

基于气象大数据的台风强度预测统计模型研究

汤婷婷^{1,2} 李晴岚² 李广鑫² 彭玉龙²

¹(中国科学技术大学 合肥 230026)

²(中国科学院深圳先进技术研究院 深圳 518055)

摘 要 台风是一种破坏力极强的灾害性天气系统,做好台风路径和强度预报是防灾减灾的关键。除了气候性因子、台风持续性因子以及环境背景场因子,文章还考虑了在近海时,受陆地影响下,台风强度演变的情况,引入了新变量,即海陆比。将 2000—2014 年西北太平洋的所有台风样本分成海盆样本和近海样本,研究它们在 12、24、36 和 48 小时间隔的强度演变规律。本研究利用 $1^{\circ} \times 1^{\circ}$ 美国国家环境预报中心/美国国家大气研究中心提供的 FNL 全球再分析资料(Final Operational Global Analysis)数据,采用逐步回归和主成分分析法的多元统计回归模型预测台风强度,并比较了两种模型在台风强度预测上的表现。综合深海盆和近海台风强度的预测结果可以看出,文章提供的近海台风强度预报方法,比国内外的其他研究更具有防台减灾的实际应用价值。

关键词 热带气旋; 强度; 预报; 逐步回归; 主成分分析

中图分类号 TG 156 **文献标志码** A

Statistical Models for Tropical Cyclone Intensity Forecasting Based on Meteorological Big Data

TANG Tingting^{1,2} LI Qinglan² LI Guangxin² PENG Yulong²

¹(University of Science and Technology of China, Hefei 230026, China)

²(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

Abstract Tropical cyclone (TC) is a destructive weather system. Accurate and timely forecast of the TC's intensity and track is vital for disaster prevention and mitigation. This study proposed statistical regression methods to forecast the TC's intensity change for 12, 24, 36 and 48 hours in the future over the Northwest Pacific Ocean. In addition to the conventional factors of climatology and persistence, this study took into account the land effect on the TC's intensity change by introducing a new factor, i.e. the ratio of sea to land, into the statistical regression models. Three sets of TC samples, ocean basin samples, offshore samples, and total TC samples for the years 2000—2014 were applied to develop the intensity forecasting models. Final operational global analysis proposed by $1^{\circ} \times 1^{\circ}$ National Centers for Environmental Prediction/National Centre for Atmospheric Research

收稿日期: 2015-11-16 修回日期: 2016-01-06

基金项目: 广东省自然科学基金自由申请项目(2015A030313742); 深圳市科技创新委员会重点项目(JCYJ20120617115926138)

作者简介: 汤婷婷, 硕士研究生, 研究方向为计算机应用; 李晴岚(通讯作者), 博士生导师, 研究方向为气候变化、天气预报和台风模拟与影响评估, E-mail: ql.li@siat.ac.cn; 李广鑫, 工程师, 研究方向为计算机应用; 彭玉龙, 硕士研究生, 研究方向为应用数学。

were used as the predictors for the environmental effects. Two methods, stepwise regression and principal component analysis, were employed to develop the TC intensity forecasting models. Due to the consideration of the ratio of sea to land, the intensity forecasting performance for offshore TCs was significantly improved. Therefore, the proposed models are valuable for practical disaster prediction.

Keywords tropical cyclone; intensity; prediction; stepwise regression; principal component analysis

1 引 言

我国是受台风灾害影响较大的国家, 台风给人民的财产安全带来了严重的损失。为了减轻台风带来的灾害程度, 需要深入了解台风的突变机制和台风的内部结构变化, 以准确预报台风的途径、强度等信息。在过去的三十年中, 由于观测技术、探测手段的不断发展, 台风路径预报已经有了稳步提升, 但台风强度预报却进展缓慢^[1], 其原因有很多, 例如: 对影响台风强度变化的物理过程了解不够、模式分辨率不高, 对台风的内部结构变化了解不足等。台风的强度突变一直是国际台风界的研究前沿, 也是台风预报业务的三大难题之一^[2]。现阶段国际上比较通用的预测台风强度的方式是统计降尺度^[3], 它的基本原理在于采用统计经验的方法建立大尺度海量气象变量与台风强度变量之间的线性或非线性联系。统计回归方法有以下优点: ①能以很高的计算效率由大尺度气象要素得到区域尺度的气象要素; ②能输出较高分辨率或站点尺度的气象要素; ③模型参数可以反映区域下垫面的特征影响。因此统计方法预报台风强度得到了广泛应用^[4]。

1979 年, Jarvinen 等^[5]研发了一种台风强度数值预测模型 (SHIFOR), 利用气候因子和持续性因子之间的数值关系, 预测大西洋海面上 72 小时内的台风强度变化。SHIFOR 模型的出现不但为今后的台风强度预测研究奠定了基础, 同时也作为评估热带气旋强度预测的一个基准。1985 年, Pike^[6]在 SHIFOR 模型的基础上, 利用地理

位势高度等因子, 首次尝试将气候影响因子放入数值预测模型, 实验效果较 SHIFOR 模型有所提高, 但成效不大。1987 年, Merrill^[7]在前者的台风数值预测模型中加入了更多的气候影响因子, 用于预测海盆和近海海岸附近的台风强度变化, 但实验效果并不明显。1994 年, DeMaria 等^[8]使用多元线性回归的方法研发 SHIPS 模型, 该模型在 Merrill 研究的基础上, 添加了更多的气候影响因子和海水表面温度来预测大西洋和东北太平洋海盆内的台风强度变化, 台风强度的预测效果有了较为明显的提高。

袁小超等^[9]在“达维”台风强度突变分析中认为水平散度、垂直涡度、温度和垂直风切变对台风强度突变具有重要的影响, 并得出垂直切变和台风强度之间有着比较明显的负相关性。徐焱等^[10]在超强台风桑美 (0608) 强度突变一文中分析到, 台风中心附近强的纬向风净流入能很好地表征台风强度的迅速增强, 而强的纬向风净流出以及高层强冷空气的侵入则表示台风将很快减弱; 此外, 台风中心附近中低层深厚的辐合层的形成以及高层辐散的迅速增强有利于台风强度迅速增强, 而台风中心附近中低层辐散的增强以及高层辐合的向下伸展则将导致台风快速减弱。梅双丽等^[11]提出, 台风前 12 小时强度变化、海表面温度和垂直风切变是影响台风迅速增强的主要因子。

在上述台风强度预测中, 多种环境影响因子之间必然存在一定程度的重叠, 因子过多或者因子间存在的复共线性关系也会影响模型的预报能力, 所以模型建立前需要对多种影响因子进行分

析并选取,降低因子之间的相关性,提高预报的准确率。农吉夫^[12]在《主成分分析在线性模型与非线性模型的应用研究》一文中提出,无论是在线性模型还是非线性模型中使用主成分分析法^[13],实验效果均有显著提升。

以上台风强度预报研究多是针对海盆中的台风,鲜少有对登陆和近海台风强度变化进行分析预报。本文采用 $1^{\circ} \times 1^{\circ}$ 美国国家环境预报中心/美国国家大气研究中心(NCEP/NCAR)提供的FNL全球分析资料(Final Operational Global Analysis,以下称“FNL全球再分析资料”)数据以及中国中央气象台提供的实际台风的观测数据,其中既包括气候因子,也包括持续性因子、环境背景场因子以及下垫面陆地的影响。本研究将使用逐步回归方法筛选显著性预报因子,分别通过逐步回归法和主成分分析法建立统计模型预测台风强度变化,并对两种实验方法的结果进行比较,为今后的台风强度预报提供参考。

2 资料与方法

2.1 资料

本文创建的统计模型主要使用2000—2014年西北太平洋领域的所有台风数据以及NCEP/NCAR提供的FNL全球再分析资料数据^[14]。NCEP的FNL数据产品格式采用了GRIB码编码格式^[15]。NCEP的FNL数据的空间分辨率为 $1^{\circ} \times 1^{\circ}$,时间间隔为6小时(国际时刻0时、06时、12时、18时),数据中包含了地表以及26个标准等压层、地表边界层和对流顶层等信息。NCEP/NCAR再分析资料采用了当今最先进的全球资料同化系统和完善的数据库,对各种资料来源(地面、船舶、无线电探空仪、测风气球、飞机、卫星等)的观测资料进行质量控制和同化处理,获得了一套完整的再分析资料集,它不仅包含的要素多、范围广,而且延伸的时段

长,是一个综合观测的海量数据集。

2.2 方法

本文主要采用多元回归分析方法^[16]建立模型。回归分析是目前气象统计分析中最为常用的一种方法,尤其在气象预报业务中为国内外台站所常用。目前国内外气象台站常用的模式输出统计量(Model Output Statistics, MOS)方法中,回归分析是最基本的方法之一。

回归分析是用来寻找若干变量之间统计关系的一种方法。利用所找到的统计关系对某一变量作出未来时刻的估计,称为预报值。本文将预报西北太平洋海域范围的台风强度,称为预报量。为了预报对象未来时刻的变化,我们选择预报量前期已发生的多个有关的气象要素,简称为预报因子,利用回归分析方法分析多个预报因子与该预报量之间的相互关系,建立统计关系的方程式,最后利用方程式对未来时刻的台风强度作出预报估计。

回归分析是一种统计模型,它包括线性回归和非线性回归。本文是基于线性回归方法来建立台风强度的预测统计模型。

2.2.1 多元线性回归的方法

在气象统计预报中,通常寻找与预报量线性关系很好的单个因子是很困难的,而且实际上某个气象要素的变化是和前期多个因子有关,因而大部分气象统计预报中的回归分析都是用多元回归技术^[17]进行。多元回归是对某一预报量 Y ,研究多个因子与其的定量统计关系。如共选取 k 个环境影响因子,记为 X_1, X_2, \dots, X_k ,则多元线性回归模型为:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k \quad (1)$$

其中, b_0 为常数项; b_1, b_2, \dots, b_k 为回归系数, b_1 为 X_1, X_2, \dots, X_k 固定时, X_1 每增加一个单位对 Y 的效应,即 X_1 对 Y 的偏回归系数;同理 b_2 为 X_2 对 Y 的偏回归系数。

多元线性回归模型的参数估计,类似于一元

线性回归方程, 在误差平方和 $\sum(x_i - \bar{x})^2$ 为最小的前提下, 用最小二乘法求解参数。以二元线性回归模型为例, 求解回归参数的标准方程组为:

$$\begin{cases} \sum y = nb_0 + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum x_1 y = b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \\ \sum x_2 y = b_0 \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \end{cases} \quad (2)$$

解此方程可求得 b_0 、 b_1 和 b_2 的数值。亦可用下列矩阵法求得, 即

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \end{bmatrix} \quad (3)$$

2.2.2 逐步回归

(1) 方法介绍

逐步回归^[18]的基本思想是按偏相关系数的大小次序将自变量逐个引入方程, 对引入方程中每个自变量的偏相关系数进行统计检验, 效应显著的自变量留在回归方程内, 继续遴选下一个自变量。如果效应不显著, 停止引入新自变量。由于新自变量的引入, 原已引入方程中的自变量由于变量之间的相互作用其效应有可能变得不显著者, 经统计检验验证后要随时从方程中剔除, 只保留效应显著的自变量。直至不再引入和剔除自变量为止, 从而得到最优的回归方程。

(2) 计算步骤

逐步回归分析的每一步都要对已引入回归方程的变量计算其偏回归系数的平方和, 然后选一个偏回归系数平方和最小的变量, 在预先给定的水平下进行显著性检验。若显著, 则该变量不必从回归方程中剔除, 这时方程中其他几个变量也都不需要剔除。如果不显著, 则该变量需要剔除, 然后按照偏回归系数的平方和由小到大依次对方程中的其他变量进行检验。将对影响不显著的变量全部剔除。最后再对未引入回归方程中的变量分别计算其偏回归系数的平方和, 并选其中偏回归系数平方和最大的一个变量, 同样在给定水平下进行显著性检验。如果显著则将该变量引

入回归方程, 这一过程一直持续下去, 直到在回归方程中的变量都不能剔除而又无新变量可以引入时为止, 这时逐步回归过程结束。

2.2.3 主成分分析法

(1) 方法介绍

主成分分析也称主分量分析, 是揭示大样本、多变量数据或样本之间内在关系的一种方法, 旨在利用降维的思想, 把多指标转化为少数几个综合指标, 降低观测空间的维数, 以获取最主要的信息。在统计学中, 主成分分析是一种简化数据集的技术。它是一个线性变换。这个变换把数据变换到一个新的坐标系统中, 使得任何数据投影的第一大方差在第一个坐标(称为第一主成分)上, 第二大方差在第二个坐标(第二主成分)上, 依次类推。主成分分析通过保留低阶主成分, 忽略高阶主成分, 减少数据集的维数, 同时保持数据集对方差贡献最大的特征。低阶成分绝大多数情况下代表了数据最重要的方面, 但也不可一概而论, 要视具体应用而定。

主成分分析是气象上多变量分析中常用的方法之一。在气象统计中, 经常要研究分析各种气象要素场, 如海温场、降冰场, 它们大多由不规则的网格点所组成。抽取这些场的某一段历史时期的资料, 构成一组以网格点为空间点随时间变化的样本, 主成分分析就是以这种样本为分析对象的。主成分分析与回归分析、差别分析不同, 它是一种分析方法而不是一种预报方法。它能够把随时间变化的气象要素场分解为空间函数部分和时间函数部分。空间函数部分概括场的地域分布地点, 这部分是不随时间变化的; 而时间函数部分则由空间点的线性组合所构成, 称为主成分, 这些主分量的前几个占有原变量总方差的很大部分。研究主分量随时间变化的规律就可以代替对原有场随时间变化的研究。

(2) 计算步骤

① 原始指标数据的标准化采集 p 维随机向量

$\mathbf{X}=(x_1, x_2, \dots, x_p)^T$, n 个样本 $\mathbf{X}_i=(x_{i1}, x_{i2}, \dots, x_{ip})^T$, $i=1, 2, \dots, n$, 其中 $n > p$ 。构造样本矩阵, 对样本矩阵进行如下标准化变换:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (4)$$

其中, $\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}$; $s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}$; $i=1, 2, \dots, n$; $j=1, 2, \dots, p$, 得到标准化矩阵 \mathbf{Z} 。

② 对标准化矩阵 \mathbf{Z} 求相关系数矩阵 \mathbf{R}

$$\mathbf{R} = [r_{ij}]_p \quad r_{ij} = \frac{\mathbf{Z}^T \mathbf{Z}}{n-1} \quad (5)$$

其中, $r_{ij} = \frac{\sum z_{kj} \cdot z_{ki}}{n-1}$; $i, j=1, 2, \dots, p$ 。

③ 解出相关系数矩阵 \mathbf{R} 的特征方程 $|\mathbf{R} - \lambda \mathbf{I}_p|$

$= 0$, 得 p 个特征根, 确定主成分按 $\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 85\%$

确定 m 值, 使得信息的利用率达到 85% 以上。对每个 λ_j , $j=1, 2, \dots, m$, 解方程组 $\mathbf{R}b = \lambda_j b$ 得单位特征向量 b_j^0 。

④ 将标准化后的指标变量转换为主成分

$$U_{ij} = z_{ij}^T b_j^0 \quad (6)$$

其中, $j=1, 2, \dots, m$; U_1 称为第一主成分, U_2 称为第二主成分, \dots , U_p 称为第 p 主成分。

⑤ 对 m 个主成分进行综合评价

对 m 个主成分进行加权求和, 即得最终评价, 权数为每个主成分的方差贡献率。

3 预测模型的建立

3.1 预报对象及样本划分

本文在进行台风强度预报建模实验时, 以 2000—2014 年西北太平洋海域发生的台风数据为样本: 将 2000—2011 年共 12 年的数据作为训练样本, 2012—2014 年共 3 年的数据作为检验

样本。考虑下垫面陆地对台风的衰减影响^[19], 本文提出一种计算海水陆地比例的方法, 通过比较 800 km 范围内海水与陆地所占面积的比例, 将样本数据划分为全部台风样本、近海台风样本和海盆台风样本, 便于针对不同领域的台风预测模型做更为精准的预测和对比。在预报建模中, 本文台风强度预测模型的预测时效分别为 12 小时、24 小时、36 小时和 48 小时。

3.2 影响因子选取

影响因子的选取对模型的预测效果起着至关重要的作用。初始因子选取过多可能会出现过拟合现象^[20], 因此需要从台风强度演变的物理过程出发, 筛选出合适的预报因子。

通过了解台风强度变化的物理机制, 考虑下垫面情况对台风强度的影响, 采用气候持续法^[21]原理。并且通过逐步回归的方法, 将变量逐个引入模型进行 F 检验, 根据 F 达到 0.05 相关显著水平的标准, 最终选取了 24 个影响因子, 具体如下表 1 所示。

将 24 个台风强度预报因子带入回归模型, 分析测试每个因子的显著性大小, 通过计算 F 值来确定该因子是否引入模型。我们分别以全部、近海及海盆这三个不同地域的台风强度预测模型为例, 表 2、表 3 和表 4 分别为全部、近海及海盆台风预测模型的环境影响变量的选择表。

从表 2、表 3 和表 4 可以看出, 全部、近海及海盆这三个不同地域的台风强度预测模型的显著影响因子有较大的差异, 但也有一些共同点:

(1) 风暴潜势 POT、POT²、温度 T200、LSHR、纬度 LAT、850~700 hpa 相对湿度 RHLO 和海温 SST 是三个模型中最为显著的环境影响因子^[22]。

(2) 除了以上所提到的显著影响因子之外, 海陆比、200~850 hpa 垂直风切变 SHR、水平散度 D200 和 500 hpa 水汽通量对全部、近海台风预测模型影响较大。

表1 环境影响因子变量名称表

Table 1 List of environmental predictors

环境因子名称缩写	中文释义	计算方法或范围
POT	最大潜在风强度	$POT = MPI - VMX$
POT ²	最大潜在风强度的平方项	
SHR	200~850 hpa 垂直风切变	
SHR	500~850 hpa 垂直风切变	
DVMX	12 h 前台风强度变化	
JDAY	儒略日	$\text{Exp}\{-[(J_d - P_d)/R_d]^2\}$
T200	200 hpa 平均温度	$r = 0 \sim 1\ 000\ \text{km}$
U200	200 hpa 纬向风	$r = 0 \sim 1\ 000\ \text{km}$
Z850	850 hpa 垂直涡度	$r = 0 \sim 1\ 000\ \text{km}$
Z500	500 hpa 垂直涡度	$r = 0 \sim 1\ 000\ \text{km}$
LSHR	$LSHR = SHR \times \sin(lat)$	
D200	200 hpa 水平散度	$r = 0 \sim 1\ 000\ \text{km}$
VMX	初始最大风暴	
Pre	台风中心气压	
LAT	纬度	
LON	经度	
SPD	风暴移动速度	$V = \sqrt{u^2 + v^2}$
海陆比	海水与陆地所占面积比例	$r = 100, 200, 300, 400\ \text{km}$
SST	海水温度	
RHLO	850~700 hpa 相对湿度	$r = 200 \sim 800\ \text{km}$
RHHI	500~300 hpa 相对湿度	$r = 200 \sim 800\ \text{km}$
USHRS	500~850 hpa 纬向风切变	$r = 200 \sim 800\ \text{km}$
USHRD	200~850 hpa 纬向风切变	$r = 200 \sim 800\ \text{km}$
水汽通量	850 hpa, 500 hpa 水汽通量	$r = 200 \sim 800\ \text{km}$

注: MPI 为台风最大可能强度; VMX 为初始台风强度, 参考 SHIPS 模型中儒略日因子的计算公式; J_d 是儒略日; P_d 是台风发生高峰期的平均值 (本文取 $P_d = 238$)^[8]; R_d 是控制本期影响的时间尺度 (本文中 $R_d = 25$)

(3) 对于海盆台风预测模型来说, 风暴移动速度 SPD 对其有着重要的影响。

3.3 模型的建立与检验

3.3.1 逐步回归模型

本文研究的不同组别台风样本数情况如表 5 所示, 此处以海盆台风数据样本为例, 对海盆台风强度 12 小时、48 小时预测建立模型, 12 小时海盆台风强度预测模型的训练样本容量 951, 检验样本容量 116; 48 小时海盆台风强度预测模型

表2 西北太平洋全部台风强度样本 12~48 小时预测环境变量选择表

Table 2 Significant predictors for the intensity prediction for all TCs in the Northwest Pacific Ocean for the coming

12-48 h

环境影响变量名称	12 h	24 h	36 h	48 h
POT	√	√	√	√
POT ²	√	√	√	√
200~850 hpa SHR	√	√	√	√
500~850 hpa SHR	√	√		
DVMX12	√	√	√	√
JDAY			√	√
T200	√	√	√	√
U200	√	√	√	
Z850	√	√	√	√
Z500	√	√	√	√
LSHR	√	√	√	√
D200	√	√	√	√
VMX	√	√	√	√
Pre				
LAT	√	√	√	√
LON	√	√	√	√
SPD	√	√	√	√
海陆比	√	√	√	√
海温度 SST	√	√	√	√
RHLO	√	√	√	√
RHHI	√	√	√	√
USHRS	√	√	√	√
USHRD			√	√
水汽通量	√	√	√	√

的训练样本容量 813, 检验样本容量 98。因对每个环境影响因子的研究范围不同, 故得到的预测方程如下所示。其中, x_i 为环境影响因子变量; Y 为台风强度变化值。

$$Y_{12} = 0.564 + 1.728x_1 - 1.053x_2 - 0.924x_5 + 0.190x_9 + 0.143x_{10} - 0.193x_{14} - 0.125x_{19} + 0.077x_{21} + 0.101x_{25} + 0.935x_{32} - 0.285x_{42} - 0.284x_{43} - 0.058x_{51} + 0.077x_{53} - 0.460x_{54} - 0.066x_{55} + 0.106x_{66} + 0.057x_{67} - 0.051x_{81} - 0.140x_{95}$$

表 3 西北太平洋近海台风强度样本 12~48 小时预测环境
变量选择表

Table 3 Significant predictors for the intensity prediction
for Offshore TCs in the Northwest Pacific Ocean for the
coming 12-48 h

环境影响变量名称	12 h	24 h	36 h	48 h
POT	√	√	√	√
POT ²	√	√	√	√
200~850 hpa SHR	√	√	√	√
500~850 hpa SHR	√	√		
DVMX12	√	√	√	√
JDAY	√	√	√	√
T200	√	√	√	√
U200		√	√	
Z850	√	√		√
Z500	√	√	√	√
LSHR	√	√	√	√
D200		√	√	
VMX			√	√
Pre				
LAT	√	√	√	√
LON		√	√	√
SPD		√	√	√
海陆比	√		√	√
海温度 SST	√	√	√	√
RHLO	√	√	√	√
RHHI	√	√	√	√
USHRS	√		√	√
USHRD		√	√	√
水汽通量	√	√	√	√

表 4 西北太平洋海盆台风强度样本 12~48 小时预测环境
变量选择表

Table 4 Significant predictors for the intensity prediction
for Basin TCs in the Northwest Pacific Ocean for the coming
12-48 h

环境影响变量名称	12 h	24 h	36 h	48 h
POT	√	√	√	√
POT ²	√	√	√	√
200~850 hpa SHR	√			
500~850 hpa SHR		√	√	√
DVMX12	√	√		
JDAY				√
T200	√	√	√	√
U200	√	√	√	√
Z850	√	√	√	√
Z500		√	√	√
LSHR	√	√	√	√
D200		√	√	
VMX		√	√	√
Pre	√			
LAT	√	√	√	√
LON				
SPD	√	√	√	√
海陆比	√	√	√	√
海温度 SST	√	√	√	√
RHLO	√	√	√	√
RHHI		√	√	√
USHRS		√	√	
USHRD		√	√	√
水汽通量		√	√	√

表 5 台风模型的样本数量比较

Table 5 Comparison of the samples number for different TC groups

预测时间	全部台风样本		近海台风样本		海盆台风样本	
	训练样本	检验样本	训练样本	检验样本	训练样本	检验样本
12 h	5 870	1 082	4 919	966	951	116
24 h	5 343	949	4 438	841	905	108
36 h	4 829	824	3 968	722	861	102
48 h	4 336	712	3 523	614	813	98

$$\begin{aligned}
Y_{48} = & -0.052 + 2.898x_1 - 1.145x_2 - 0.073x_7 + \\
& 1.600x_{12} - 2.975x_{13} + 1.333x_{14} + 0.243x_{15} - \\
& 0.816x_{17} + 0.186x_{22} + 0.388x_{28} - 0.361x_{29} + \\
& 0.218x_{32} + 0.789x_{41} - 0.230x_{43} + 0.183x_{45} - \\
& 0.533x_{47} + 0.107x_{50} - 0.129x_{51} - 1.051x_{54} - \\
& 0.126x_{56} + 0.134x_{66} + 0.154x_{69} + 0.326x_{73} - \\
& 0.626x_{74} + 0.380x_{78} - 0.221x_{79} + 0.362x_{80} - \\
& 0.278x_{81} + 0.430x_{87} + 0.439x_{88} - 0.248x_{91} - \\
& 0.307x_{95} + 0.033x_{96}
\end{aligned}$$

3.3.2 主成分法线性回归模型

通过主成分分析法处理后的因子代入到线性回归模型中, 同样以海盆台风数据样本为例, 对海盆台风强度 12 小时、48 小时预测建立模型, 得到主成分法线性回归模型方程为:

$$\begin{aligned}
Y_{12} = & 0.131 + 0.016z_1 + 0.103z_2 - 0.017z_3 + \\
& 0.055z_7 - 0.070z_8 + 0.063z_9 - 0.040z_{11} - \\
& 0.086z_{15} + 0.104z_{20} + 0.242z_{21} - 0.070z_{22} - \\
& 0.085z_{24} + 0.089z_{25} + 0.059z_{26} - 0.088z_{29} + \\
& 0.088z_{30} + 0.090z_{35} + 0.090z_{36} + 0.146z_{39} - \\
& 0.242z_{40} + 0.174z_{41} - 0.212z_{42} + 0.171z_{45} + \\
& 0.185z_{46} - 0.565z_{48} + 0.331z_{50} - 0.442z_{51} - \\
& 0.549z_{55} - 0.319z_{56} + 0.319z_{59} + 0.493z_{67} + \\
& 0.396z_{69} + 0.501z_{70} + 0.815z_{71} + 0.606z_{78} + \\
& 0.532z_{80} + 1.752z_{81} + 0.816z_{82} - 1.669z_{84} + \\
& 0.822z_{85} + 2.567z_{95} \\
Y_{48} = & 0.189 - 0.028z_1 + 0.185z_2 + 0.022z_3 + 0.032z_4 + \\
& 0.048z_5 - 0.024z_6 + 0.101z_7 + 0.175z_8 - \\
& 0.038z_9 - 0.061z_{10} - 0.023z_{11} - 0.045z_{12} - \\
& 0.031z_{13} - 0.049z_{15} - 0.047z_{17} + 0.113z_{21} + \\
& 0.143z_{22} + 0.177z_{24} + 0.097z_{25} + 0.095z_{26} - \\
& 0.141z_{27} - 0.112z_{28} - 0.090z_{30} - 0.177z_{34} + \\
& 0.169z_{35} + 0.116z_{36} - 0.248z_{38} - 0.289z_{40} - \\
& 0.524z_{41} + 0.209z_{46} + 0.516z_{48} - 0.329z_{49} - \\
& 0.405z_{50} - 0.180z_{51} + 0.493z_{52} + 0.214z_{53} - \\
& 1.032z_{54} + 0.237z_{55} - 0.667z_{57} - 0.729z_{61} + \\
& 0.657z_{65} + 0.382z_{68} - 1.152z_{70} - 0.423z_{73} + \\
& 0.723z_{74} + 1.438z_{79} - 1.319z_{81} - 2.064z_{83} + \\
& 1.478z_{86} + 1.233z_{88}
\end{aligned}$$

其中, 以上各方程中, Y_{12} 和 Y_{48} 分别代表 12 小时和 48 小时的台风强度变化; z_i 是各预报方程通过主成分分析提取的主成分变量。

4 预测模型的结果分析与评价

本文对线性回归模型的评价主要分为两方面: 一是对训练样本的拟合情况进行检验, 如果拟合情况较好, 说明该模型对于训练样本已经掌握了隐含的内在规律, 可以较好地反映数据之间的关系; 二是对检验样本的预测情况做检验, 这也是检测预测模型是否具有泛化能力的关键。因此, 分别用拟合样本和检验样本的各一个台风例子: 2009 年的“芭玛”和 2012 年“韦森特”台风, 来评价模型的强度预报情况, 图 1 和图 2 展示的是提前 12 小时和提前 24 小时预报“芭玛”台风强度情况; 图 3、图 4 展示的是提前 12 小时和提前 24 小时预报“韦森特”台风强度情况。

由图 1~图 4, 从实际台风强度模拟的两个例子可以看出, 两种预报模型都能很好地把握台风强度的演变过程, 提前 12 小时比提前 24 小时的强度预报误差要小; 一般说来, 12 小时提前预测台风强度, 逐步回归方法比主成分预报方法性能要好。

更进一步, 从统计角度用参数方法对预报模型性能进行评判, 采用的三个评判系数分别为确定性系数 (R^2)、平均绝对误差 (MAE) 和均方根误差 (RMSE), 误差单位为 m/s。三种台风样本模型对比如表 6~表 8 所示, 其中模型 A 代表多元线性回归模型, 模型 B 代表主成分回归模型。

通过对比分析表 6~表 8 数据, 可以得出:

(1) 从整体来说, 模型 B (主成分线性回归模型) 的训练样本拟合程度较好, 但是检验样本的预测效果并不理想, 较模型 A (多元线性回归模型) 的预测结果差。

这说明主成分分析法并没有因消除预报因子

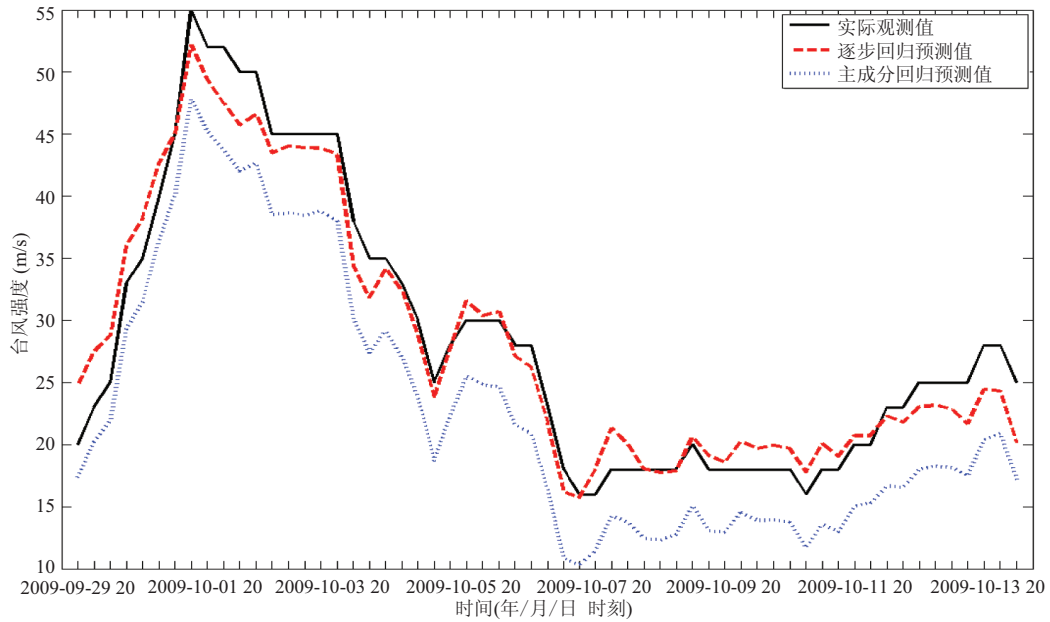


图 1 提前 12 小时的台风强度预测图 (2009 年 10 月芭玛)

Fig. 1 TC intensity forecast for Typhoon Parma (October 2009) 12 hours in advance

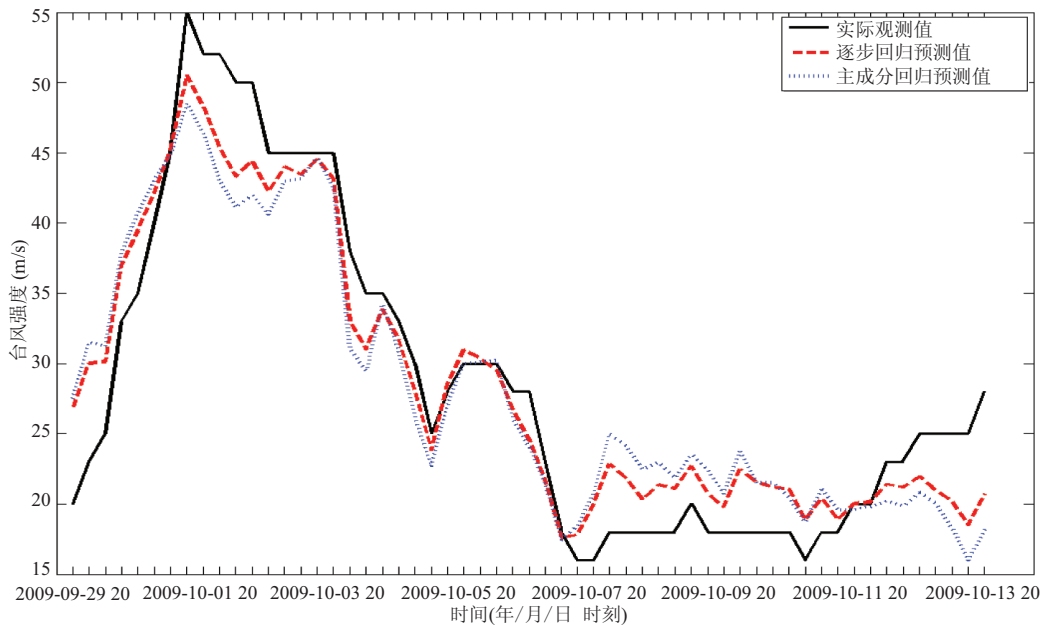


图 2 提前 24 小时的台风强度预测图 (2009 年 10 月芭玛)

Fig. 2 TC intensity forecast for Typhoon Parma (October 2009) 24 hours in advance

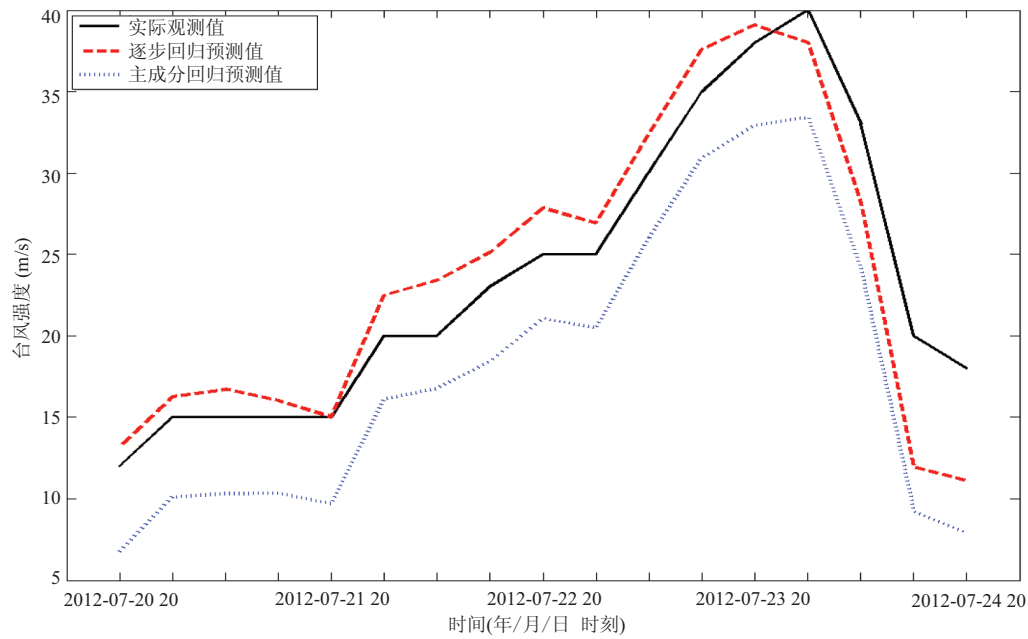


图 3 提前 12 小时的台风强度预测图(2012 年 7 月韦森特)

Fig. 3 TC intensity forecast for Typhoon Vicente (July 2012) 12 hours in advance

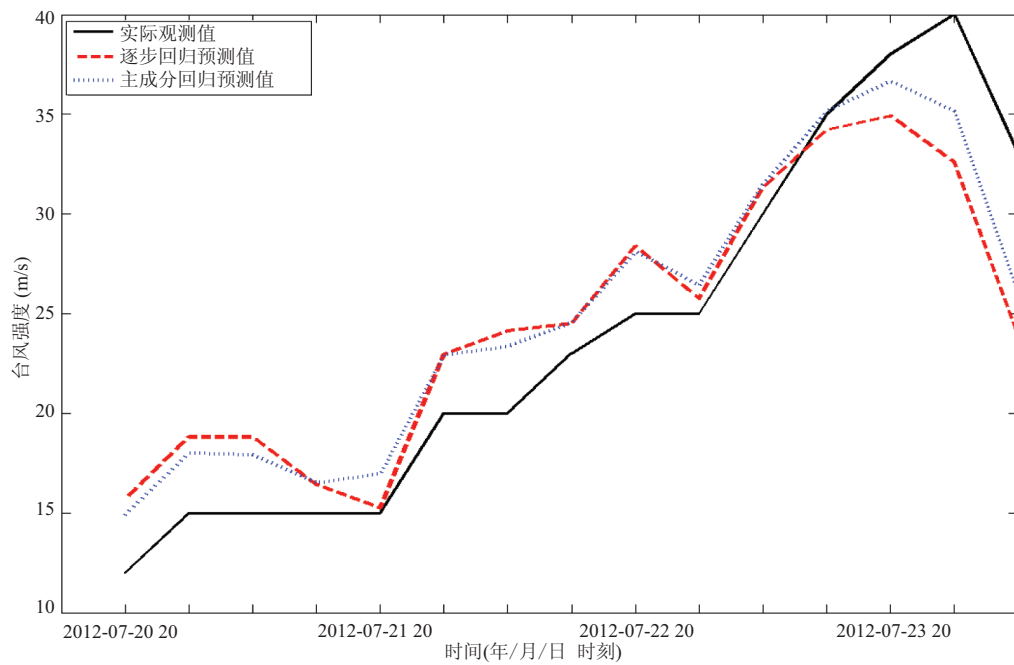


图 4 提前 24 小时的台风强度预测图(2012 年 7 月韦森特)

Fig. 4 TC intensity forecast for Typhoon Vicente (July 2012) 24 hours in advance

表 6 全部台风样本模型的训练样本和检验样本的参数比较

Table 6 Comparison between the calibration and validation of the intensity forecast for total TC samples

预测 时间	模型 类型	训练样本			检验样本		
		R^2	MAE	RMSE	R^2	MAE	RMSE
12 h	模型 A	0.47	2.73	3.70	0.49	2.99	4.16
	模型 B	0.47	1.74	2.35	0.49	6.14	7.08
24 h	模型 A	0.57	4.25	5.55	0.57	4.70	6.37
	模型 B	0.57	3.16	4.12	0.57	4.80	6.74
36 h	模型 A	0.63	5.31	6.83	0.61	6.04	8.07
	模型 B	0.63	4.94	6.33	0.61	6.20	8.22
48 h	模型 A	0.66	6.07	7.72	0.64	6.90	8.90
	模型 B	0.66	5.45	6.93	0.65	7.43	9.45

表 7 近海台风样本模型的训练样本和检验样本的参数比较

Table 7 Comparison between the calibration and validation of the intensity forecast for offshore TC samples

预测 时间	模型 类型	训练样本			检验样本		
		R^2	MAE	RMSE	R^2	MAE	RMSE
12 h	模型 A	0.46	2.75	3.74	0.48	3.09	4.31
	模型 B	0.47	1.74	2.37	0.48	6.23	7.20
24 h	模型 A	0.57	4.26	5.55	0.55	4.95	6.68
	模型 B	0.58	3.16	4.12	0.55	5.01	6.95
36 h	模型 A	0.63	5.28	6.75	0.59	6.38	8.39
	模型 B	0.64	4.90	6.26	0.59	6.54	8.53
48 h	模型 A	0.66	6.05	7.65	0.64	7.22	9.33
	模型 B	0.67	5.41	6.81	0.63	7.64	9.63

表 8 海盆台风样本模型的训练样本和检验样本的参数比较

Table 8 Comparison between the calibration and validation of the intensity forecast for ocean basin TC samples

预测 时间	模型 类型	训练样本			检验样本		
		R^2	MAE	RMSE	R^2	MAE	RMSE
12 h	模型 A	0.54	2.53	3.28	0.64	2.16	2.79
	模型 B	0.57	1.57	2.03	0.48	2.53	3.21
24 h	模型 A	0.69	3.60	4.57	0.69	3.93	4.79
	模型 B	0.70	2.56	3.30	0.63	4.13	5.30
36 h	模型 A	0.74	4.42	5.60	0.68	5.16	6.54
	模型 B	0.75	4.05	5.11	0.64	5.28	6.55
48 h	模型 A	0.75	5.08	6.49	0.67	5.04	6.71
	模型 B	0.77	4.82	6.16	0.61	6.20	7.89

之间的内在重叠性而取得较好的预测效果, 虽然主成分线性回归模型的训练样本拟合程度较好, 但是它的实际预测结果并不乐观。

(2) 对于全部、近海及海盆台风样本三种模型来说, 海盆台风强度预测模型的预测效果最好。这可能由于海盆领域的台风受外界环境干扰较小, 台风生存环境较为单一, 故海盆领域台风强度预测效果明显好于其他两种情况下的预测模型。

(3) 三个预测模型的预测误差都随着预测时效增加(12 小时、24 小时、36 小时、48 小时)而逐渐增大。

(4) 对比国内外其他台风强度预测模型, 在海盆台风强度的预测上, 本文的预测结果和国外其他模型的预测结果相当, 12 小时海盆台风强度预测平均绝对误差为 2.16 m/s, 24 小时海盆台风强度预测平均绝对误差为 3.93 m/s, 36 小时海盆台风强度预测平均绝对误差为 5.16 m/s, 48 小时海盆台风强度预测平均绝对误差为 5.04 m/s。但是, 本文充分考虑了陆地对近海台风的衰减作用, 因此对于近海台风强度的预测, 以及对于近海和登陆台风带来的灾害防御, 本文提出的模型有非常重要的实用价值。

5 结 论

本文利用 $1^\circ \times 1^\circ$ NCEP/NCAR 提供的 FNL 全球再分析资料数据以及中国中央气象台提供的实际台风的观测数据, 考虑到台风近海时下垫面陆地对台风的衰减作用, 将 2000—2014 年的台风样本分成全部台风、近海台风、海盆台风三个数据集, 用多元线性回归方法筛选西北太平洋台风强度变化的预报因子, 应用逐步回归方法和主成分分析法建立统计模型。通过比较两种模型在西北太平洋热带气旋强度预报中的表现, 我们可以看出, 两种方法都能够有效地预测西北太平

洋台风的强度变化。虽然主成分回归模型的训练样本拟合程度较好,但是逐步回归模型的预测结果更稳定,因此在实际预测中,我们推荐使用逐步回归模型方法作为台风强度变化的预报方法。因本文考虑了陆地对近海台风强度的影响,可以为近海和登陆台风的强度演变提供更精准的预测,对于近海和登陆台风的灾害预防具有重要的指导意义。

参 考 文 献

- [1] DeMaria M, Sampson CR, Knaff JA, et al. Is tropical cyclone intensity guidance improving? [J]. American Meteorological Society, 2014, 3: 387-398.
- [2] 陈联寿, 孟智勇. 我国热带气旋研究十年进展 [J]. 大气科学, 2001, 25(3): 420-432.
- [3] 范丽军, 符淙斌, 陈德亮. 统计降尺度法对未来区域气候变化情景预估的研究进展 [J]. 地球科学进展, 2005, 20(3): 320-329.
- [4] 李湘阁, 胡凝. 实用气象统计方法 [M]. 北京: 气象出版社, 2015.
- [5] Jarvinen BR, Neumann CJ. Statistical forecasts of tropical cyclone intensity for the North Atlantic basin [R]. Miami: NOAA Technical Memorandum NWS National Hurricane Center-10, 1979.
- [6] Pike AC. Geopotential heights and thicknesses as predictors of Atlantic tropical cyclone motion and intensity [J]. Monthly Weather Review, 1985, 113: 931-939.
- [7] Merrill RT. An experiment in statistical prediction of tropical cyclone intensity change [R]. Miami: NOAA Technical Memorandum NWS National Hurricane Center - 34, 1987.
- [8] DeMaria M, Kaplan J. A statistic hurricane intensity prediction scheme (SHIPS) for the Atlantic basin [J]. Weather and Forecasting, 1994, 9: 209-220.
- [9] 袁小超, 谭季青. “达维”台风强度突变的诊断分析 [J]. 科技通报, 2011, 27(4): 489-502.
- [10] 徐燧, 王东法, 郑沛群. 超强台风桑美(0608)强度突变的预报因子分析 [C]// 第十五届全国热带气旋科学讨论会, 2009.
- [11] 梅双丽, 江静. 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, 2012, 28(1): 1-11.
- [12] 农吉夫. 主成分分析在线性模型与非线性模型的应用研究 [J]. 广西民族大学学报, 2012, 18(4): 30-34.
- [13] 刘永和, 郭维栋, 冯锦明, 等. 气象资料的统计降尺度方法综述 [J]. 地球科学进展, 2011, 26(8): 837-847.
- [14] 邓小花, 翟盘茂, 袁春红. 国外几套再分析资料的对比与分析 [J]. 气象科技, 2010, 38(1): 1-8.
- [15] 方庆文, 潘永地, 郑锋, 等. 利用 GRADS 读取 NCEP 再分析资料格点数据 [J]. 贵州气象, 2004, 28(5): 34-36.
- [16] 黄嘉佑. 气象统计分析与预报方法 [M]. 北京: 气象出版社, 2004.
- [17] DeMaria M, Kaplan J. Further improvements to the statistical hurricane intensity prediction scheme (SHIPS) [J]. Weather Forecasting, 2004, 20(4): 531-543.
- [18] Knaff J, Sampson CR, DeMaria M. An operational statistical typhoon intensity prediction scheme for the Western North Pacific [J]. Weather and Forecasting, 2005, 20(4): 688-699.
- [19] Fitzpatrick PJ. Understanding and forecasting tropical cyclone intensity change with the typhoon intensity prediction scheme (TIPS) [J]. Weather and Forecasting, 1997, 12(4): 826-846.
- [20] Wang B, Zhou X. Climate variation and prediction of rapid intensification in the western North Pacific [J]. Meteorology Atmospheric Physics, 2008, 99(1): 1-16.
- [21] DeMaria M, Kaplan J. An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and Eastern North Pacific basin [J]. Weather and Forecasting, 1999, 14(3): 326-337.
- [22] 孔宁谦, 陈润珍. 用统计动力方法作盛夏南海中北部热带气旋强度预报 [J]. 广西气象, 2006, 27(1): 4-5.