*，维度为$2^{k-1}d_1 \times 2^{k-1}d_2$，对应于网格$G_k$，包含$2^{2(k-1)}$个常数值块，对应于通过平均$G_k$中的值获得的那些值，$k = 1, 2, \ldots, K$。注意$G_1^* = G_1$。在我们的变体方法中，网格定义K个像素分辨率，用于p个协变量，pK个协变量网格用于获取惩罚估计，如第3节所述。因此，这个变体方法提供了一种方法来选择每个协变量在具有与事件发生更高关联的像素分辨率上。

从此，我们将把这个变体称为多分辨率程序。

在巴罗科罗拉多岛数据中，我们设置$d_1 = 25$, $d_2 = 50$，和$K = 3$。因此，$G_1 = 25 \times 50, G_2 = 50 \times 100$，和$G_3 = 100 \times 200$是所有像素分辨率的网格。在最细的像素分辨率$G_1$上的协变量值是使用第2.2节中描述的求积方案计算的。在每个较粗的像素分辨率G_k上，协变量值从100 × 200网格上平均超过$2^{k-1} \times 2^{k-1}$子网格，$k = 1, 2$。

表4包含使用与之前相同的协变量，但在三个像素分辨率上同时计算的回归系数估计。结果显示，所有八个协变量都在恰好一个像素分辨率上被选择，除了P. cordulatum和P. panamense，再次，E. nesiotica在任何像素分辨率下都不被选择。B. pendula被吸引到在两个较粗像素分辨率上的P. cordulatum和在最细和最粗像素分辨率上的P. panamense。在不同像素分辨率上选择协变量似乎突出了B. pendula和其他树种之间吸引范围的有趣差异。此外，

7. 结论

我们已经开发了一种用于空间泊松点过程的变量选择和参数估计的正则化方法。我们的方法使用自适应Lasso惩罚，并通过拉普拉斯近似和

LARS算法实现。我们的模拟研究表明，该方法在合理条件下表现良好，特别是在较大的空间域和较多事件的情况下。

我们将该方法应用于BCI森林普查数据，研究了Beilschmiedia pendula树的分布与环境因素和其他树种的关系。我们发现B. pendula在海拔较高、坡度较陡的区域更为丰富，并且与某些树种（如P. cordulatum和P. panamense）有正相关，而与其他树种（如T. rosea和E. oerstediana）有负相关。

此外，我们提出了一种多分辨率方法，允许在不同像素分辨率上同时选择协变量。这种方法能够捕捉不同空间尺度上的关系，提供了对空间点过程更全面的理解。

未来的研究方向包括扩展该方法以处理更复杂的空间点过程模型，如Cox过程和Gibbs过程，以及开发更高效的计算算法以处理大规模数据集。

参考文献

1. Baddeley, A., Turner, R., 2000. Practical maximum pseudolikelihood for spatial point patterns. Australian and New Zealand Journal of Statistics 42, 283–322.
2. Berman, M., Turner, T.R., 1992. Approximating point process likelihoods with GLIM. Applied Statistics 41, 31–38.
3. Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. John Wiley & Sons, Hoboken, NJ.
4. Condit, R., 1998. Tropical Forest Census Plots: Methods and Results from Barro Colorado Island, Panama and a Comparison with Other Plots. Springer-Verlag, Berlin.
5. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. The Annals of Statistics 32, 407–499.
6. Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348–1360.

7. Fan, J., Peng, H., 2004. Nonconcave penalized likelihood with a diverging number of parameters. The Annals of Statistics 32, 928–961.
8. Møller, J., Waagepetersen, R.P., 2004. Statistical Inference and Simulation for Spatial Point Processes. Chapman & Hall/CRC, Boca Raton, FL.
9. Miller, A., 2002. Subset Selection in Regression, second ed. Chapman & Hall/CRC, Boca Raton, FL.
10. Huang, H.-C., Hsu, N.-J., Theobald, D.M., Breidt, F.J., 2010. Spatial lasso with applications to GIS model selection. Journal of Computational and Graphical Statistics 19, 963–983.
11. Hubbell, S.P., Foster, R.B., 1983. Diversity of canopy trees in a neotropical forest and implications for conservation. In: Sutton, S.L., Whitmore, T.C., Chadwick, A.C. (Eds.), Tropical Rain Forest: Ecology and Management. Blackwell Scientific Publications, Oxford, pp. 25–41.
12. Hubbell, S.P., Condit, R., Foster, R.B., 2005. Barro Colorado Forest Census Plot Data. URL: https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci.
13. Jensen, J.L., Møller, J., 1991. Pseudolikelihood for exponential family models of spatial point processes. The Annals of Applied Probability 1, 445–461.
14. Møller, J., Waagepetersen, R.P., 2007. Modern statistics for spatial point processes. Scandinavian Journal of Statistics 34, 643–684.
15. Renner, I.W., Warton, D.I., 2013. Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. Biometrics 69, 274–281.
16. Rathbun, S.L., Cressie, N., 1994. Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. Advances in Applied Probability 26, 122–154.

17. Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B 58, 267–288.
18. Wang, H., Li, R., Tsai, C.-L., 2007. Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika 94, 553–568.
19. Wang, H., Li, B., Leng, C., 2009. Shrinkage tuning parameter selection with a diverging number of parameters. Journal of the Royal Statistical Society: Series B 71, 671–683.
20. Waagepetersen, R.P., 2007. An estimating function approach to inference for inhomogeneous Neyman–Scott processes. Biometrics 63, 252–258.
21. Zhu, J., Huang, H.-C., Reyes, P.E., 2010. On selection of spatial linear models for lattice data. Journal of the Royal Statistical Society: Series B 72, 389–402.
22. Zou, H., 2006. The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101, 1418–1429.
23. Zou, H., Li, R., 2008. One-step sparse estimates in nonconcave penalized likelihood models. The Annals of Statistics 36, 1509–1533.
24. Zou, H., Hastie, T., Tibshirani, R., 2007. On the "degrees of freedom" of the lasso. The Annals of Statistics 35, 2173–2192.