科学通报 2010 年 第 55 卷 第 33 期：3205 ~ 3213
论 文 www.scichina.com csb.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS

基于元模型的高致病性禽流感传播风险研究

曹春香[①], 徐敏[①③], 常超一[①③], 薛勇[①], 钟少波[①], 方立群[②], 曹务春[②], 张颢[①], 高孟绪[①③], 何祺胜[①③], 赵坚[①③], 陈伟[①③], 郑盛[①③], 李小文[①]

① 中国科学院遥感应用研究所, 遥感科学国家重点实验室, 北京 100101;
② 军事医学科学院微生物流行病研究所, 北京 100071;
③ 中国科学院研究生院, 北京 100049
E-mail: cao413@irsa.ac.cn

2010-06-10 收稿, 2010-08-19 接受
国家自然科学基金(408711730)、国家重点基础研究发展计划(2007CB714404)和国家科技重大专项(2008ZX10004-012)资助项目

摘要 高致病性禽流感的爆发与环境影响因子及候鸟的迁徙有很强的相关性, 基于 MODIS 的 NDVI 产品数据, 提取了禽流感的一些主要环境影响因子, 利用逻辑斯谛回归的方法生成了高致病性禽流感爆发的风险图, 并利用贝叶斯最大熵模型分析了高致病性禽流感疫情的时空相关性. 结果表明, 水体和高速公路是影响 HPAI 爆发和传播的关键因子, 水禽候鸟是高致病性禽流感传播的主要传染源. 此外, 高致病性禽流感具有很高的时间和空间自相关性. 基于这些发现, 对 2004 年 1 月 23 日~2 月 17 日中国大陆地区高致病性禽流感的爆发风险进行了模拟, 构建了基于环境影响因素的高致病性禽流感预测模型.

关键词
高致病性禽流感
元模型
遥感
地理信息系统
逻辑斯谛回归
时空自相关性

流行性感冒(简称流感)病毒的研究目前已经取得了显著的进步. 流感病毒属于正黏病毒科病毒, 根据核蛋白(NP)和基质蛋白(MP)的不同, 流感病毒共分 3 个类型, 即甲(A)、乙(B)、丙(C)型. 甲型流感病毒能够感染家禽、野生鸟类、人、猪以及海洋生物等, 目前为止还没有发现有其他类型的流感病毒感染鸟类, 因此, 我们提到的禽流感一般都是指由甲型流感病毒感染的鸟禽类疾病综合征. 此外, 流感病毒还可以根据抗原结构进一步划分为 H 类和 N 类, H 代表血细胞凝集素(Hemagglutinin), N 代表神经氨酸酶(Neuramidinase). 到目前为止, 在流感病毒中共发现 16 种 H 类和 9 种 N 类抗原, 这些子类病毒经过组合可以形成新的病毒. 理论上, 这些抗原能够组合形成多达 144 种甲型流感病毒, 而实际上远远比这少得多. 当人被不同种类的病毒感染后, 这些病毒有可能重新交叉变异形成新型病毒, 并且在人与人之间迅速传播, 这将导致非常严重的社会经济及公共卫生问题. 根据禽流感的病原性种类, 禽流感通常可以划分为高致病性禽流感和低致病性禽流感. 低致病性禽流感是非致命性的, 并不表现明显的临床症状, 只会引起一般性的呼吸道或生殖道疾病; 而高致病性禽流感则会导致严重疾病, 其死亡率接近 100%[1~3].

从 2003 年年底以来, 禽流感就开始在世界多个国家频繁爆发. 2004~2005 年间中国大陆地区多个省市相继爆发禽流感疫情, 除我国以外, 亚洲还有其他 10 个国家也爆发了禽流感疫情. 仅仅从 2004 年 1 月 23 日~2 月 26 日期间, 我国就爆发了多达 49 例疫情, 分布在 16 个省市自治区. 2005 年 5 月大量青海湖野鸟(斑头雁)死于高致病性禽流感病毒 H5N1, 直至 2005 年 11 月 22 日, 在我国大陆地区总共爆发了 51 例高致病性禽流感案例. 2005 年 11 月底, 我国还出现了第一例人感染禽流感病例[4]. 图 1 显示了我国大陆地区自 2004 年 1 月~2005 年 11 月间的禽流感月发病数.

初步的研究表明, 高致病性禽流感的爆发具有很强的时空聚集性, 并且和环境影响因素具有紧密

英文版见: Cao C X, Xu M, Chang C Y, et al. Risk analysis for the highly pathogenic avian influenza in mainland China using meta-modeling. Chinese Sci Bull, 2010, 55: 4165—4175, doi: 10.1007/s11434-010-4225-x

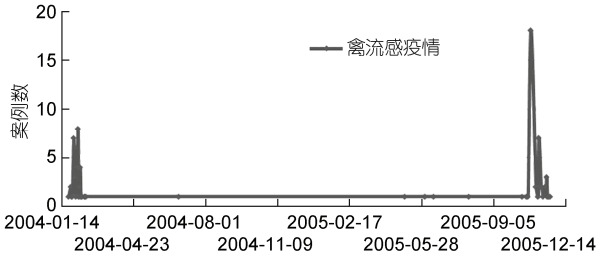


图 1 中国大陆地区禽流感病例月发病数分布图

的联系, 大多数禽流感疫情的爆发地都与大型水体相邻, 并且与水禽密切相关[5]. 已有文献证实了 H5N1 流感病毒的传播主要是通过迁徙鸟类在湖泊间的停留[6]. 实际上, 这里的水禽不仅仅是指野生鸟类, 同时也包括了迁徙鸟类和家禽(如鸭). 迁徙鸟类在途中停留的水体很有可能是禽流感病毒从迁徙鸟类到家禽之间传播的重要载体. 另外, 气候变化尤其是气温的变化也是影响禽流感爆发的一个因素. 因此, 尽管还没有直接的证据来解释环境因素跟禽流感爆发之间的关系, 但是我们有理由相信这两者之间存在很强的联系. 在早期的一些研究中, 遥感和地理信息系统已经成功地应用于疟疾和血吸虫病的分析, 在传染病研究领域显示出其巨大的应用潜力[7,8]. 遥感可以获取禽流感研究所需的环境相关数据, 而地理信息系统可以提供空间分析工具用来解释高致病性禽流感疫情与候鸟迁徙路线及环境之间的关系. 已有研究利用遥感和地理信息系统技术对禽流感进行了定性地分析, 表明气候、鸟类迁徙、交通运输、水体等多种因素都对高致病性禽流感的爆发有显著的影响[9,10]. 然而, 哪些环境影响因素影响禽流感的爆发成为理解禽流感传播的本质问题. 近期有研究表明, 水稻的栽种范围与禽流感的发生呈正相关性[11], 这说明植被是禽流感爆发的一个环境影响因素. 除了植被以外, 地表温度、家禽的畜养环境等因素也引起了相关学者对于高致病禽流感传播特征识别的兴趣, 这些工作主要是利用空间分析功能以及探索性空间数据分析技术, 如叠置分析、缓冲区分析、直方图、散点图、空间聚类等. 本文只是选取了部分重要的环境影响因素以减少因子的自相关性. 考虑自相关性的逻辑斯谛回归模型广泛应用于流行病学研究, 尤其是二值回归模型[12,13]. 结合现代空间及时间统计技术可以提高仿真计算的精度, 因此, 贝叶斯最大熵的方法可以用来进行二分量逻辑斯谛回归模型的有效性验证及不确定性分析[14~16]. 这两个模型的连接显示了环境影响与时空自相关的体系结合, 提出了一种由环境影响因子、时间、空间以及传染性决定的高致病性禽流感扩散的严格数学科学模拟框架[17]. 这些研究都为应用遥感和地理信息系统来分析高致病性禽流感扩散的理论基础, 因此, 定量地分析高致病性禽流感与其影响因子之间的关系具有非常重要的意义.

1 研究方法

1.1 数据处理

(ⅰ) 预处理. 高致病性禽流感的发病数据来自于中国农业部(http://www.agri.gov.cn/)于 2004 年的月报数据. 数据集包括疫情爆发村庄所在的位置以及爆发时间等. 利用国家测绘局提供的 1∶1000000 比例尺中国数字地图对爆发地点进行精确的地理编码(以爆发疫情的村庄中心为爆发地点). 疫情爆发的时间以疑似病例(该病例最终被确诊)报道或者直接确诊(没有经过疑似病例的报道)的时间为准.

本研究选择 MODIS 数据作为遥感数据源. 美国航空航天局(NASA)提供了 MODIS 的各种陆地、大气、海洋产品供全球用户免费下载. 所有 MODIS 数据在应用前经拼接等预处理后都能覆盖中国大陆及其周边地区(72°~136°E, 16°~56°N). 选取了 8 种环境影响因子来作为模型的输入变量, 包括河流、湖泊、水库、国道、铁路、植被、地表温度和湿度. 河流、湖泊、水库、国道和铁路数据来源于国家 1:1000000 基础比例尺数字地图, 相应的 MODIS 陆表温度和湿度产品数据是从美国航空航天局地球科学分布式数据档案中心(http://disc.gsfc.nasa.gov/)下载的. 植被指数数据应用的是相应时期覆盖中国大陆地区的 SPOT 植被指数产品, 图 2 显示的是数据处理的技术流程图.

(ⅱ) 统计分析. 在本研究中, 通过考虑空间和时间权重的二分量逻辑斯谛回归模型分析了高致病性禽流感的发生率和一些环境要素之间的关系. 利用遥感数据反演得到地表温度、近红外水汽、归一化植被指数, 通过地理信息系统方法计算得到水体(河流、湖泊、水库)以及交通干线(国道、铁路)的邻近度

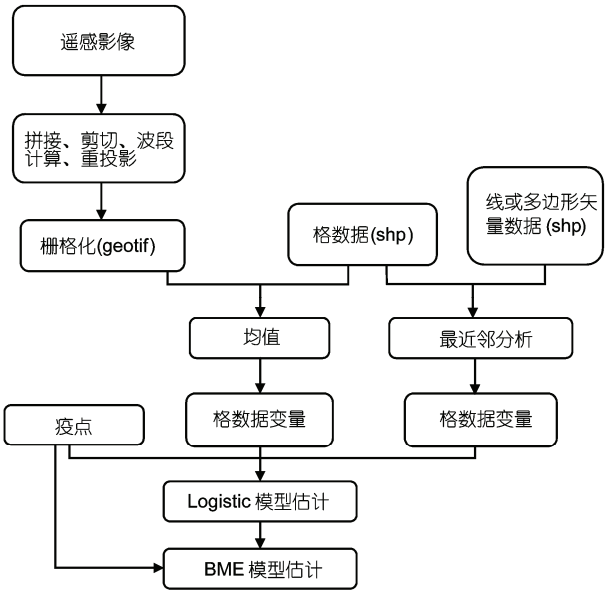


图 2　数据处理技术流程图

(图 3). 考虑到各空间环境数据、候鸟的迁徙路线和禽流感的发生率的关系, 我们研究并定义了 HPAI 发生率的空间权重.

(ⅲ) 空间分析策略. 为了与 MODIS 卫星影像的空间分辨率保持一致, 我们把矢量化的数字地图细分为空间网格(0.5°×0.5°), 在上面标出 HPAI 的爆发点, 并将其作为空间分析的虚拟对象. 我们将其中与流行病学研究中的控制案例(参考案例)相似的对象挑出[18], 这样做我们可以降低选择性偏差, 避免繁琐且无效的数据处理和计算工作. 为了达到目的, 我们设计了如下的采样策略: 首先, 我们将研究区

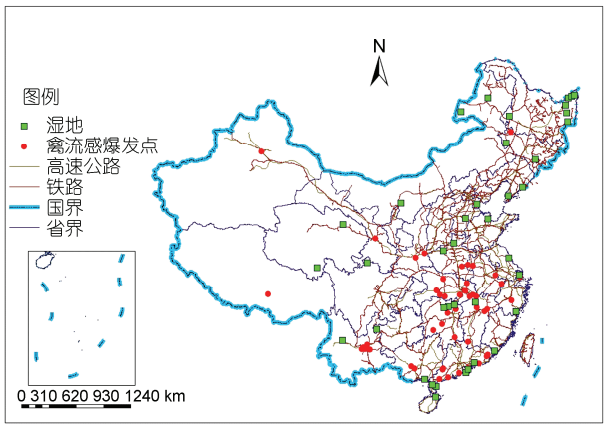


图 3　高速公路、铁路、湿地与 HPAI 爆发情况的空间分布图

(中国大陆)按行和列分割成细小的空间网格(此处间隔为 1°). 然后, 将包含 HPAI 发生点的每个网格提取出来作为案例集, 用于接下来的讨论研究, 如分析 H5N1 病例演变模式和建立预测程序. 此外, 我们还提取出两倍数量的网格作为控制组用于验证建模结果的准确性.

1.2　方法与模型

(ⅰ) 一般逻辑斯谛回归. 对于二值变量的一般线性回归来说, 使用非线性函数来分析这种类型的变量效果比较好. 逻辑斯谛函数是一种被广泛使用的函数类型. 相应的回归模型称为逻辑斯谛回归模型[19]. 令事件发生的条件概率为 $P(y_i=1|x_i)=p_i$, 则逻辑斯谛回归模型假定这种条件概率可以表达如下:

$$p_i=\frac{1}{1+\mathrm{e}^{-(\alpha+\beta x_i)}}=\frac{\mathrm{e}^{(\alpha+\beta x_i)}}{1+\mathrm{e}^{(\alpha+\beta x_i)}}, \tag{1}$$

式中, p_i是第 i 个案例发生的条件概率. 它是协变量的非线性函数, 且能够被转换为线性函数.

根据公式(1), 一个事件发生几率与不发生几率的比值为

$$\frac{p_i}{1-p_i}=\mathrm{e}^{\alpha+\beta x_i}, \tag{2}$$

这个比值称为优势. 通过对优势作自然对数变换, 我们可以得到一个线性函数, 即

$$\ln\left(\frac{p_i}{1-p_i}\right)=\alpha+\beta x_i. \tag{3}$$

(ⅱ) 二分量逻辑斯谛模型. 然而, 由于候鸟影响导致的空间变异性, 这种经典的逻辑斯谛模型对于空间数据的分析是不合适的. 因而专门针对空间数据分析的二分量逻辑斯谛回归模型被提出来[20]. 这个模型是把空间权重函数考虑到经典逻辑斯谛回归模型中, 它能处理空间统计分析中内在的空间变量效应. 通过几十年的发展, 逻辑斯谛回归模型已经得到了广泛的应用.

与经典逻辑斯谛回归模型相比, 二分量逻辑斯谛回归模型可以更好地表达爆发的 0/1 特性(发生与否). 二分量逻辑斯谛回归模型通过如下的公式表达 y_i 的条件概率:

$$p_i=P(y_1^{2_i}\mid y_j(j\neq i))=\frac{\mathrm{e}^{\eta_i}}{1+\mathrm{e}^{\eta_i}}, \tag{4}$$

式中，$\eta_i=\alpha+\boldsymbol{\beta}'\boldsymbol{X}_i+\gamma\hat{y}_i$，$\hat{y}_i=\sum_{j=1}^{n}y_j w_{i,j}$，并且，$\boldsymbol{\beta}=\begin{pmatrix}\beta_1 & \beta_2 & \cdots & \beta_p\end{pmatrix}'$是协变量的参数矢量，$\boldsymbol{X}_i=\begin{pmatrix}x_{i1} & x_{i2} & \cdots & x_{ip}\end{pmatrix}'$是协变量矢量，$\gamma$是对应空间协变量$\hat{y}_i$的参数.

但是对于类似研究中的时间序列数据，需要综合考虑空间函数和时间函数. 此处，我们令

$$w_{ij}=\varpi_{ij}\omega_{ij}, \tag{5}$$

这样，这个公式和逻辑斯谛模型一致. 然而，公式末端的参量的意义是完全不同的. 为了构建 BME 软数据，我们计算得到了每个网格上事件发生的概率.

1.3 建模与参数估计

在我们的研究中，开始阶段所有考虑的变量理论上都是可接受的. 在建模的过程中，我们需要做的是识别和选择那些能很好预测响应变量的候选变量. 一般来说，我们首先拟合一个初步模型，再根据结果修改产生新的模型. 每个模型都有一个不同的数学含义：二分量逻辑斯谛模型考虑不同的环境因素，而 BME 模型则聚焦于空间和时间统计值上. 然后我们对两个模型作对比，通过不断重复这样的调整和对比过程，我们确定最终的模型. 在第一步中，如果有多个协变量，则工作会费时费力. 这种情况下，我们可以让计算机自动选择协变量. 在实际应用中，构成模型的所有变量都应该通过统计 F 检验.

由于考虑空间和时间函数的部分破坏了变量独立的假设，最大似然方法不再适用于该模型的参数估计. 从原理上我们应该选择其他参数估计方法，如蒙特卡洛马尔科夫链方法，然而，为了方便起见，我们使用了 SPSS 13(SPSS 公司，1989~2004 年)中的二分量逻辑斯谛程序. 因此实际上我们使用的是伪似然方法[21].

1.4 与候鸟迁徙路径相关的加权空间函数

在我们的研究中，候鸟被认为是影响高致病性禽流感爆发的主要因素. 图 4 描绘了中国大陆 4 条公认的候鸟迁徙路线[22]：(1) 从内蒙古自治区、甘肃省、青海省到印度和巴基斯坦；(2) 从东北和华北到东南亚；(3) 从东北和华北到日本、菲律宾和澳大利亚；(4) 从蒙古和西伯利亚到华南地区. 在这些迁徙路线上，我们引入了 20 个中途停留点到模型中并进行了计算.

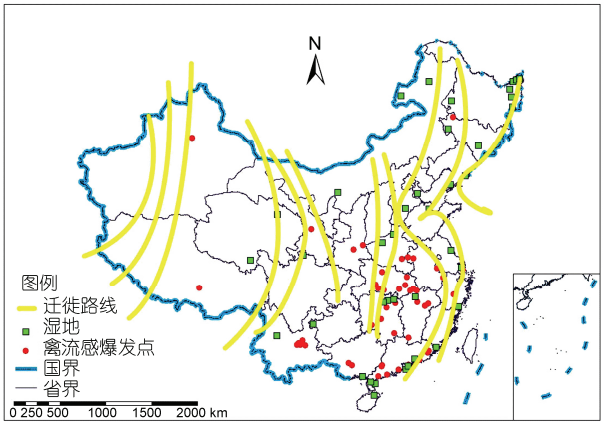


图 4 禽流感爆发点叠加主要候鸟迁徙路线和重要湿地

候鸟迁徙路线摘自国际湿地组织收集和分析的数据)的分布图，引自“野生鸟类甲型流感病毒的全球分布”.

此外，根据对候鸟迁徙特性的研究，这些迁徙鸟类很少直接从出发地不间断地飞到目的地. 由于路途遥远，它们中途需要停靠在一些地方休息. 相比那些它们直接飞过不停留的地区，这些停靠点特别明显地受到影响. 这些停靠点一般都是那些能够为迁徙鸟类提供充足食物的地方，它们很少受到人类的影响，能够提供合适的栖息环境. 这些地方主要是湿地，如湖泊、水库、河流等. 当我们把中国大陆境内主要湿地分布和中国境内主要的候鸟迁徙路线叠加到 2004~2005 年间 HPAI 爆发点分布图时，我们可以看到疾病绝大部分发生在那些湿地的周边地区. 然而，同样明显的是，有些湿地周围很少有甚至没有 HPAI 的爆发.

加权空间函数的确定主要有两种途径：利用先验知识和根据已有数据空间模式进行探索性分析. 通过禽流感爆发与候鸟迁徙路线的相关性研究结果，我们可以进行如下几点假设：(1) 同一条迁徙路线上的停留点及其附近地区之间禽流感发生具有高度相关性. (2) 不在迁徙路线上的站点之间相关性与其距迁徙路线的远近有关. (3) 不同的迁徙路线上的停留点之间相关性很小，应该根据它们之间的空间距离来度量.

根据上面的假设，我们给出如下几条确定禽流感加权空间函数的规则：(1) 如果两站点为同一站点，

则权重为 0. (2) 如果两站点是同一候鸟迁徙路线上的停留点, 那么它们之间的空间权重为 1. (3) 如果两个站点至少一个不在迁徙路线上, 分别计算两个站点距同一迁徙路线上最近停留点的距离, 然后根据这个距离确定它们之间的权重, 如取距离和的倒数. (4) 其他情况计算两站点间距离, 根据这个距离确定它们之间的权重, 如两站点间的欧几里得距离的倒数.

如果规定相同迁徙路线上的两个停留点之间的距离为 0, 则依据上面定义的距离 S, 空间权重可以表示为

$$w_{ij} = \varpi_{ji} = \frac{1}{1+S}. \tag{6}$$

1.5 加权时间函数

某个时间的空间模式通常与早期的空间模式相关. 传染病的传播过程通常表现出这种特征. 究其原因, 疫点和非疫点之间的交互致使空间模式以不同的方式发展. 对这一点可以这样理解: 目前禽流感的爆发情况与过去一段时间禽流感的爆发情况有关系, 而且不同时间点的疫情发生的关联大小与它们之间的时间间隔有关. 通常间隔越小, 相关性越大. 据此, 我们可以将这种时间维的自相关通过随时间间距增加而递减的函数来表达, 例如取如下的形式:

$$w_{ij} = \varpi_{ji} = \frac{1}{\mathrm{e}^{|t_i - t_j|}}. \tag{7}$$

1.6 元模型——贝叶斯最大熵处理

二分量逻辑斯谛模型能够表达和评价传染中的有关过程, 而贝叶斯最大熵可以分析时空变异性, 这变异性同等影响高致病性禽流感的风险. 考虑到对参数和研究应用的不同, 最佳的解决方法是结合两个模型, 两个模型的共同模拟比单个模型降低了不确定性. 贝叶斯最大熵方法通过结合多学科知识库和高阶统计来估计不确定性数据, 它是一种非常有效的地学统计方法. 所有事件(爆发和潜在爆发)被认为是随机场的实现(发展潜力), 这个随机场把时间轴增加到了二维空间中. 所有事件被分成两类——硬数据和软数据. 硬数据指精确的变量, 如观测值和记录值, 标示为 $\chi(\rho) = \chi(s,t)$. 软数据具有连接硬数据和潜在实现, 并且可被多元概率密度函数(PDF)描述的特征, 如

$$F_\chi(\chi)d_\chi = \mathrm{Prob}[\chi^- < P < \chi^+ + d_\chi], \tag{8}$$

式中, Prob.[.]是概率算子, 软数据的概率可以根据它们的累积密度函数来计算. 软数据是根据多学科知识来建立的, 形成结合硬数据的基础知识库的数据集. 随机场的随机期望可以定义为

$$E[g(\chi)] = \int d_\chi g(\chi)\mathrm{prob}F_\chi(\chi) = m_\chi. \tag{9}$$

如果已知均值, 所有的实现可以被分成随机期望和残差, 也就是

$$\chi = \chi_d + m_\chi, \tag{10}$$

式中, χ_d 代表每个实现的残差. 基于残差, 可以利用协方差来定量化时空变异性. 距离决定空间滞后和时间滞后. 空间滞后是两个空间点之间的空间距离, 时间滞后是两个时间点的间隔. 时空协方差可以通过计算获得, 计算中变程和基台值参数是具有传染性含义的. 基台意味着协方差降到平稳值, 时空变程意味着空间和时间尺度自相关的程度. 然后, 一个数学模型可以用来匹配计算的时空协方差. 为了最大化贝叶斯熵, 贝叶斯最大熵基于贝叶斯条件化等式在每个估计实现 χ_k 上计算后验概率密度函数.

$$F(\chi_k) = A^{-1}\int d_\chi f(\mathrm{map})F_\chi(\chi), \tag{11}$$

式中, A 是归一化系数, f(map)表示从时空协方差模型中获得的权重.

虽然二分量逻辑斯谛模型通过了置信区间为95%的 F 统计检验, 但是它在估计中带来了很大的不确定性. 所以结果可以认为是软数据. 高致病性禽流感爆发是硬数据. 通常, 贝叶斯最大熵估计极大地提高了二分量逻辑斯谛模型的精度.

2 结果和验证

本研究中, 主要选择了国道临近度、铁路临近度、水库临近度、湖泊临近度、河流临近度、地表温度、近红外水汽、植被指数 8 种危险因素. 响应结果是高致病性禽流感的出现状态(存在或不存在), 它们是二值变量. 对于出现的栅格, 则值为 1, 否则值为 0. 表 1 描述了协变量和响应的概要.

2.1 单因素分析

对导致高致病性禽流感发生的可疑危险因素进

行单因素分析, 结果显示, 在所调查的 8 种因素中, 其中 6 个与高致病性禽流感发生有关联, 表 2 给出了详细的分析结果. 有 5 个 P 值在 0.005 以下, 即具有非常显著的意义. 这些十分显著因素包括所有调查的交通和水体因素, 这与我们前面推测的水鸟的活动以及与交通相关的禽类贸易对流感的发生有影响是一致的.

2.2 二分量逻辑斯谛回归模型分析

图 5 是利用上述模型对高致病性禽流感的发生概率预测图. 通过使用二分量逻辑斯谛回归模型对高致病性禽流感爆发的模拟结果可利用贝叶斯最大熵方法进行分析. 显然, 二分量逻辑斯谛回归模型提供禽流感爆发的主要环境要素, 同时也给未爆发疾病区域的模拟带来不确定性, 而贝叶斯最大熵方法的主要目的是排除不确定性同时提高模拟的精度.

表 1 变量和意义

变量	描述	单位	类型
协变			
National	与所有国道最近距离	km	连续
Reservoir	与所有水库最近距离	km	连续
Lake	与所有湖泊最近距离	km	连续
Railway	与所有铁路最近距离	km	连续
River	与所有双线河最近距离	km	连续
lst	MODIS 地表温度	℃	连续
Wv	MODIS 近红外水汽	10~1 cm	连续
NDVI	归一化植被指数	10~1	连续
响应			
Present	是否疫点		二元分类

表 2 单因素分析结果

变量	OR 值	OR 值(95%CI)	P 值
National	0.949	0.928~0.97	<0.001
Railway	0.995	0.992~0.998	<0.05
Reservoir	0.983	0.974~0.992	<0.001
Lake	0.927	0.900~0.956	<0.001
River	0.997	0.995~0.999	<0.05
lst	0.974	0.927~1.024	0.304
Wv	0.999	0.967~1.031	0.941
NDVI	0.957	0.806~1.137	0.616

表 3 考虑时空要素的二分量模型分析结果

变量	系数	OR 值	OR 值(95%CI)	P 值
National	–0.113	0.893	0.842~0.948	<0.001
Reservoir	–0.011	0.989	0.976~1.002	<0.1
Lake	–0.039	0.962	0.930~0.994	<0.05
Sigma	1.849	6.353	2.502~16.131	<0.001
Constant	1.464	4.325	–	<0.01

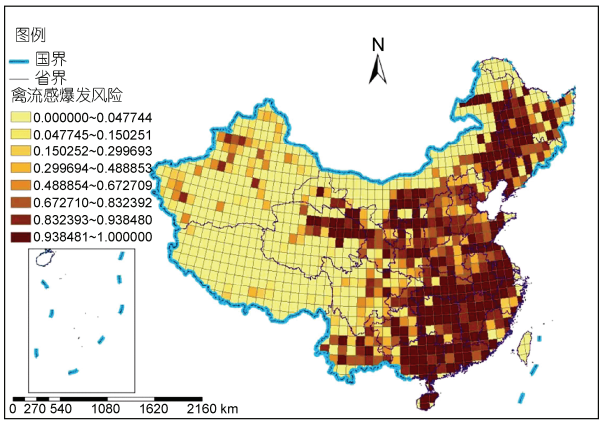


图 5 2004 年 1 月 31 日中国大陆逻辑斯谛预测结果

2.3 贝叶斯最大熵元模型分析

通过圆形缓冲区选择爆发点硬数据周围的模拟点, 这些点组成软数据. 软数据集与影响高致病性禽流感爆发的环境因子高度相关, 并具有一系列的不确定性. 由于自然残差的缘故, 不确定性呈正态分布(高斯概率密度函数分布). 2004 年 1 月 23 日~2 月 17 日之间, 共有 50 个硬数据和 2238 个软数据.

图 6 显示研究期间所有数据的空间分布. 所有数据残差符合正态分布, 并具有时空相关性. 计算贝叶斯最大熵三维随机场的协方差的结果如图 7.

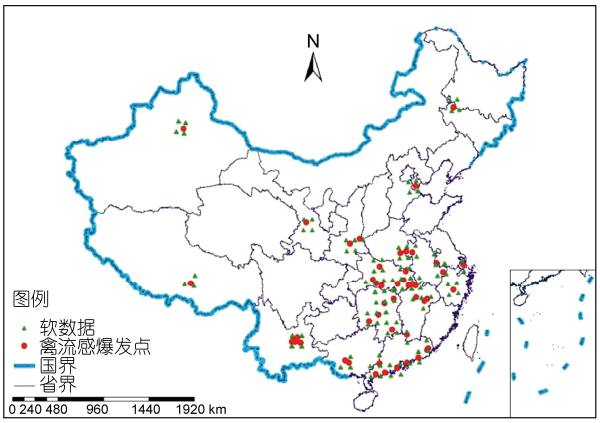


图 6 2004 年中国大陆高致病性禽流感分布图

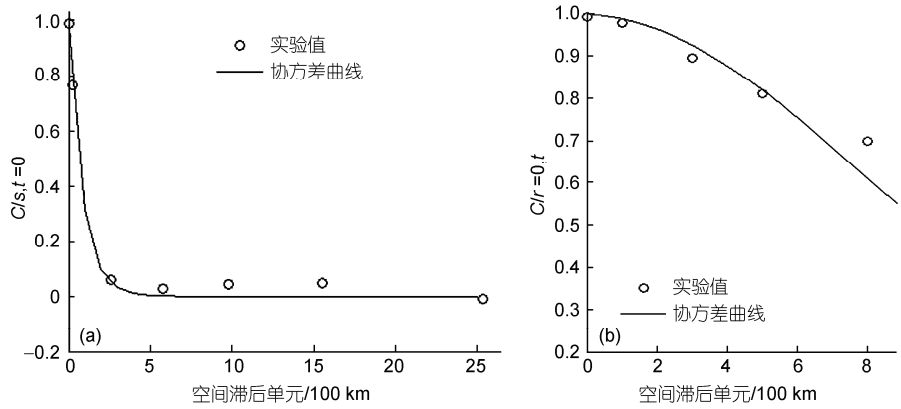


图 7　贝叶斯最大熵方法三维随机场的协方差

(a) 空间尺度上实验和模型的协方差; (b) 时间尺度上实验和模型的协方差

最后，本文构建一个理论协方差模型来拟合该协方差.

$$c_{\chi R}(r=|s-s'|,\tau=|t-t'|)$$
$$=c_1\exp\left(\frac{-3r}{a_{r1}}\right)\exp\left(\frac{-3\tau^2}{at_1^2}\right)+c_2\exp\left(\frac{-3r}{a_{r2}}\right)\exp\left(\frac{-3\tau}{at_2}\right), \quad (13)$$

式中, c_1= 0.08, c_2 = 0.005, a_{r1} = 234.5 km, a_{r2} = 45 km, at_1 = 22 d, at_2 = 4.6 d.

由于我们的协方差模型显示了疫情时空变异性，其参数具有流行病学意义. 由高致病性禽流感的时空结构可以得出，疫情风险是受 2 个时空模式影响: (1) 主要的空间和时间溢出; (2) 干扰. 因此，为了制图目的，我们按时间顺序提取每个估计点的贝叶斯最大熵的协方差均值并验证其准确性.

2.4 贝叶斯最大熵元模型的验证

随机选取高致病性禽流感的爆发点和空白点数据集来验证模型，分别从硬数据和软数据中随机选取验证点共计 55 个(图 8).

为了测度结果图的准确性，我们利用最简单的方法确定误差，即模型估计值与观测值间的差值. 若有结果则认为其值为 1，无结果输出则当作 0，根据绝对风险概率偏离观测值的差值，计算得到误差值(图 9). 其中 X 轴表示观值精度. 我们随机选取 1 d 进行验证(图 10). 计算得到 2004 年 1 月 31 日的结果图，并与简单利用二分量逻辑斯谛模型得到结果的进行比较，发现元模型的总体精

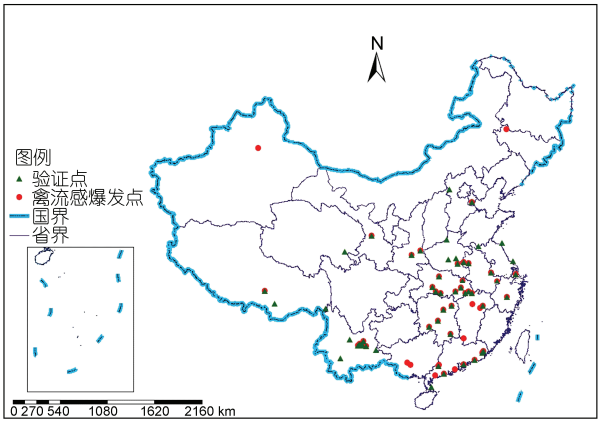


图 8　从硬数据和软数据中随机选取的验证点分布图

度提高了 8.8%. 同时，结果表明 BME 能明显提高精度的稳定性，变异误差降低了 38.8%.

3　结论与展望

本文结合二分量逻辑斯谛回归与贝叶斯最大熵方法的元模型方法，对中国大陆高致病性禽流感的爆发进行了分析与预测. 从不同的 MODIS 数据产品中提取了分析环境要素的基础数据，通过分析候鸟迁徙路线和高致病性禽流感的爆发及传播特性之间的关系，定义了时空权重函数; 利用遥感数据和地理信息系统技术，确定了许多与高致病性禽流感爆发相关的潜在环境变量，结合定义的时空权重构建了逻辑斯谛回归模型. 基于这一模型，我们可以发现影响高致病性禽流感爆发

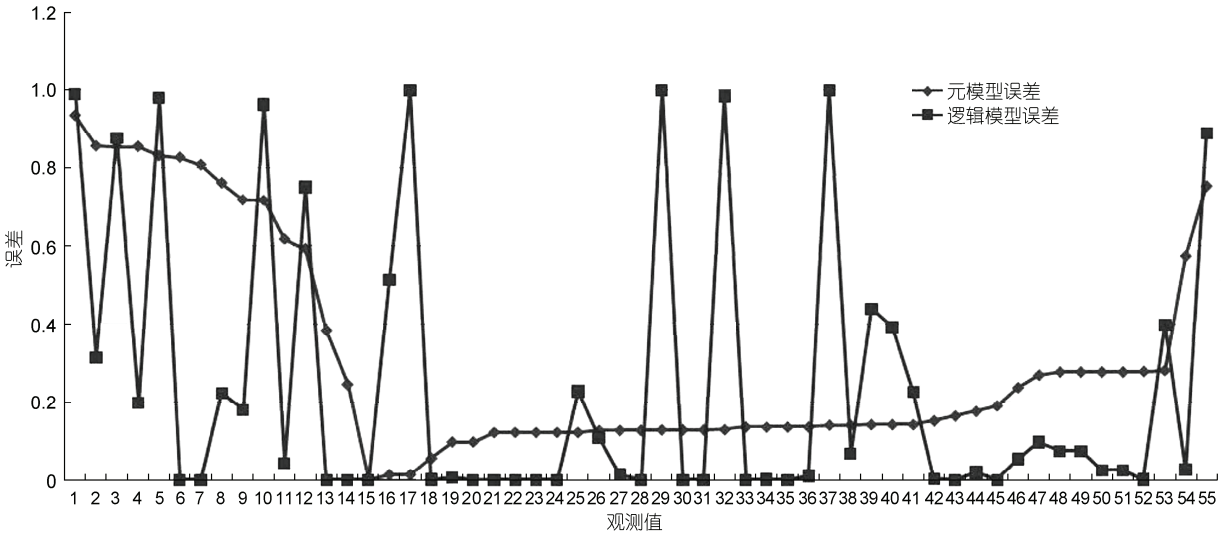


图 9 逻辑斯谛回归模型与元模型验证点的绝对误差对比

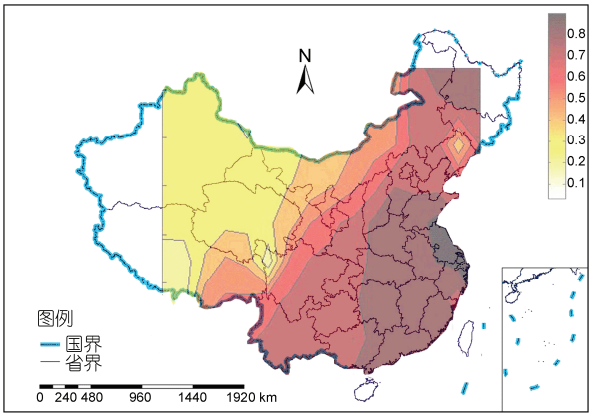


图 10 2004 年 1 月 31 日元模型在中国大陆的预测

的主要变量，从而进行甲型禽流感传染病的爆发预测，以进行有效预防控制.

本文确定了高致病性禽流感的一些主要影响因素. 国道临近度、水库临近度及与湖泊临近度是与高致病性禽流感的爆发相关在统计学上表现显著的主要环境要素. 水体临近度和与高致病性禽流感的爆发之间的密切关联是非常显著的，这或许可以用候鸟的迁徙特性来解释，因为湖泊、水库等水体是候鸟迁徙的主要中途停留地. 水体作为候鸟迁徙路线的组成部分，对于许多家禽类也是很有必要. 当许多候鸟成为流感病毒的宿主，并能够传染给其他鸟类时，这些地方会成为候鸟群体间发生严重交叉感染的最可能地点. 家禽是以传统方式在水体中饲养的，当暴露于受感染的野生鸟类时，会造成候鸟与家禽间的交叉感染. 在某种程度上，这使我们更加确信候鸟和水鸟是高致病性禽流感的重要传染源，而此前研究中对候鸟在高致病性禽流感的传播的作用已有讨论. 研究也证明了与国道的临近度也与高致病性禽流感的爆发有关. 国道是进行家禽贸易的重要运输路线. 在运输、贸易与流通中，感染的家禽可能会传染给其他禽类. 虽然认为病毒在家禽中的流行易于得到控制，禽类作为感染源之一的重要性仍然不应该忽视.

选取的要素(与国道的临近度、与水体的临近度、与湖泊的临近度、与铁路的临近度、与河流的临近度、地表温度、NDVI 和近红外水汽)与高致病性禽流感的爆发间有相对较高的优势比. 我们将这些现象主要归结为水体与候鸟和邻近的家禽之间，以及运输路线和家禽之间的高关联性上. 尽管研究中所有这些因素看似都是有效的，仍应该考虑要素间的自相关性，如临近水库的地方通常也与河流或湖泊相临. 在研究中，由于这 3 种因素(与国度临近度、与水库的临近度和与湖泊的临近度)具有相对较小的 P 值，我们最终选择取了这 3 种因素作为主要因素来构建我们的模型. 对于高致病性禽流感的爆发与传播，仍有更多更复杂的

环境要素有待研究.

对时空多变性的分析也可以利用这 3 种主要要素进行解释. 在 234.5 km 的距离以内, 两个爆发病例间有较高相关性. 这一距离可能是两个城市间进行常规家禽贸易的距离或是鸟类每天迁徙路线的长度. 在这一距离内, 当某一病例爆发后, 鸟类的运输与飞行具有将病毒引入传统市场(这种市场通常在较小的地方内进行自由贸易)和靠近水体(候鸟可能在此停留觅食与休息)的家禽农场的较高可能性. 由于城市与水体的不同分布方式, 致使这一传染病的流行空间范围有 45 km^2 的波动. 另外, 两个病例的爆发可能按时间先后在 22 d 内相继发生. 对于时间范围来说, 一个重要的原因是高致病性禽流感的季节性, 即常发生在冬季与春季. 在这两个季节里的 22 d 可能伴随相对固定的天气状况形成较稳定的气候, 这将造成病毒感染与扩散的可能性. 尽管对潜伏期的确切认识仍有争议, 高致病性禽流感病毒存在的潜伏期已经扩大到较低温度.

尽管我们的研究在进行高致病性禽流感爆发的评估时考虑了主要的环境因素和候鸟, 并对 2004 年间高致病性禽流感的发生进行了模拟, 但我们没有对高致病性禽流感的演变进行很好的预测. 高致病性禽流感与其他的影响决定因素间的关系将会随时间而变化. 在今后的研究中, 我们将探究更多的因素, 尤其是社会经济因素, 如分布位置与范围、地形、地貌、风向、社会经济状况和水体温度等. 我们期望发展一种时空交互模型, 具有综合社会、经济、环境等多种因素的能力, 以对于高致病性禽流感的爆发与传染进行模拟和预测.

参考文献

1 Alexander D J. A review of avian influenza in different bird species. Vet Microbiol, 2000, 74: 3—13
2 Alexander D J. An overview of the epidemiology of avian influenza. Vaccine, 2007, 25: 5637—5644
3 Swayne D E, Suarez D L. Highly pathogenic avian influenza. Rev Sci Tech, 2000, 19: 463—482
4 Zhu Q Y, Qin E D, Wang W, et al. Fatal infection with Influenza A (H5N1) virus in China. New Engl J Med, 2006, 354: 2731—2732
5 钟少波. GIS 和遥感应用于传染病流行病学研究——以乙肝和高致病性禽流感为例. 博士学位论文. 北京: 中国科学院遥感应用研究所, 2006. 67—83
6 Chen H, Smith G J, Zhang S Y, et al. Avian flu: H5N1 virus outbreak in migratory waterfowl. Nature, 2005, 436: 191—192
7 Rogers D J, Randolph S E, Snow R W, et al. Satellite imagery in the study and forecast of malaria. Nature, 2002, 415: 710—715
8 Xu B, Gong P, Biging G, et al. Snail density prediction for schistosomiasis control using Ikonos and ASTER images. Photogramm Eng Rem Sens, 2004, 70: 1285—1294
9 Guo J P, Xue Y, Cao C X, et al. Study on the highly pathogenic avian influenza epidemic using land surface temperature from MODIS data. Int Geosci Rem Sens Symp, 2005, 5: 3599—3602
10 方立群, 曹春香, 陈国胜, 等. 地理信息系统应用于中国大陆高致病性禽流感的空间分布及环境因素分析. 中华流行病学杂志, 2005, 26: 839—842
11 Gilbert M, Chaitaweesub P, Parakamawongsa T, et al. Free-grazing ducks and highly pathogenic avian influenza, Thailand. Emerg Infectious Dis, 2006, 12: 227—234
12 Brownstein J S, Holford T R, Fish D. Effect of climate change on lyme disease risk in North America. Ecohealth, 2005, 2: 38—46
13 Gumpertz M L, Graham J M, Ristaino J B. Autologistic model of spatial pattern of Phytophthora epidemic in bell pepper: Effects of soil variables on disease presence. J Agric Biol Environ Stat, 1997, 2: 131—156
14 Christakos G. Spatiotemporal information systems in soil and environmental sciences. Geoderma, 1998, 85: 141—179
15 Christakos G. Modern Spatiotemporal Geostatistics. New York: Oxford University Press, 2000. 116—215
16 Christakos G, Bogaert P, Serre M L. Temporal GIS. New York: Springer, 2002. 155—166
17 Wang J F, Christakos G, Han W G, et al. Data-driven exploration of “spatial pattern-time process-driving forces” associations of SARS epidemic in Beijing, China. J Public Health, 2008, 30: 234—244
18 Woodward M. Epidemiology: Study Design and Data Analysis. Boca Raton: Chapman & Hall/CRC, 1999. 145—190
19 Hosmer D W, Stanley L. Applied Logistic Regression. 2nd ed. New York: Wiley, 2000
20 Besag J. Spatial interaction and the statistical analysis of lattice systems. J Roy Statist Soc B, 1974, 36: 192—236
21 Cressie N A C. Statistics for Spatial Data. New York: Wiley, 1994
22 Kanai K, Ueta M, Germogenov N, et al. Migration routes and important resting areas of Siberian cranes (Grus leucogeranus) between northeastern Siberia and China as revealed by satellite tracking. Biol Conserv, 2002, 106: 339—346