氣候與環境研究
Climatic and Environmental Research

最佳子集多元线性回归模型在热带气旋风圈变化预报中的应用

饶晨泓 陈光华 陈可鑫 朱志伟

Application of Best-Subsets Multiple Linear Regression Models in Forecasting the Gale-Force Wind Radii of Tropical Cyclones

RAO Chenhong CHEN Guanghua CHEN Kexin ZHU Zhiwei

在线阅读 View online: https://doi.org/10.3878/j.issn.1006-9585.2020.20055

您可能感兴趣的其他文章 Articles you may be interested in

利用人工神经网络模型预测西北太平洋热带气旋生成频数

Prediction of Frequency of Tropical Cyclones Forming over the Western North Pacific Using An Artificial Neural Network Model

气候与环境研究. 2019, 24(3): 324 https://doi.org/10.3878/j.issn.1006-9585.2019.18110

两类El Niño型对西北太平洋季风槽及热带气旋生成的可能影响

Possible Impacts of Two Types of El Nio Events on the Western North Pacific Monsoon Trough and Tropical Cyclogenesis

气候与环境研究. 2018, 23(2): 150 https://doi.org/10.3878/j.issn.1006-9585.2017.17055

西北太平洋季风槽的季节和年际变化特征及其与热带气旋生成大尺度环境因子的联系

Seasonal and Interannual Variability of the Western North Pacific Monsoon Trough and Its Relationship to Large-Scale Environmental Factors

气候与环境研究. 2017, 22(4): 418 https://doi.org/10.3878/j.issn.1006-9585.2016.16065

基于MTCSWA风场资料对西北太平洋热带气旋风场结构的气候统计特征研究

Statistical Characteristics of Wind Field Structures of Tropical Cyclones over the Western North Pacific Based on MTCSWA Data

气候与环境研究. 2020, 25(6): 588 https://doi.org/10.3878/j.issn.1006-9585.2020.19122

海南岛附近区域热带气旋低层风场的变化特征

Characteristics of Tropical Cyclone-Induced Low-Level Wind Fields over Hainan

气候与环境研究. 2018, 23(3): 299 https://doi.org/10.3878/j.issn.1006-9585.2018.17180

FGOALS海洋同化试验对西北太平洋夏季SST—降水关系的模拟评估

Summer Rainfall-SST Relationships in the Western North Pacific Simulated by the FGOALS Model with Ocean Assimilation

气候与环境研究. 2018, 23(2): 139 https://doi.org/10.3878/j.issn.1006-9585.2017.17006

关注微信公众号，获得更多资讯信息

饶晨泓, 陈光华, 陈可鑫, 等. 2021. 最佳子集多元线性回归模型在热带气旋风圈变化预报中的应用 [J]. 气候与环境研究, 26(1): 115−122. RAO Chenhong, CHEN Guanghua, CHEN Kexin, et al. 2021. Application of Best-Subsets Multiple Linear Regression Models in Forecasting the Gale-Force Wind Radii of Tropical Cyclones [J]. Climatic and Environmental Research (in Chinese), 26 (1): 115−122. doi:10.3878/j.issn.1006-9585.2020.20055

最佳子集多元线性回归模型在热带气旋风圈变化预报中的应用

饶晨泓[1,2] 陈光华[1] 陈可鑫[1,2] 朱志伟[3]

1 中国科学院大气物理研究所云降水物理与强风暴重点实验室，北京 100029
2 中国科学院大学，北京 100049
3 南京信息工程大学，南京 210044

摘 要 基于最佳路径（IBTrACS）数据集和欧洲中期天气预报中心（ECMWF）的再分析（ERA-Interim）数据，建立了西北太平洋上（Western North Pacific, WNP）热带气旋（Tropical Cyclone, TC）的七级风圈（R17）变化的最佳子集多元线性回归（bs-MLR）模型。首先根据 2001～2014 年 6～11 月 TC 初始半径（R_{17_0}）的第 1～25、26～50、51～75、76～100 个百分位点将 TC 分为 4 类，建立针对各类 TC 的 bs-MLR 模型，再利用 2015 年 6～11 月的全部 TC 对模型的预报效果进行检验。结果表明：对 TC 生命周期中任意时刻的未来 12 小时 R17（R_{17_12}）进行预报时，当 R_{17_0} 小于 92.6 km 及 R_{17_0} 在 111.1～138.9 km 范围内时，模型对于 R_{17_12} 的趋势预报和大小预报均具有较好的效果；对 TC 生命周期中任意时刻未来 24 小时 R17（R_{17_24}）进行预报时，当 R_{17_0} 在 111.1～138.9 km 范围内时，模式对 R_{17_24} 的趋势预报的效果较好。整体而言，bs-MLR 模型对于 R_{17_12} 的预报准确性高于对 R_{17_24}。

关键词 最佳子集多元线性回归 热带气旋 风圈预报 西北太平洋
文章编号 1006-9585(2021)01-0115-08 **中图分类号** P458.1+24 **文献标识码** A
doi:10.3878/j.issn.1006-9585.2020.20055

Application of Best-Subsets Multiple Linear Regression Models in Forecasting the Gale-Force Wind Radii of Tropical Cyclones

RAO Chenhong[1,2], CHEN Guanghua[1], CHEN Kexin[1,2], and ZHU Zhiwei[3]

1 *Key Laboratory of Cloud–Precipitation Physics and Severe Storms*, *Institute of Atmospheric Physics*, *Chinese Academy of Sciences*, *Beijing* 100029
2 *University of Chinese Academy of Sciences*, *Beijing* 100049
3 *Nanjing University of Information Science and Technology*, *Nanjing* 210044

Abstract Based on the International Best Track Archive for Climate Stewardship dataset and European Centre for Medium-Range Weather Forecasts reanalysis data, the best-subsets multiple linear regression (bs-MLR) models were established by forecasting the gale-force wind radii (R17) of Tropical Cyclones (TCs) in the western North Pacific region.

收稿日期 2020-04-24；**网络预出版日期** 2020-11-12
作者简介 饶晨泓，女，1998 年出生，硕士研究生，主要从事热带气旋研究。E-mail: raochenhong@163.com
通讯作者 陈光华，E-mail: cgh@mail.iap.ac.cn
资助项目 国家重点研发计划项目 2017YFA0603901，国家自然科学基金项目 41975071、41775063
Funded by National Key Research and Development Program of China (Grant 2017YFA0603901), National Natural Science Foundation of China (Grants 41975071 and 41775063)

First, TCs from June to November 2001–2014 were divided into four categories according to the 1−25, 26−50, 51−75, and 76−100 percentiles of the initial sizes (R_{17_0}), and the bs-MLR models for TCs in each category were established. Then all TCs from June to November 2015 were used to test the estimated effectiveness of the bs-MLR models. The results showed that, when R_{17_0} was less than 92.6 km or R_{17_0} was between 111.1 km and 138.9 km, the models had better performances in forecasting the values and changing tendencies of R17 in the next 12 h (R_{17_12}) for any moment of TC life cycle. When R_{17_0} was between 111.1 km and 138.9 km, the models had better performances in forecasting the values and changing tendencies of R17 in the next 24 h (R_{17_24}) for any moment of TC life cycle. Overall, bs-MLR models had higher accuracy in forecasting R_{17_12} than R_{17_24}.

Keywords Best-subsets Multiple Linear Regression, Tropical cyclone, Wind radii forecast, Western North Pacific

1 引言

热带气旋（Tropical cyclone, TC）是地球上最具破坏性的天气系统之一，它会引起诸如风暴潮、台风浪、大风、强降雨等一系列灾害的发生，而西北太平洋（Westem North Pacific, WNP）又是全球TC生成数量最多、生成范围最广的海域之一（Elsberry, 2004）。WNP平均每年有20多个TC生成，约占全球TC生成总数的1/3（袁金南等, 2008），而其中约1/4的TC登陆我国并造成严重影响，因此提高对TC特征指标预报的准确性将有利于提高沿海地区防灾减灾的能力，有助于保障人民的生命财产安全。

Holland and Merrill（1984）将TC的结构变化归为强度（intensity）变化、尺度（size）变化和壮度（strength）变化3种，其中，TC强度用TC风场的极大值表征；TC尺度用TC风场的水平范围表征，一般指七级风圈半径（最大风速半径外，径向风速衰减为17 m/s时距离TC中心的半径；记为R17，单位km）；R17范围内的区域称为内核，风场在内核区域的平均角动量为壮度。相较于TC的其他特征量，目前对TC尺度的研究相对较少，这主要是因为TC尺度的业务预报开展的时间较短，并且TC尺度的观测数据也较少。但是，对TC尺度进行研究却具有十分重要的意义，因为TC尺度可以提供TC灾害的面积，从而更好地预报TC的结构，为TC近海或登陆前后的防灾减灾工作提供帮助（吴磊, 2013）。

除了TC自身信息因子外，目前相当一部分研究工作也强调了大尺度环境场因子对TC尺度的影响（李崇银, 1983; 陈联寿和刘式适, 1997; Khairoutdinov and Emanuel, 2013; Chavas and Emanuel, 2014; Chavas et al., 2016）。Liu and Chan（2002）指出TC尺度变化与天气环流形势相关；Lee et al.（2010）进一步指出大部分的小尺度TC形成于东风波内，而中到大尺度TC的形成常常与季风有关。Maclay et al.（2008）通过更高分辨率的飞机观测资料发现，影响TC尺度变化的主要物理过程与双眼墙替换和大尺度环境场的外部强迫有关，而环境强迫中最重要的因子是垂直风切变（VWS）。通过数值模拟，Wang（2009）强调了外围螺旋雨带潜热释放对TC内核尺度增长具有重要意义，因此他推测，TC内核尺度对相对湿度具有一定的敏感性。Hill and Lackmann（2009）基于位势涡度（PV）的观点也指出相对湿度对TC外围风场的扩展至关重要。Xu and Wang（2010）通过数值模拟进一步验证了相对湿度对TC内核尺度增长具有重要的意义。Lin et al.（2015）研究表明，TC降雨区面积主要与相对海表温度（局地海表温度与热带平均海表温度之差）有关，而降雨量则随绝对海表温度的增加而增加。此外，200 hPa散度和850 hPa涡度可表征大尺度环境强迫场对TC的影响，200 hPa等压面上的温度可表征TC环流高层流出层的温度（Lee et al., 2015）。在前人研究的基础上，本文综合考虑了TC自身信息和大尺度环境场因子对R17的影响，在TC的生命周期中，研究R17的变化情况，并对R17的变化趋势及大小进行预报。

建立统计学模型是TC特征参数预报的重要手段之一。多元线性回归模型在对TC强度的预报研究中已得到应用（Neumann, 1972; Jarvinen and Neumann, 1979; Chu, 1994）。金龙等（2008）选择复共线性关系小的预报因子组合利用逐步回归方法预报台风移动路径。从理论上讲，逐步回归只选出了一个局部最优子集，但是这个子集不一定是所有子集中最优的。此外，逐步回归方法在每一次回归时都引进方差贡献最大的那个因子，这种标准往往会导致选择过多的因子，从而使预报效果不稳定。为了避免逐步回归方法的弊端，Demuth et al.（2006）

将 Miller（2002）提出的最佳子集多元线性回归方法运用于 TC 强度的估计，当有 n 个备选因子时，所有可能的回归模型共有 2^n-1 个，当满足 1）所有因子通过 95% 信度水平、2）平均绝对误差（MAE）最小时，则该因子组合为最佳子集因子组合。为了保证结果的显著性，他还对每个可能的回归方程进行了交叉验证。

本文将尝试运用最佳子集多元线性回归方法研究 TC R17 的变化，对 R17 的两个预报量的变化趋势及大小进行预报。规定观测记录当前时刻为初始时刻，两个预报量分别为：未来 12 小时 R17 的大小（R_{17_12}）和未来 24 小时 R17 的大小（R_{17_24}）。对于预报因子的选择，本文综合考虑了 TC 自身信息（位置、强度）及大尺度环境场因子（热力学因子、动力学因子、水汽因子）对 R17 的影响，以进一步提高对 TC R17 变化趋势及大小预报的水平。

2　数据与方法

2.1　数据

2.1.1　数据的来源

本文所用的 TC 数据来自第四代最佳路径（International Best Track Archive for Climate Stewardship，IBTrACS）数据集（Knapp et al., 2010），具体为 TC 中心的纬度（LAT）、经度（LON）、最大持续风速（MSW）、最低海平面气压（MSLP）、离岸距离以及 4 个象限（东北 NE、东南 SE、西南 SW、西北 NW 象限）的 R17（R17_NE、R17_SE、R17_SW、R17_NW）。将西北太平洋上（0º～40ºN，140ºE～180º）2001～2015 年 6～11 月的数据作为研究数据，时间分辨率为 6 h，即在 00:00（协调世界时，下同）、06:00、12:00、18:00 各有一次观测记录 (https://www.ncdc.noaa.gov/ibtracs/[2020-03-10])。

大尺度环境场数据来自欧洲中期天气预报中心（European Centre for Medium-Range Weather Forecasts，ECMWF）的再分析（ERA-Interim）数据，具体为 925 hPa、700 hPa 的比湿，海表温度，850 hPa、200 hPa 的风场，200 hPa 的散度，850 hPa 的相对涡度以及 200 hPa 的温度。时间、空间的范围和及时间的分辨率同 TC 数据，水平分辨率为 0.75º（纬度）×0.75º（经度）(https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim[2020-03-10])。

2.1.2　物理量的计算

本文综合考虑了 TC 自身信息因子以及大尺度环境场因子对 TC R17 变化的影响，选取的 10 个预报因子及其具体描述如表 1 所示，其中 LON、LAT、MSW、MSLP 直接来自 IBTrACS 数据集，6 个大尺度环境场因子的计算方法如下：首先，寻找距离 TC 中心最近的 ERA-Interim 格点作为极点建立极坐标系，每隔 45º方位角取一条极坐标半径；然后，取海表温度的径向格点分辨率为 0.1º，剩余 5 个大尺度环境场因子的径向格点分辨率为 1º，将 ERA-Interim 格点数据插值到极坐标网格中；最后，对于 925 hPa、700 hPa 的比湿，850 hPa、200 hPa 的风场，200 hPa 的温度，为了去除 TC 环流对环境场的影响，取 TC 半径 2º～7º之间的圆环区域，

表 1　10 个预报因子的简称、单位与描述

Table 1　Abbreviations, units, and descriptions of the ten predictors

类别	简称	单位	描述
位置	LAT	ºN	纬度
	LON	ºE	经度
强度	MSW	m/s	TC近中心最大持续风速
	MSLP	hPa	TC中心最低海平面气压
水汽因子	Q	kg/kg	925～700 hPa上r=2º～7º范围内的平均比湿
热力学因子	T200	℃	200 hPa上r=2º～7º范围内的温度
	SSTA	℃	r=0～0.5º范围内的海表面温度距平
动力学因子	VWS	m/s	850～200 hPa上r=2º～7º范围内的垂直风切变
	div200	s^{-1}	200 hPa上r=0～9º范围内的散度
	vor850	s^{-1}	850 hPa上r=0～9º范围内的相对涡度

注：半径定义为r；SSTA为各观测记录的局地海表温度与所有观测记录平均海表温度的差。

对于 200 hPa 的散度，850 hPa 的相对涡度，取 TC 半径 9º以内的圆形区域，对于海表温度，取 TC 半径 0.5º以内的圆形区域，计算极坐标系中不同区域的平均值（DeMaria and Kaplan, 1994; Maclay et al., 2008; 梅双丽和江静, 2012）。

平均比湿（Q）定义为

$$Q=\frac{q_{925}+q_{700}}{2}, \qquad (1)$$

其中，q_{925}、q_{700} 分别表示 925 hPa、700 hPa 高度场上 TC 半径 r=2º～7º范围内的区域平均湿度。

在前人的研究中 VWS 有多种定义方法（Gray, 1968; 陈联寿和丁一汇, 1979），本文采用了 Palmer and Barnes（2002）求取 VWS 的方法，将 VWS 定义为

$$\mathrm{VWS}=\sqrt{\left(\bar{u}_{200}-\bar{u}_{850}\right)^2+(\bar{v}_{200}-\bar{v}_{850})^2}, \qquad (2)$$

其中，$\bar{u}_{200}$和$\bar{u}_{850}$、$\bar{v}_{200}$和$\bar{v}_{850}$分别表示 200 hPa 和 850 hPa 高度场上 TC半径 r=2º～7º范围内的区域平均纬向风、经向风。

2.2 方法

考虑到不同初始尺度的 TC 存在较大的个体差异，为了使回归结果具有更强的针对性，本文按照初始尺度的大小将 TC 分为 4 类，针对每类中单个 TC 的各预报量运用最佳子集多元线性回归方法。该方法将遍历 n 个预报因子所有可能的组合并建立回归方程，例如 n=3，将产生 $2^n-1=7$ 个多元线性回归方程：3 个包含 1 个预报因子、3 个包含 2 个预报因子和一个包含 3 个预报因子的回归方程。通过一定的标准，分别在 4 类中挑选最优因子组合建立最佳子集的多元线性回归方程。

我们挑选预报因子的标准是基于 F 检验以及各种因子组合建立回归方程所产生的 MAE，具体步骤如下：首先，对每个 TC 所记录的 R17 数据与 10 个预报因子的不同组合之间分别建立回归方程，确定方程系数，并对各预报因子进行 F 检验。然后，通过计算不同因子组合的回归方程的 MAE 来考察回归效果。为了得到每个 TC 每种因子组合方程更加稳定的 MAE，我们在计算 MAE 时采用了 5-折交叉验证的方法，具体步骤如下：假设第 g（g≤4）类中共有 K_g 个 TC，对于每个 TC 都有 $2^{10}-1=1023$ 种因子组合方式，计算第 m 种因子组合方程的 MAE 时，1）由于采用 5-折交叉验证，因此将预报量和预报因子的所有数据大致均分为 5 组；2）分别选取其中第 i（i≤5）组作为测试组，剩余 4 组作为训练组，利用训练组的样本数据进行拟合，得到针对第 i 个测试组的回归方程，接着利用测试组的样本数据计算回归方程的平均绝对误差（MAE(g,k,m,i)，其中 g、k、m、i 分别表示第 g 类的第 k 个 TC、第 m 种因子组合、第 i 个测试组）；3）根据如下公式：

$$\mathrm{MAE}(g,k,m)=\frac{1}{5}\sum_{i=1}^{5}\mathrm{MAE}(g,k,m,i), \quad g=1,2,3,4,$$
$$k=1,2,3,\ldots,K_g, \quad m=1,2,3,\ldots,1023, \qquad (3)$$

计算 5 个测试组 $mae_{(g,k,m,i)}$ 的平均值作为第 g 类的第 k 个 TC 第 m 种因子组合方程的 $\mathrm{MAE}_{(g,k,m)}$。最后，对于第 g 类第 m 种因子组合，计算 K_g 个 TC 中预报因子全部通过 95% 信度水平（F 检验）的回归方程对应的 MAE 的平均值 $\mathrm{MAE}_{(g,m)}$，选取 $\mathrm{MAE}_{(g,m)}$ 中最小值所对应的因子组合作为第 g 类 TC 的最佳子集因子组合。

2.3 数据处理

2.3.1 预报因子和预报量数据的筛选

根据 GB/T 19201～2006《热带气旋等级》国家标准，将 TC 按照 MSW 的大小分为热带低压（TD）、热带风暴（TS）、强热带风暴（STS）、台风（TY）、强台风（STY）和超强台风（SuperTY）6 个等级。本文只选用强度大于 TS 级别，即 MSW≥17.2 m/s 的 TC 观测记录作为研究对象。Knaff et al.（2016）对两个路径不同的 TC 进行研究后发现，当 TC 登陆后，地形的分布差异将给 TC 风场结构不对称变化带来不同的影响，为了减小地形对 TC 风场结构的影响，本文只对位于海上的 TC R17 的变化进行研究，即筛选离岸距离≥500 km 的 TC。在 Demuth et al.（2004）的研究中，R17 取 R17_NE、R17_SE、R17_SW、R17_NW 的平均值，包括了那些不存在风半径的象限。为了降低通过象限平均得到的 R17 数据的噪音，本文去除了存在风半径缺测的 TC 观测记录，同时也避免了 R17 计算值偏小的现象。

将 WNP 上 2001～2014 年 6～11 月的数据作为训练样本，2015 年 6～11 月的数据作为测试样本。为了提升回归的效果，本文还对预报因子和预报量数据进行了质量控制：1）删除训练样本与测试样本中存在缺测值的观测记录；2）将训练样本与测试样本中各预报因子和预报量数据分别按照升序排列，删除小于第 1 个百分位点和大于第 99 个

表 2　预报 R_{17_12} 和 R_{17_24} 时训练样本和测试样本中符合筛选条件的 TC 个数及观测记录数量

Table 2　Number of TCs and observation records in the training set and test set that matched the screening conditions when forecasting R_{17_12} and R_{17_24}

	训练样本		测试样本	
	TC个数	观测记录数量	TC个数	观测记录数量
R_{17_12}	30	861	11	277
R_{17_24}	26	715	11	265

百分位点的极端观测记录；3）删除训练样本中观测记录数量少于 20 个的 TC，删除测试样本中观测记录数量少于 10 个的 TC。经过筛选后，训练样本和测试样本中，对 R_{17_12} 进行预报时的 TC 个数和观测记录数量大于或等于对 R_{17_24} 进行预报时的 TC 个数和观测记录数量（表 2）。

2.3.2　预报因子与预报量相关性的考察

本文基于前人研究，考虑了 TC 自身信息以及大尺度环境场对 R17 的影响，选取了 10 个预报因子，并进一步考察了这些预报因子与预报量之间的相关性，以确保在构建回归模型时能取得较好的回归效果。预报因子与预报量之间相关性的衡量方法如下：1）当预报某一预报量时，在第 k 个 TC（共 K 个 TC）的生命周期中，分别计算每个预报因子与预报量的相关系数 r_{kd}（k=1,2,...,K，d=1,2,..., 10；r_{kd} 表示第 k 个 TC 的第 d 个预报因子与预报量的相关系数）；2）对于第 d 个预报因子，计算 K 个 TC 中通过 95% 信度水平（t 检验）的相关系数绝对值的平均值 $\overline{|r_d|}$。从图 1 中可以看到，所有的 $\overline{|r_d|}$ 值均大于 0.5，说明这 10 个预报因子与 R_{17_12} 和 R_{17_24} 之间的相关关系较好，对预报量的变化具有一定的影响。

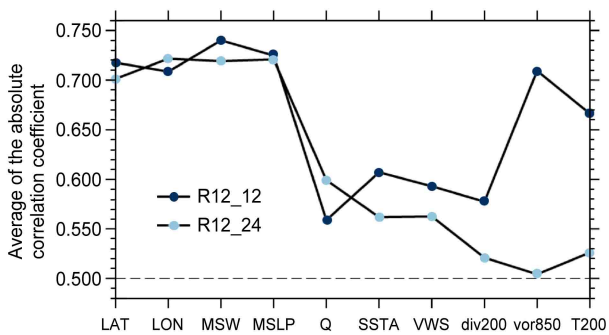

图 1　预报 R_{17_12} 和 R_{17_24} 时各预报因子与预报量相关系数绝对值的平均值

Fig. 1　Average of the absolute correlation coefficient between each predictor and predictand when forecasting R_{17_12} and R_{17_24}

3　模型构建

除了大尺度环境场因子和 TC 自身信息外，前人的研究中（Cocks and Gray, 2002; Xu and Wang, 2010）还强调了初始涡旋大小对 TC 尺度的影响，因此本文根据 TC 初始涡旋尺度对 TC 进行分类，从而得到针对不同初始涡旋尺度的某一类 TC 的最佳子集多元线性回归（bs-MLR）模型。将经过 2.3.1 筛选后，TC 观测序列中的第一个 R17 作为此 TC 的初始尺度（R_{17_0}），将 R_{17_0} 按照升序排列，分别以 R_{17_0} 的第 25 个、第 50 个、第 75 个百分位点作为分界阈值，将 TC 分为 4 类。分类结果如表 3 所示，预报 R_{17_12} 和 R_{17_24} 时，R_{17_0} 的第 25 个、第 50 个、第 75 个百分位点分别对应为 92.6 km、111.1 km、138.9 km。

在 4 个分类中，针对各类 TC 的各预报量运用最佳子集多元线性回归方法，得到各类 TC 的最佳子集因子组合；然后，将各类中所有 TC 的最佳子集因子组合方程对应的系数求平均，确定出不同 R_{17_0} 分类时针对某一预报量的回归方程。由表 4 可知，各类 TC 回归方程的最佳预报因子个数在 2～5 个左右，因子数目较少，可有效避免预报效果不稳定的现象发生。

4　结果分析

得到了针对不同 R_{17_0}、不同预报量的回归方程后，本文选取了 2015 年 6～11 月的全部 11 个 TC（观测记录数量大于 10 个的 TC）：Chan-hom、Nangka、Halola、Soudelor、Goni、Atsani、Kilo、Krovanh、Dujuan、Champi、In-fa，对 bs-MLR 模型的预报效果进行检验。分别用 MAE、符合指数（d）和相关系数（R）的大小来评估 bs-MLR 模型对于预报量的预报情况（表 5）。符合指数 d 用 Willmott（1982）提出的值位于 0～1 之间的无量纲数表示，d 的值越大表示符合程度越高，公式如下：

$$d = 1 - \left[\sum_{i=1}^{N} (F_i - O_i)^2 \Big/ \sum_{i=1}^{N} \left(|F_i - \overline{O}| + |O_i - O| \right)^2 \right], \quad (4)$$

其中，F_i 表示第 i 次预报值，Q_i 表示第 i 次观测值，$\overline{O}$ 表示观测平均值，N 为单个 TC 的观测记录次数。

表 3 预报 R_{17_12} 和 R_{17_24} 时的分类情况及观测数量

Table 3 Categories of R_{17_12} and R_{17_24} with their observed numbers

	观测数量			
	第一类(R_{17_0}<92.6 km)	第二类(92.6≤R_{17_0}<111.1 km)	第三类(111.1≤R_{17_0}<138.9 km)	第四类(R_{17_0}≥138.9 km)
R_{17_12}	5	7	10	8
R_{17_24}	4	7	7	8

表 4 预报 R_{17_12} 和 R_{17_24} 时各类 TC 的回归方程

Table 4 Regression equations of each TC category when forecasting R_{17_12} and R_{17_24}

预报量	回归方程
R_{17_12}	$R_{17_12\,(1)}$=−3836.50+3.45LAT+14.73MSW+3.60MSLP
	$R_{17_12\,(2)}$=357.38+3.32MSW−21447.39Q+3.93VWS
	$R_{17_12\,(3)}$=1521.51−1.75MSLP+30607.96Q+691563.4vor_{850}
	$R_{17_12\,(4)}$=168.87+3.23LAT+995334.2vor_{850}
R_{17_24}	$R_{17_24\,(1)}$=2571.15+0.09LON−2.10MSLP−16445.6Q−4.72SSTA+2.60VWS
	$R_{17_24\,(2)}$=334.76−3.36LON+3.92MSW+0.24MSLP−3.68SSTA
	$R_{17_24\,(3)}$=515.20−3.53LON+22493.27Q−29.91SSTA
	$R_{17_24\,(4)}$=509.41−0.38LON+1.45MSW−647425div_{200}+3.89T_{200}

注：$R_{17_12\,(1)}$表示第一类R_{17_12}，其他预报量的含义以此类推。

表 5 测试样本误差分析

Table 5 Error analysis of the test set

预报量	类别	TC名称	MAE/km	d	R
R_{17_12}	一	Chan-hom	33.94	0.86	0.89*
		Nangka	56.78	0.77	0.61*
		Halola	92.91	0.45	0.91*
	二	Atsani	108.22	0.56	0.68*
		Krovanh	86.30	0.16	0.12
	三	Soudelor	56.14	0.82	0.90*
		Goni	73.91	0.55	0.86*
	四	Kilo	30.79	0.62	0.76*
		Dujuan	62.91	0.57	0.93*
		Champi	59.00	0.45	0.19
		In-fa	79.60	0.29	−0.08
R_{17_24}	一	Chan-hom	124.96	0.43	0.79*
		Nangka	156.17	0.35	0.17
		Halola	202.28	0.24	0.76*
	二	Atsani	125.18	0.58	0.90*
		Krovanh	54.64	0.13	−0.23
	三	Soudelor	29.16	0.80	0.82*
		Goni	91.08	0.34	0.65*
	四	Kilo	50.18	0.48	0.66*
		Dujuan	41.72	0.67	0.95*
		Champi	43.07	0.53	0.50*
		In-fa	106.07	0.27	−0.09

注：*表示该相关系数通过95%信度水平。

由表 5 可知，当预报 R_{17_12} 时，4 类 TC 的预报效果不同。其中预报效果最好的是第三类 TC，Soudelor 和 Goni 的 R 值分别为 0.90 和 0.86，d 值分别为 0.82 和 0.55，说明该类 TC 的回归方程不仅对 R_{17_12} 变化趋势的预报效果较好，对于 R_{17_12} 的大小也具有一定的预报能力。其次，第一类 TC 的预报效果也较好，Chan-hom、Nangka 和 Halola 的 R 值分别为 0.89、0.61、0.91，说明该类 TC 的回归方程对 R_{17_12} 的变化趋势具有一定的预报能力，其中 Chan-hom 和 Nangka 的 MAE 分别为 33.94 km 和 56.78 km，和其他 TC 相比值较小，d 值分别为 0.86 和 0.77，说明该类 TC 的回归方程对于 R_{17_12} 的大小也具有一定的预报能力。对于第二类 TC，Atsani 的 R 值为 0.68，d 值为 0.56，但是 Krovanh 的 R 值未通过 95% 信度水平，并且 d 值仅为 0.16，因此该类 TC 的回归方程对 R_{17_12} 的变化趋势和 R_{17_12} 的大小的预报能力较差。对于第四类 TC，Kilo 和 Dujuan 的 R 分别为 0.76 和 0.93，d 值分别为 0.62 和 0.57，但是 Champi 和 In-fa 的 R 都未通过 95% 信度水平，并且 d 值都较小，因此该类 TC 的回归方程对 R_{17_12} 的变化趋势和 R_{17_12} 的大小的预报能力也较差。总体来说，运用 bs-MLR 模型对 R_{17_12} 进行预报时，R_{17_0} 小于 92.6 km 和 R_{17_0} 位于 111.1～138.9 km 之间的 TC，回归方程对 R_{17_12} 变化趋势的预报效果较好，对 R_{17_12} 的大小具有一定的预报能力。

当预报 R_{17_24} 时，4 类 TC 的预报效果与预报 R_{17_12} 时有所不同。对于第三类 TC，Soudelor 和

Goni 的 R 值分别为 0.82 和 0.65，虽然 Soudelor 的 d 值为 0.82，但是 Goni 的 d 值仅为 0.34，说明该类 TC 的回归方程对 R_{17_24} 的变化趋势具有一定的预报能力，但是对 R_{17_24} 大小的预报能力较差。对于第一、二、四组 TC，每一类中皆存在 1 个 TC 的 R 未通过 95% 信度水平，因此这些 TC 的回归方程对 R_{17_24} 的变化趋势的预报能力较差，并且每一类中至少存在 1 个 TC 的 MAE 较大，d 值较小，因此这些 TC 的回归方程对 R_{17_24} 的大小的预报能力也较差。总体来说，运用 bs-MLR 模型对 R_{17_24} 进行预报时，R_{17_0} 位于 111.1～138.9 km 之间的 TC，回归方程对 R_{17_24} 的变化趋势具有一定的预报能力，但是对 R_{17_24} 的大小不具有预报能力。

通过对比 R_{17_12} 和 R_{17_24} 的预报效果可以发现，预报 R_{17_12} 时，可对两类 TC 的变化趋势和大小进行预报；而预报 R_{17_24} 时，却只能对一类 TC 的变化趋势进行预报，R_{17_12} 的预报效果优于 R_{17_24} 的预报效果，说明预报时效越短，趋势预报的准确性越高。

5　结论和讨论

本文运用最佳子集多元线性回归方法，针对 2001～2014 年 6～11 月 WNP 上强度大于 TS 级别的 TC，选取 LAT、LON、MSW、MSLP、Q、SSTA、VWS、div_{200}、vor_{850}、T_{200} 这 10 个预报因子，对于不同 R_{17_0} 的各类 TC 构建 bs-MLR 模型。首先，以 TC R_{17_0} 的第 25、50、75 个百分位点作为分界阈值，将所有 TC 分为 R_{17_0} 不同的 4 类；然后，针对各类 TC 的各预报量运用最佳子集多元线性回归方法，得到各类 TC 的最佳子集因子组合；最后，将各类中所有 TC 的最佳子集因子组合方程所对应的系数求平均，确定出不同 R_{17_0} 分类时针对某一预报量的回归方程。利用 2015 年的 11 个 TC 对各组回归方程的预报效果进行检验。研究结果表明：1）bs-MLR 模型对 R_{17_12} 进行预报时，R_{17_0} 小于 92.6 km 和 R_{17_0} 位于 111.1～138.9 km 之间的 TC，回归方程对 R_{17_12} 变化趋势的预报效果较好，对 R_{17_12} 的大小具有一定的预报能力；2）bs-MLR 模型对 R_{17_24} 进行预报时，R_{17_0} 位于 111.1～138.9 km 之间的 TC，回归方程对 R_{17_24} 的变化趋势具有一定的预报能力，但是对 R_{17_24} 的大小不具有预报能力；3）预报时效越短，预报的准确性越高。

在未来的研究中，可以考虑将最佳子集多元线性回归方法结合动力模式进行动力统计预报，但是目前还有一些因素制约着最佳子集多元线性回归方法预报的准确性，比如：1）是否还存在更合理的 TC 的分类指标；2）影响 WNP TC R17 变化的因素有很多，是否存在更重要的预报因子未被选入，是否存在重要的物理过程未被考虑；3）本文的研究方法是否适用于 TC 强度变化、频数预报等其它特征参量的研究，这些都需要未来更近一步的研究。

参考文献（References）

Chavas D R, Emanuel K. 2014. Equilibrium tropical cyclone size in an idealized state of axisymmetric radiative–convective equilibrium [J]. J. Atmos. Sci., 71(5): 1663−1680. doi:10.1175/JAS-D-13-0155.1

Chavas D R, Lin N, Dong W H, et al. 2016. Observed tropical cyclone size revisited [J]. J. Climate, 29(8): 2923−2939. doi:10.1175/JCLI-D-15-0731.1

陈联寿, 丁一汇. 1979. 西太平洋台风概论 [M]. 北京: 科学出版社, 107–109.　Chen Lianshou, Ding Yihui. 1979. An Introduction to the Western Pacific Typhoon (in Chinese) [M]. Beijing: Science Press, 107–109.

陈联寿, 刘式适. 1997. 关于热带气旋尺度变化机制的初步分析 [J]. 热带气象学报, 13(2): 105−111.　Chen Lianshou, Liu Shikuo. 1997. A preliminary analysis on mechanism of size change in tropical cyclone [J]. Journal of Tropical Meteorology (in Chinese), 13(2): 105−111. doi:10.1007/s00376-997-0061-6

Chu J H. 1994. A regression model for the western north pacific tropical cyclone intensity forecasts [R]. NRL Memo. Rep. 7541-94-7215.

Cocks S B, Gray W M. 2002. Variability of the outer wind profiles of Western North Pacific typhoons: Classifications and techniques for analysis and forecasting [J]. Mon. Wea. Rev., 130(8): 1989−2005. doi:10.1175/1520-0493(2002)130<1989:VOTOWP>2.0.CO;2

DeMaria M, Kaplan J. 1994. A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin [J]. Wea. Forecasting, 9(2): 209−220. doi:10.1175/1520-0434(1994)009<0209:ASHIPS>2.0.CO;2

Demuth J L, DeMaria M, Knaff J A, et al. 2004. Evaluation of Advanced Microwave Sounding Unit tropical-cyclone intensity and size estimation algorithms [J]. J. Appl. Meteor., 43(2): 282−296. doi:10.1175/1520-0450(2004)043<0282:EOAMSU>2.0.CO;2

Demuth J L, DeMaria M, Knaff J A. 2006. Improvement of advanced microwave sounding unit tropical cyclone intensity and size estimation algorithms [J]. J. Appl. Meteor. Climatol., 45(11): 1573−1581. doi:10.1175/JAM2429.1

Elsberry R L. 2004. Monsoon-related tropical cyclones in East Asia [M]//Chang C P. East Asian Monsoon. Singapore: World Scientific, 463–498.

Gray W M. 1968. Global view of the origin of tropical disturbances and

storms [J]. Mon. Wea. Rev., 96(10): 669−700. doi:10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2

Hill K A, Lackmann G M. 2009. Influence of environmental humidity on tropical cyclone size [J]. Mon. Wea. Rev., 137(10): 3294−3315. doi:10.1175/2009MWR2679.1

Holland G J, Merrill R T. 1984. On the dynamics of tropical cyclone structural changes [J]. Quart. J. Roy. Meteor. Soc., 110(465): 723−745. doi:10.1002/qj.49711046510

Jarvinen B R, Neumann C J. 1979. Statistical forecasts of tropical cyclone intensity for the North Atlantic basin [R]. NOAA Technical Memorandum NWS NHC-10, 22pp.

金龙, 黄小燕, 史旭明. 2008. 复共线性关系对逐步回归预报方程的影响研究 [J]. 气象学报, 66(4): 547−554. Jin Long, Huang Xiaoyan, Shi Xuming. 2008. A study on impact of multicollinearity on stepwise regression prediction equation [J]. Acta Meteor Sinica (in Chinese), 66(4): 547−554. doi:10.3321/j.issn:0577-6619.2008.04.008

Khairoutdinov M, Emanuel K. 2013. Rotating radiative−convective equilibrium simulated by a cloud-resolving model [J]. J. Adv. Model. Earth Syst., 5(4): 816−825. doi:10.1002/2013MS000253

Knaff J A, Slocum C J, Musgrave K D, et al. 2016. Using routinely available information to estimate tropical cyclone wind structure [J]. Mon. Wea. Rev., 144(4): 1233−1247. doi:10.1175/MWR-D-15-0267.1

Knapp K R, Kruk M C, Levinson D H, et al. 2010. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data [J]. Bull. Amer. Meteor. Soc., 91(3): 363−376. doi:10.1175/2009BAMS2755.1

Lee C S, Cheung K K W, Fang W T, et al. 2010. Initial maintenance of tropical cyclone size in the Western North Pacific [J]. Mon. Wea. Rev., 138(8): 3207−3223. doi:10.1175/2010MWR3023.1

Lee C Y, Tippett M K, Camargo S J, et al. 2015. Probabilistic multiple linear regression modeling for tropical cyclone intensity [J]. Mon. Wea. Rev., 143(3): 933−954. doi:10.1175/MWR-D-14-00171.1

李崇银. 1983. 环境流场对台风发生发展的影响 [J]. 气象学报, 41(3): 275−284. Lee Chongyin. 1983. The effect of ambient flow on generation and development of typhoon [J]. Acta Meteor. Sinica (in Chinese), 41(3): 275−284. doi:10.11676/qxxb1983.032

Lin Y L, Zhao M, Zhang M H. 2015. Tropical cyclone rainfall area controlled by relative sea surface temperature [J]. Nat. Commun., 6: 6591. doi:10.1038/ncomms7591

Liu K S, Chan J C L. 2002. Synoptic flow patterns associated with small and large tropical cyclones over the Western North Pacific [J]. Mon. Wea. Rev., 130(8): 2134−2142. doi:10.1175/1520-0493(2002)130<2134:SFPAWS>2.0.CO;2

Maclay K S, DeMaria M, Vonder Haar T H. 2008. Tropical cyclone inner-core kinetic energy evolution [J]. Mon. Wea. Rev., 136(12): 4882−4898. doi:10.1175/2008MWR2268.1

梅双丽, 江静. 2012. 西北太平洋热带气旋迅速增强特征及其影响因子 [J]. 热带气象学报, 28(1): 1−11. Mei Shuangli, Jiang Jing. 2012. Characteristics and influence factors of rapid intensification of tropical cyclones in the Western North Pacific [J]. J. Trop. Meteor. (in Chinese), 28(1): 1−11. doi:10.3969/j.issn.1004-4965.2012.01.001

Miller A. 2002. Subset Selection in Regression (2nd ed.)[M]. Boca Raton: Chapman & Hall, 256.

Neumann C B. 1972. An alternate to the HURRAN (hurricane analog) tropical cyclone forecast system [R]. NOAA Tech. Memo. NWS SR-62, 23pp.

Palmer C K, Barnes G M. 2002. The effect of vertical wind shear as diagnosed by the NCEP/NCAR Reanalysis data on Northeast Pacific hurricane intensity [C]// Preprints, 25th Conf. on Hurricanes and Tropical Meteorology, San Diego, CA: American Meteorological Society, 122–123.

Wang Y Q. 2009. How do outer spiral rainbands affect tropical cyclone structure and intensity? [J]. J. Atmos. Sci., 66(5): 1250−1273. doi:10.1175/2008JAS2737.1

Willmott C J. 1982. Some comments on the evaluation of model performance [J]. Bull. Amer. Meteor. Soc., 63(11): 1309−1313. doi:10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2

吴磊. 2013. 西北太平洋热带气旋尺度特征的研究 [D]. 南京大学硕士学位论文, 1–2. Wu Lei. 2003. A study on the characteristics of tropical cyclone size over the Western North Pacific [D]. M. S. thesis (in Chinese), Nanjing University, 1–2.

Xu J, Wang Y Q. 2010. Sensitivity of the simulated tropical cyclone inner-core size to the initial vortex size [J]. Mon. Wea. Rev., 138(11): 4135−4157. doi:10.1175/2010MWR3335.1

袁金南, 林爱兰, 刘春霞. 2008. 60 年来西北太平洋上不同强度热带气旋的变化特征 [J]. 气象学报, 66(2): 213−223. Yuan Jinnan, Lin Ailan, Liu Chunxia. 2008. Change characters of tropical cyclones with different intensities over the Western North Pacific during the last 60 years [J]. Acta Meteor. Sinica (in Chinese), 66(2): 213−223. doi:10.3321/j.issn:0577-6619.2008.02.008