

陈璇, 游小宝, 郑崇伟, 等. 2019. 一种新的线性回归模型及其应用示例 [J]. 大气科学, 43 (2): 389–400. Chen Xuan, You Xiaobao, Zheng Chongwei, et al. 2019. A new linear regression model and its application [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 43 (2): 389–400, doi:10.3878/j.issn.1006-9895.1808.18108.

# 一种新的线性回归模型及其应用示例

陈璇<sup>1, 2, 3</sup> 游小宝<sup>2</sup> 郑崇伟<sup>1, 2, 4</sup> 孙威<sup>5</sup> 谢胜浪<sup>6</sup>

1 华东师范大学河口海岸学国家重点实验室, 上海 200062

2 中国科学院大气物理研究所大气科学和地球流体力学数值模拟国家重点实验室 (LASG), 北京 100029

3 解放军 75839 部队, 广州 510510

4 海军大连舰艇学院, 辽宁省大连 116018

5 陆军工程大学, 南京 210007

6 解放军 95080 部队, 广东省汕头 515049

**摘 要** 回归分析是统计分析中常用的方法之一。传统的回归模型不具备全域分析能力, 而变量场之间的关系多采用 SVD (Singular Value Decomposition) 进行分析, 与传统的回归分析有所脱节。更为广义的线性回归模型是传统线性回归模型的延拓, 在标量情况下, 该模型可转化为传统线性回归模型。该模型的基本特征包含乘法不可互易性、等价于传统线性回归 (因子项为标量时)、可分析性、延拓性、降维特征及容错性等。该模型解决了传统的线性回归模型不具备全域分析能力及模型表达能力受限于模型维数的现实问题。本文采用了 NCEP (National Centers for Environmental Prediction) 降水、高度场、风场月平均资料及国家气候中心西太平洋副热带高压指数资料, 利用该模型和传统回归方案进行对比分析, 分析结果表明, 该模型具有一定的实用参考价值。

**关键词** 线性回归 统计 更为广义

**文章编号** 1006-9895(2019)02-0389-12

**中图分类号** P456

**文献标识码** A

**doi:**10.3878/j.issn.1006-9895.1808.18108

## A New Linear Regression Model and Its Application

CHEN Xuan<sup>1, 2, 3</sup>, YOU Xiaobao<sup>2</sup>, ZHENG Chongwei<sup>1, 2, 4</sup>, SUN Wei<sup>5</sup>, and XIE Shenglang<sup>6</sup>

1 State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai 200062

2 State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029

3 The 75839 Army of the PLA, Guangzhou 510510

4 Dalian Naval Academy, Dalian, Liaoning, 116018

5 Army Engineering University of PLA, Nanjing 210007

6 The 95080 Army of the PLA, Shantou, Guangdong, 515049

**Abstract** Regression analysis is one of the commonly used methods in statistical analysis. However, traditional regression models have less ability for global analysis, and the relationship between variables is often analyzed by methods like the SVD (Singular Value Decomposition), which lack connections with traditional regression analysis. A

**收稿日期** 2018-01-15; **网络预出版日期** 2018-08-22

**作者简介** 陈璇, 男, 1989 年出生, 主要从事海洋数值模拟和统计方向研究。E-mail: chzffx@qq.com

**通讯作者** 郑崇伟, E-mail: chinaoceanzcw@sina.cn

**资助项目** 河口海岸学国家重点实验室开放基金 SKLEC-KF201707, 国家自然科学基金项目 41490642、51709243

**Funded by** State Key Laboratory of Estuarine and Coastal Research (Grant SKLEC-KF201707), National Natural Science Foundation of China (Grants 41490642, 51709243)

MGLRM (more generalized linear regression model) is a continuation of traditional linear regression model. In the case that both the predictand and the predictors are scalars, the MGLRM can be transformed into the traditional linear regression model. The MGLRM's basic features include non-commutative multiplication, equivalence to traditional linear regression as predictors in the model are scalars, analysis, extension, dimension-reduction, and robustness, etc. The MGLRM solves problems in traditional linear regression models that have less ability for global analysis and limited expressive ability due to the dimensions of the regression equation. In this paper, the MGLRM and the traditional regression model are applied for statistical analysis of monthly average data of precipitation, height, and wind fields from the NCEP (National Centers for Environmental Prediction) and the western Pacific subtropical high index data from the National Climate Center. Comparison of the results show that the MGLRM has practical implications.

**Keywords** Regression, Statistics, More generalized

## 1 引言

统计分析是气象分析预报的基础方法之一,对于数值预报释用、海洋能源开发(郑崇伟等, 2017; 郑崇伟, 2018a, 2018b, 2018c)等均有着重要的意义。分析两个量(场)之间的关系是统计分析的一个重要课题。对于气候分析和预测来说,如何有效的分析两个量,尤其是两个场之间的相关关系尤其重要。在针对两个场的相关性分析中多使用典型相关分析、SVD 等方案,这些方案往往将资料按时空序列进行排列,分析完成后还原。

这些统计方案的限制往往较为显著,例如在 SVD 的讨论中,其限制较为明显:要求两个变量场互为正交变换,两场之一的协方差矩阵为单位矩阵,此时的 SVD 分解才为 SVD 模态,否则只是近似耦合模态(魏凤英, 2007);只有同一对 SVD 模态的时间系数序列存在相关,否则不相关(吴洪宝和吴蕾, 2010)。此类分析方案所得出的模态具有重要的分析价值,然而其所受时间序列长度的影响也较为突出,尤其是时间序列的尾段。

统计方案无法对一个天气过程或影响要素进行系统性分析,基于动力—统计相结合的方案多采用降尺度手段,该方案主要采用包括主成分分析、经验正交函数分解等具有提取主分量方案的统计方法,此类方法保留了主要信息,舍弃了次要信息以简化模型,但所舍弃的次要信息究竟有无影响尚存争议。但统计方案依然有着显著的优势:相关研究(陈璇等, 2017)表明自回归分析对于简单信号(谐波信号、周期性有限正切信号、线性信号等以及任意多不同的谐波信号叠加)来说具有非常好的描述性,可以保证不收敛且自回归的回归系数不唯一(可以证明);夹杂误差的理想信号对于回归而言影响并不是很大。

传统的回归处理方式侧重于统计理论和对随机误差的处理过程,而不具备将动力特征引入。例如,我们在气候分析过程中引入各种指数,但指数仅是对于要素场分布的一种侧面描述,同一指数下,要素场分布可能具有不同的分布特征(袁媛和李崇银, 2009),而这种特征对于气候形态来讲影响巨大,但这一点则无法在传统的统计分析中引入。

据此,陈璇等(2017)提出了一个更为广义的线性回归模型,该模型具有乘法不可交换性、等价于常规线性回归、可分析性、延拓性、降维特征及容错性等六个特点。该方案将所考虑的因子作为要素场,以副热带高压(简称副高)和 500 hPa 高度场为例,该方案将 500 hPa 高度场作为因子以替代副高各个指数代入回归方程。从动力角度上来说,500 hPa 高度场可以更为完善地表征副高各个特点;从统计意义上而言,该方案基于最小二乘,具有全局最优特性。

综上,本文将在第二节中介绍这种新的线性回归模型,并在第三节中给出应用示例,以期能为相关统计分析提供新的思路。

## 2 方法简介

以两个变量为例,假设,  $\mathbf{P}$  表示预报量,  $\mathbf{H}$ 、 $\mathbf{W}$  表示预报因子,如果预报量和预报因子间存在较好的线性关系时,则可以构建场回归方案:

$$\mathbf{P}' = \mathbf{L}_1(\mathbf{A}\mathbf{H}\mathbf{B}) + \mathbf{L}_2(\mathbf{C}\mathbf{W}\mathbf{D}) + \mathbf{E}, \quad (1)$$

其中,  $\mathbf{P}'$  表示预报量( $\mathbf{P}$ )的估计量;  $\mathbf{A}$ 、 $\mathbf{B}$ 、 $\mathbf{C}$ 、 $\mathbf{D}$ 、 $\mathbf{E}$  为回归方程系数矩阵,可由最小二乘法得出;  $\mathbf{L}_1$  及  $\mathbf{L}_2$  为同一线性规则,可以由最小二乘法得出。对于单一预报因子的场回归方案,以  $\mathbf{H}$  为例,式(1)将写为:  $\mathbf{P}' = \mathbf{L}_1(\mathbf{A}\mathbf{H}\mathbf{B}) + \mathbf{E}$ 。此时,依据陈璇等(2017)的结论,我们可以舍弃  $\mathbf{L}_1$  项,即:  $\mathbf{P}' = \mathbf{A}\mathbf{H}\mathbf{B} + \mathbf{E}$ 。

此类方案不要求预报因子与预报量场之间有相同的维数，从式（1）可以看出，该方案具有乘法不可交换性；当将预报量与预报因子均化为标量时，式（1）则变为  $p' = L_1(ahb) + L_2(cwd) + e$ ，式中小写斜体为标量，由最小二乘规则和标量式的乘法规则可化，该式可化为

$$p' = a'h + b'w + e'. \quad (2)$$

式（2）即一般线性回归。对比式（2）和式（1）可以发现，式（1）具有可分析性，即可以分析因变量场（如  $H$ ）的任意变化对预报量场的影响；由于式（1）具有乘法不可交换性，从而导致式（1）具有不同的表达形式，即延拓性；由于式（1）将变量场直接作为因子考虑，从而避免了时空展开式较高的时空维数所造成的计算困难；由于计算所得的系数是矩阵，单点或小区域的误差被扩散到一个较小的值分布，从而使式（1）具有鲁棒性特征。

两个时间序列的相关系数计算表达式为

$$CC = \frac{\sum (x_t - \bar{x}_t)(y_t - \bar{y}_t)}{\sqrt{\sum (x_t - \bar{x}_t)^2 \sum (y_t - \bar{y}_t)^2}}, \quad (3)$$

式中，CC 表示时间序列  $y$  与  $x$  的相关系数， $t$  表示时间，“ $-$ ”表示均值。对比一元线性回归表达式  $y = ax + b$ ，将其代入式（3），不难得出  $y$  与  $x$  的相关系数表达式：

$$CC = \frac{\sum (y'_t - \bar{y}'_t)(y_t - \bar{y}_t)}{\sqrt{\sum (y'_t - \bar{y}'_t)^2 \sum (y_t - \bar{y}_t)^2}}. \quad (4)$$

据式（4），依据  $P' = AHB + E$ ，可得两个场  $P$ 、 $H$  的相关性表达式：

$$CC_{i,j} = \frac{\sum (p'_{i,j,t} - \bar{p}'_{i,j,t})(p_{i,j,t} - \bar{p}_{i,j,t})}{\sqrt{\sum (p'_{i,j,t} - \bar{p}'_{i,j,t})^2 \sum (p_{i,j,t} - \bar{p}_{i,j,t})^2}}, \quad (5)$$

式中，下标  $i, j$  表示场中各个点位的指示量。式（5）可以表征两个变量场之间相关关系，其对应检验可以采用  $t$  检验方案。

依据式（1），还可以计算两个量场  $H$ 、 $W$  对于  $P$  的贡献率：

$$c_i = \frac{|L_i|}{|L_1| + |L_2|}, \quad (6)$$

式中， $i=1, 2$ ， $c_i$  表示  $H$ （ $i=1$ ）、 $W$ （ $i=2$ ）对于  $P$  的贡献率，此时，计算  $L_1$  及  $L_2$  时，式（1）中的相关项均采用标准化处理。相较于其他贡献率计算方案而言，式（6）的贡献率在统计意义上较为明确。

在此，我们给出场回归系数估计方法。

式（1）的最初形式是一个矩阵方程，可以通过 Sylvester 方程（ $AX + XB = C$ ，式中， $A$ 、 $B$ 、 $C$  为已知矩阵， $X$  为待求矩阵）估计矩阵方程里的系数。

对于矩阵形式的回归（场回归）而言，回归的总误差可以以观测  $P$  与预测  $P'$  之差平方的迹来表示： $\text{trace}[(P' - P)(P' - P)^T] + \text{trace}[(P' - P)^T(P' - P)]$ 。其中， $\text{trace}()$  为迹算子，求取矩阵的对角线上元素的累加和，上标  $T$  为转置算子。

这里，可以通过三步估计本文的回归形式（ $P' = AHB + E$ ）中的系数矩阵：第一步，先行估算回归形式（ $P' = AH$ ）中的回归系数矩阵  $A$ ；第二步，再估算回归形式（ $P' = AHB$ ）中的回归系数矩阵  $B$ ；最后，计算回归形式（ $P' = AHB + E$ ）中的  $E$ 。

通过矩阵求导法则（Chen, 2018），对于回归形式（ $P' = AH$  或者  $P' = HB$ ）来说，对其总误差关于回归系数矩阵求导而言，可以发现所得到的方程为 Sylvester 方程。据此，则可以很容易地估算出回归方程中回归系数矩阵的各项大小。

### 3 资料数据及研究区域

中长期降水预测是汛期预测的重要内容，同时也是气候预测关注的重点，本文的主要研究区域和内容为中国陆地区域降水，如图 1 所示，图中绿色粗实线所标记区域中我国陆地区域为主要研究区域。

由于短期气候预测问题既可能受初值影响，又可能受外部强迫影响，故而常见的降水预测研究方案或基于动力框架（王会军等，2007），或对中长期降水的相关影响因子有深入的研究基础：陈丽娟等（2003）采用 500 hPa 高度场距平作为降尺度模型的因子项进行月尺度降水预报回报试验取得了较好的效果，黄玮等（2013）也将类似方法及 500 hPa 高度场资料应用于云南区域延伸期降水预报，魏凤英和黄嘉佑（2010）基于降尺度技术、偏最小二乘回归和 500 hPa 高度场资料的研究表明此类方案和因子可以应用在我国东部夏季降水预测中。我国夏季降水的相关研究（黄嘉佑等，2004）表明，影响我国夏季降水的要素是多方面的，其中副高是一个重要因子，而副高的进退与我国雨带的分布密切相关（俞亚勋等，2013）。

南海夏季风伴随有充沛的水汽，是夏季降水中不可忽视的重要因子。吴尚森等（2003）分析了南

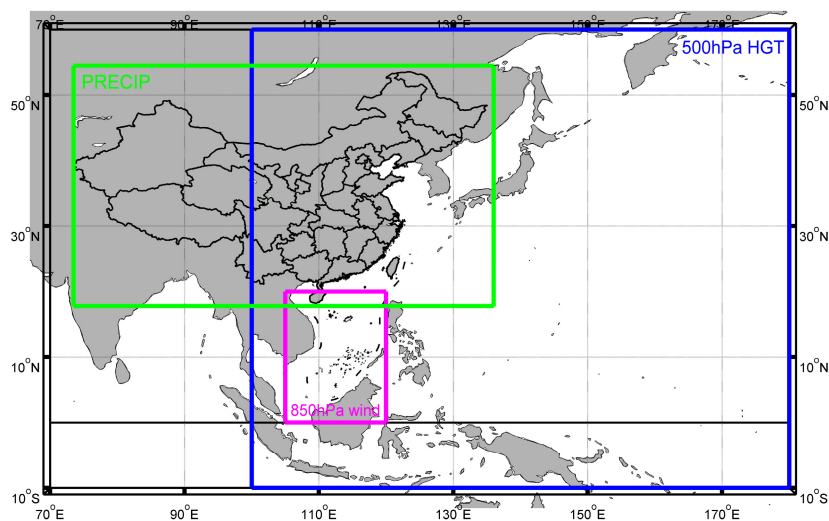


图1 本文降水的研究区域(绿色框区, PRECIP)及主要影响因子(500 hPa 位势高度场、850 hPa 风场)的区域分布。500 hPa HGT 表示 500 hPa 位势高度场(蓝色框区), 850 hPa Wind 表示 850 hPa 风场(粉色框区)

Fig.1 The distributions of the study area (green rectangle, PRECIP) and the main factors (500-hPa geopotential height, 850-hPa wind). 500-hPa HGT indicates 500-hPa geopotential height (blue rectangle), 850 hPa Wind represent 850-hPa wind (pink rectangle)

海夏季风与我国汛期降水的关系, 结果表明南海夏季风指数与江南区降水和华南后汛期降水有显著正相关; 郑彬和施能(2006)的研究表明 20 世纪 70 年代后南海夏季风对华南区域降水有着显著的影响; 杨林等(2007)分析了东亚夏季旱涝与南海夏季风的密切关系。涉及延伸期降水预测或者夏季降水分析的研究多考虑 500 hPa 副高和 850 hPa 南海风场。据此, 选取以下两个因子作为影响季节降水的主要因子: 图 1 中蓝色粗实线包络区域的 500 hPa 高度场(严华生等, 2003; 黄嘉佑等, 2004; 姚愚等, 2004; 赵俊虎等, 2011, 2014)及红色粗实线包络区域(南海)的 850 hPa 风场(姚永红和钱永甫, 2001; 黄菲等, 2007; 靳莉君和赵平, 2012)。

建模数据拟采用的降水资料、高度场资料、风场资料为 NCEP 再分析月平均资料 (<http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html>[2018-01-05]), 资料空间分辨率为  $2.5^{\circ} \times 2.5^{\circ}$ 。所采用的西太平洋副热带高压(简称西太副高)指数资料来源于国家气候中心 (<http://cmdp.ncc.cma.gov.cn/160/74.php>[2018-01-05]), 所选取的指数包含有: 西太副高面积指数 ( $110^{\circ}\text{E} \sim 180^{\circ}$ )、西太副高强度指数 ( $110^{\circ}\text{E} \sim 180^{\circ}$ )、西太副高脊线指数 ( $110^{\circ} \sim 150^{\circ}\text{E}$ ) 及西太副高北界指数 ( $110^{\circ} \sim 150^{\circ}\text{E}$ )。资料时间跨度为 1979 年 1 月至 2016 年 6 月。

本文所构建的季节降水相关分析模型包括对

比模型和基于更为广义的线性回归季节降水模型(以下简称场回归模型), 其中对比模型采用西太副高指数, 在第四节中, 先行对比仅基于副高的一般回归模型(普通方案)和场回归方案效果, 在此之后, 着重分析考虑南海 850 hPa 风场和 500 hPa 高度场的场回归模型效果。

## 4 对比检验

利用 NCEP 和国家气候中心所提供的西太副高指数前 400 组数据, 构建回归方程 (1) 和 (2), 利用后 50 组数据进行后验比对; 在综合评估中, 综合分析场回归方案的拟合效果, 并分析高度场和风场的贡献及季节尺度降水模拟的效果。

### 4.1 普通方案和场回归方案

在本节中, 本文将呈现西太副高面积指数模型(mod\_a)、西太副高强度指数模型(mod\_b)、西太副高脊线指数模型(mod\_c)及西太副高北界指数模型(mod\_d)四个指数模型及场回归模型(mod\_e)构建的降水回归方程效果(试验分类见表 1)。作为对比, 场回归模型的因子项仅为图 1 粗蓝线包络的矩形区域内 500 hPa 高度场, 即, 采用单一预报因子的场回归方案:  $P' = AHB + E$ , 其中,  $P'$  为降水场,  $H$  为图 1 粗蓝线包络的矩形区域内 500 hPa 高度场,  $A$ 、 $B$ 、 $E$  为回归系数矩阵。各个回归模型的降水场与 NCEP 后验组的相关性如图 2 及图 3 所示, 其平均绝对误差如图 4 及图 5 所示, 图中

所示的相关系数均已通过显著性检验。

表 1 对比试验

Table 1 Comparative experiments

| 模型    | 因变量                               | 回归方程              |
|-------|-----------------------------------|-------------------|
| mod_a | 西太副高面积指数 ( $I_{area}$ )           | $p'=aI_{area}+b$  |
| mod_b | 西太副高强度指数 ( $I_{inten}$ )          | $p'=aI_{inten}+b$ |
| mod_c | 西太副高脊线指数 ( $I_{ridge}$ )          | $p'=aI_{ridge}+b$ |
| mod_d | 西太副高北界指数 ( $I_{north}$ )          | $p'=aI_{north}+b$ |
| mod_e | 500 hPa 高度场 ( $H$ ) (图 1 蓝粗线包围区域) | $P'=AHB+E$        |

注:  $p'$  为格点降水量,  $a$ 、 $b$  为回归系数。

从图 2~5 可以看出, 西太副高面积指数和西太副高强度指数所拟合的全国降水明显劣于西太副高脊线指数和西太副高北界指数的效果, 而在湖南、江西、福建西部及两广北部等地区, 西太副高面积指数和西太副高强度指数所拟合的降水要优于另两个指数; 这一点在误差上也有所体现。对比各个模型的结果, 无论是误差还是相关系数, 基于指数的一般回归模型均劣于场回归模型。

#### 4.2 场回归方案综合评估及因子贡献分析

在 4.1 节中, 对比了基于指数的季节降水模型和场回归模型, 对比结果表明场回归方案要明显优于对比试验, 其中主要原因在于指数无法系统地表征一个量的系统性影响。副高的四个指数只是从四个侧面描绘副高特征, 从拟合降水的效果来说, 这四个指数可归为两组。此外, 具有相同指数的副高, 可能具有不同的形式分布, 而这种分布对于降水而言可能至关重要, 而副高指数却无法全面地展示相应的副高特征。

为了较好地研究中国陆地区域降水的两个重要影响因子——副高和南海风场, 在降水量当中的贡献, 本文将南海夏季风也考虑进去, 此时的预报方程为:  $P'=L_1(AHB)+L_2(C_1UD_1)+L_3(C_2VD_2)+E$ , 其中,  $U$ 、 $V$  分别为图 1 红线包围区域纬向风场和经向风场,  $A$ 、 $B$ 、 $C_1$ 、 $D_1$ 、 $C_2$ 、 $D_2$ 、 $E$  为回归系数矩阵,  $L_1$ 、 $L_2$ 、 $L_3$  为上文所述线性规则。其拟合结果如图 6 和图 7 所示。从图中可以看出, 相较于 4.1 节中模型, 考虑了南海风场的季节降水模型在湖

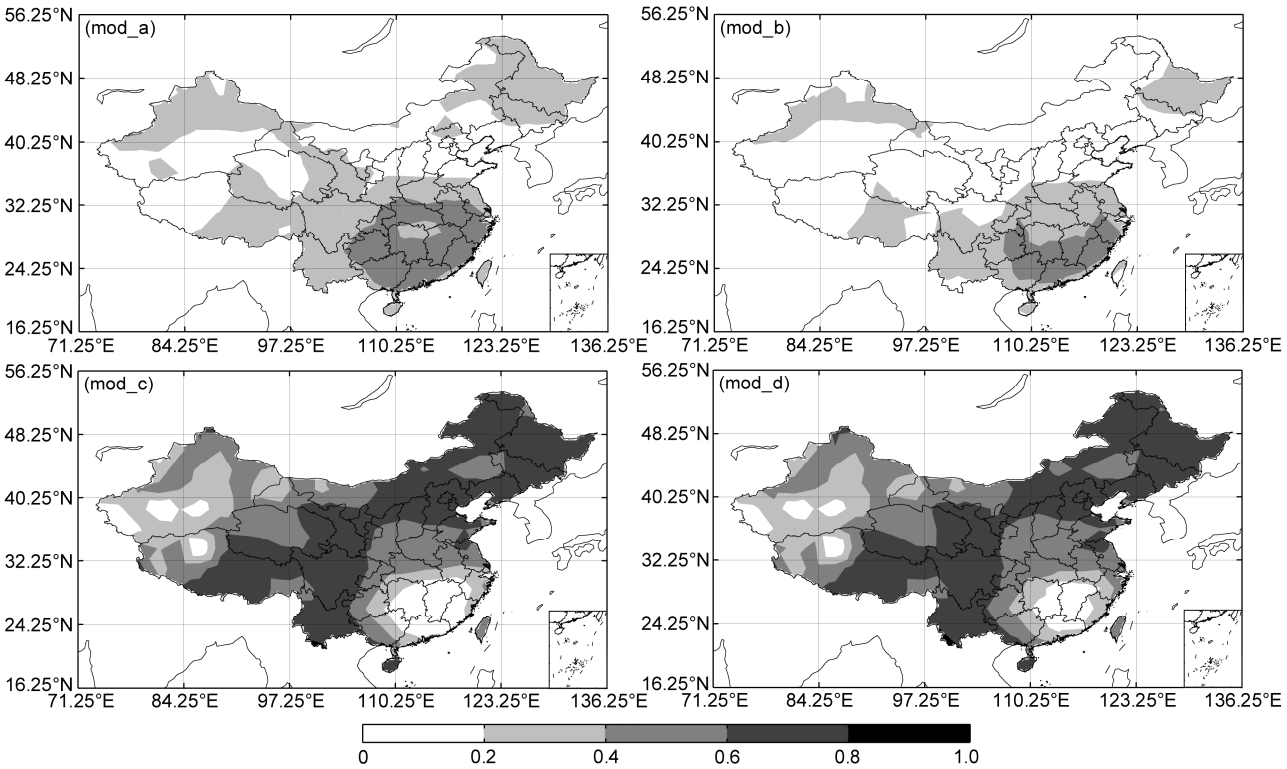


图 2 NCEP 再分析资料的降水场与西太副高面积指数模型 (mod\_a)、西太副高强度指数模型 (mod\_b)、西太副高脊线指数模型 (mod\_c) 及西太副高北界指数模型 (mod\_d) 的对比试验回归结果的相关系数分布

Fig. 2 The correlation coefficients of precipitation from NCEP reanalysis data and from the regressions with mod\_a (the predictor of regression equation is the index of the area of the subtropical high over the western Pacific), mod\_b (the predictor of regression equation is the index of the strength of the subtropical high over the western Pacific), mod\_c (the predictor of regression equation is the index of the ridge line of the subtropical high over the western Pacific), mod\_d (the predictor of regression equation is the index of the northern extend of the subtropical high over the western Pacific)

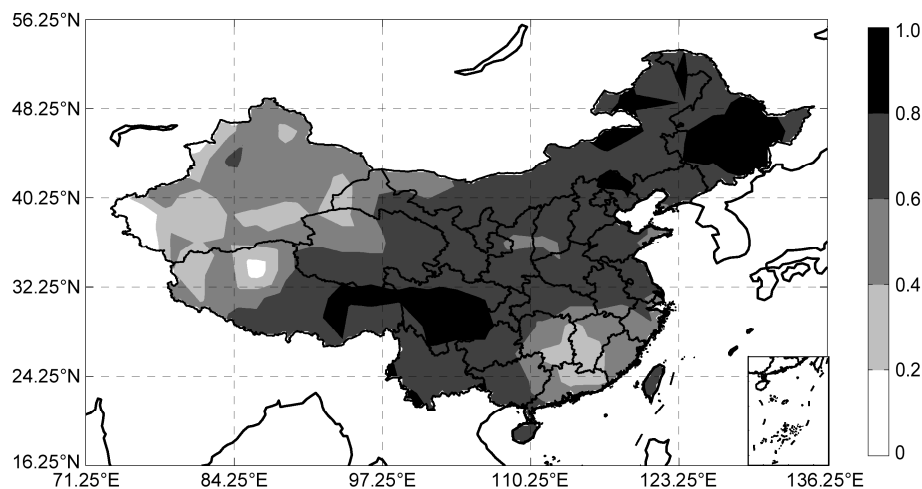


图3 NCEP再分析资料的降水场与场回归模型(mod\_e, 因子项仅含500 hPa高度场)回归结果相关系数分布

Fig. 3 The correlation coefficients of precipitation from NCEP and from the regressions with mod\_e (more generalized linear regression model, the predictor in the model is only 500-hPa geopotential height)

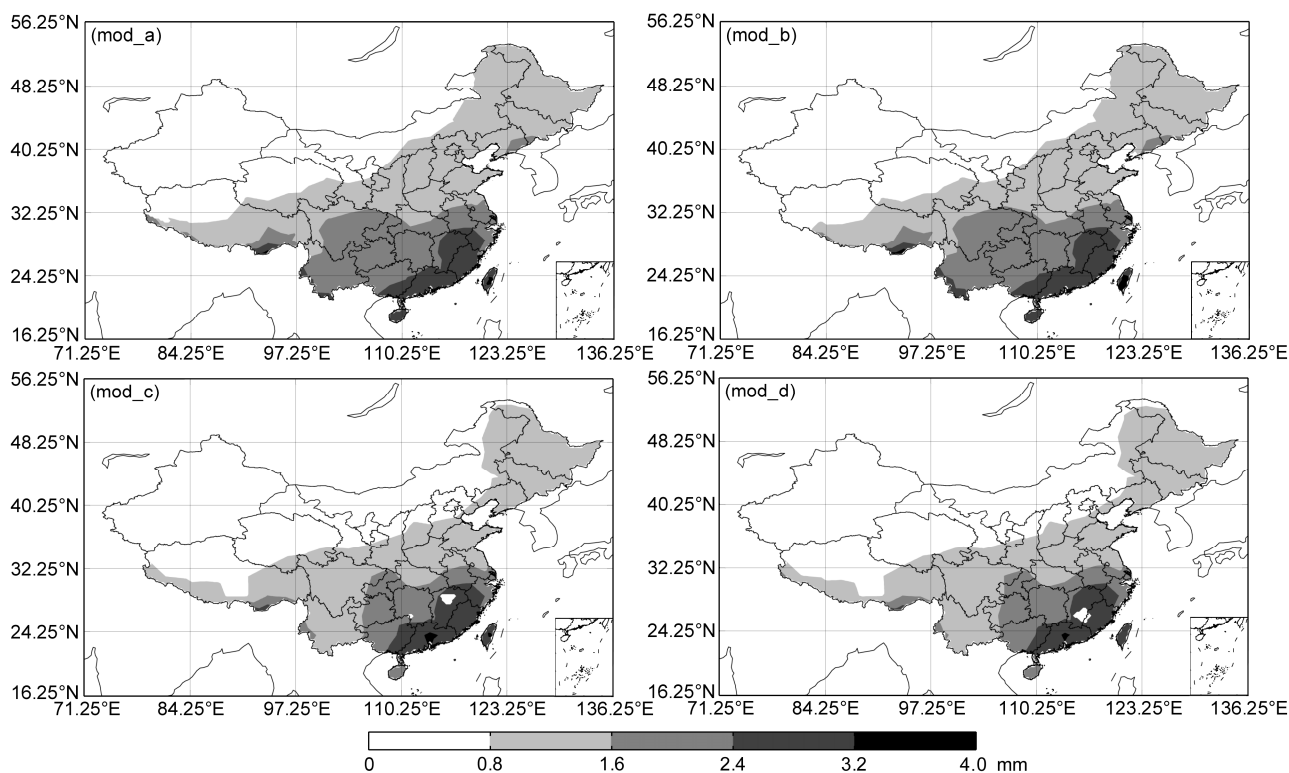


图4 同图2, 但为平均绝对误差分布

Fig. 4 As in Fig. 2, but for the mean absolute errors

南、江西、福建西部及两广北部等地区有了显著的改进, 且境内大部分区域相关性大于0.6, 平均在0.7左右, 东北地区 and 西南地区的相关性高于0.8, 同时误差整体下降(见图7)。

从我国降水分布来讲, 我国的雨带有着明显的季节性时空分布特征, 受多种系统共同影响, 夏季

降水较为集中, 很多地区存在雨热同期的特点。一般而言, 西太副高与我国雨带的移动有着密切关系, 当只考虑500 hPa高度场时, 在夏季降水集中时期, 长江中下游以南区域受副高控制, 此时如果只采用副高指数作为因子项, 模型对于降水的拟合将受限于样本关于副高和对应雨带之间关系的信

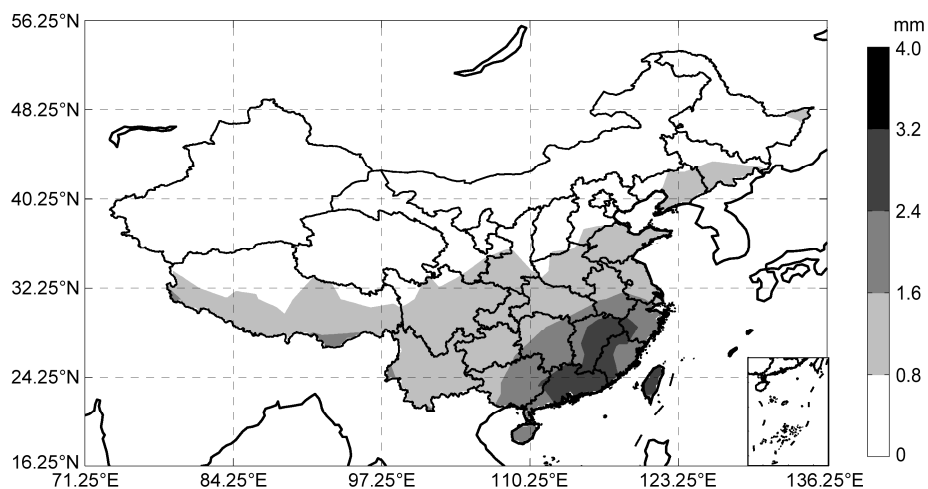


图 5 同图 3, 但为平均绝对误差分布

Fig. 5 As in Fig. 3, but for the mean absolute errors

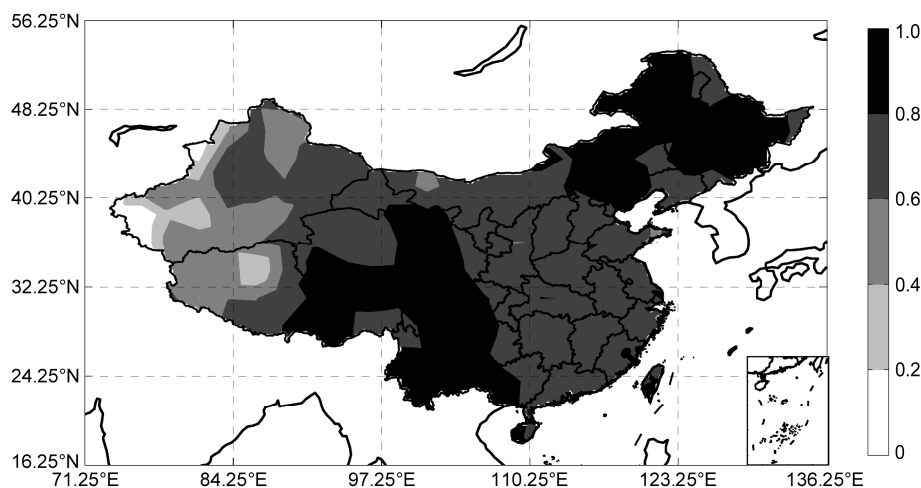


图 6 NCEP 再分析资料的降水场与场回归模型 (mod\_e, 因子项为 500 hPa 高度场和 850 hPa 风场) 回归结果相关系数分布

Fig. 6 The correlation coefficients of precipitation from NCEP and from the regressions with mod\_e (the predictors in the model are 500-hPa geopotential height and 850-hPa wind)

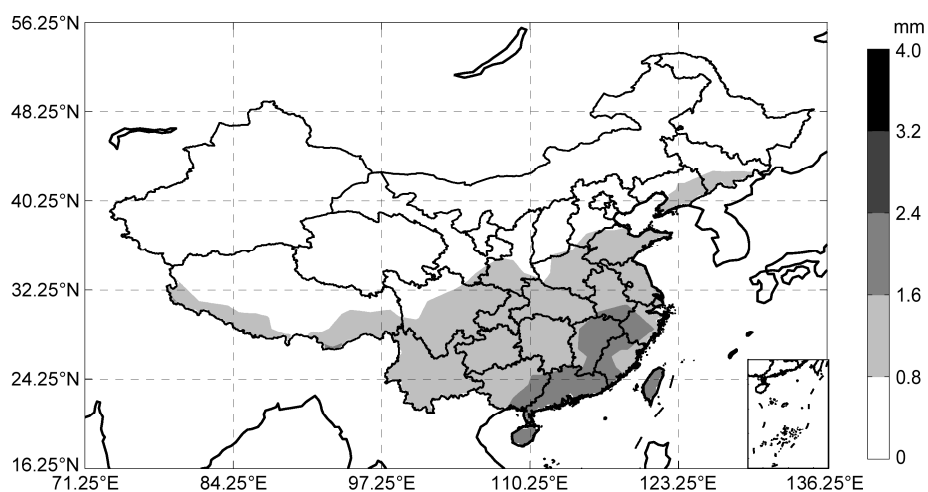


图 7 同图 6, 但为平均绝对误差分布

Fig. 7 As in Fig. 6, but for the mean absolute errors

息含量。而在上文所述区域中,对应主要降水时期,副高既与雨带密切相关,又不利于其控制区域降水发生,故而在统计模型中副高对降水的影响在此区域不是很显著。而南海夏季风与西太副高强度指数和面积指数存在显著的相关性并改善了这一区域的结果(图2)。对比表明,总体而言一般线性回归没有场回归效果好(图2~5)。

当在场回归方程中添加了南海 850 hPa 风场作为一个因子项,则可以将南海方向的水汽运输影响及副高的总体影响一起考虑进上文所述区域;从回归分析角度而言(Haykin, 2009),对于该区域来说,500 hPa 高度场不足以完全描述该区域降水特征,而 850 hPa 风场对于该区域降水而言则相当于一个关于降水量多少的分类要素,起到分类器的作用,进而改进预报效果。

此外,本文分析南海风场纬向分量( $u$ )、经向分量( $v$ )及 500 hPa 高度场在基于更为广义的线性回归降水模型中的贡献率,三者累积贡献率为 1(每个分量的贡献率在 0 到 1 之间,三者累加和等于 1),分析结果如图 8 所示。图中从左至右依次为 500 hPa 高度场贡献率、南海 850 hPa 风场  $u$  分量贡献率和南海 850 hPa 风场  $v$  分量贡献率。从三者比重而言,南海 850 hPa 风场  $v$  分量仅在海南至长江中下游一片有显著影响,以长江中下游地区贡献率最大,500

hPa 高度场则在其余地区有较高贡献率。此外,相较于  $v$  分量, $u$  风量在北方地区亦有相对较高贡献。

依据本小节结果,构建相应季节的平均降水,并与 NCEP 相应时次比对,其结果如图 9 所示,图中 CC 表示相关系数,ME 表示平均误差, RMSE 表示均方根误差, SIGMA 表示标准差, MAE 表示平均绝对误差, SI 表示预报 SI (skill index) 评分,图中点线为相应统计量的平均值, SI 阈值为 0.7, 用黑实线表示。从图 9 可以看出,对于季节降水,该模型的整体相关性平均达 0.9 以上,最低在 0.78 以上,拟合的偏差总体偏低约 0.06 mm,离散程度在 0.4 mm 以下,平均为 0.1 mm 左右,平均绝对误差在 0.5 mm 左右,预报 SI 评分平均值在 0.5 左右,最大值在 0.6 左右,具备一定参考性。

为了更全面地验证模型效果,我们采用交叉检验的方式对场回归模型进行检验。将建模数据依次分为三组,每组 150 个时次,建模因子和预测因子不变,共设立三个对比试验:试验 1 (exp\_1), 第一组为对比检验组,后两组为建模组;试验 2 (exp\_2), 第二组为对比检验组,第一和第三组为建模组;试验 3 (exp\_3), 第三组为对比检验组,前两组为建模组。交叉检验结果如图 10~12 所示:预测结果与观测组的相关性、平均误差、均方根误差等均与图 9 一致。另外,在交叉检验的试验中,

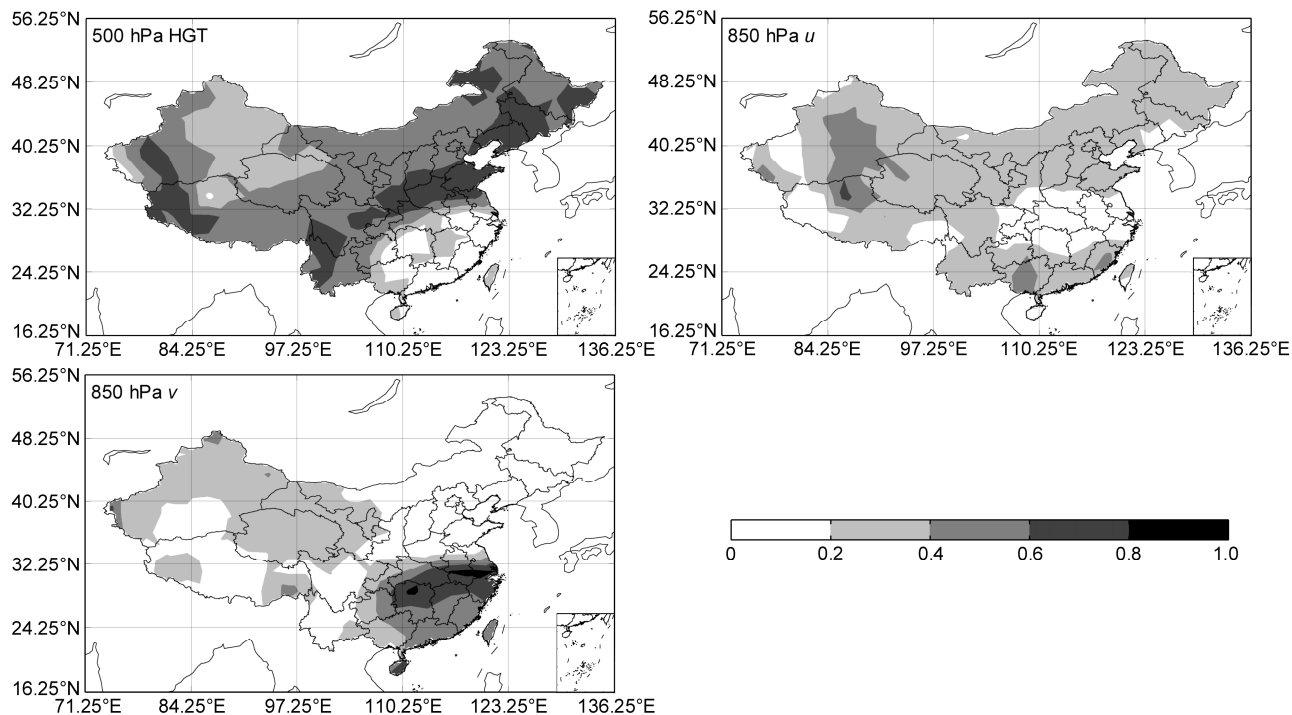


图 8 含 500 hPa 高度场和 850 hPa 风场作为预报因子的场回归模型各分量贡献率

Fig. 8 The contribution rates of individual components in mod\_e with 500-hPa geopotential height and 850-hPa wind as predictors

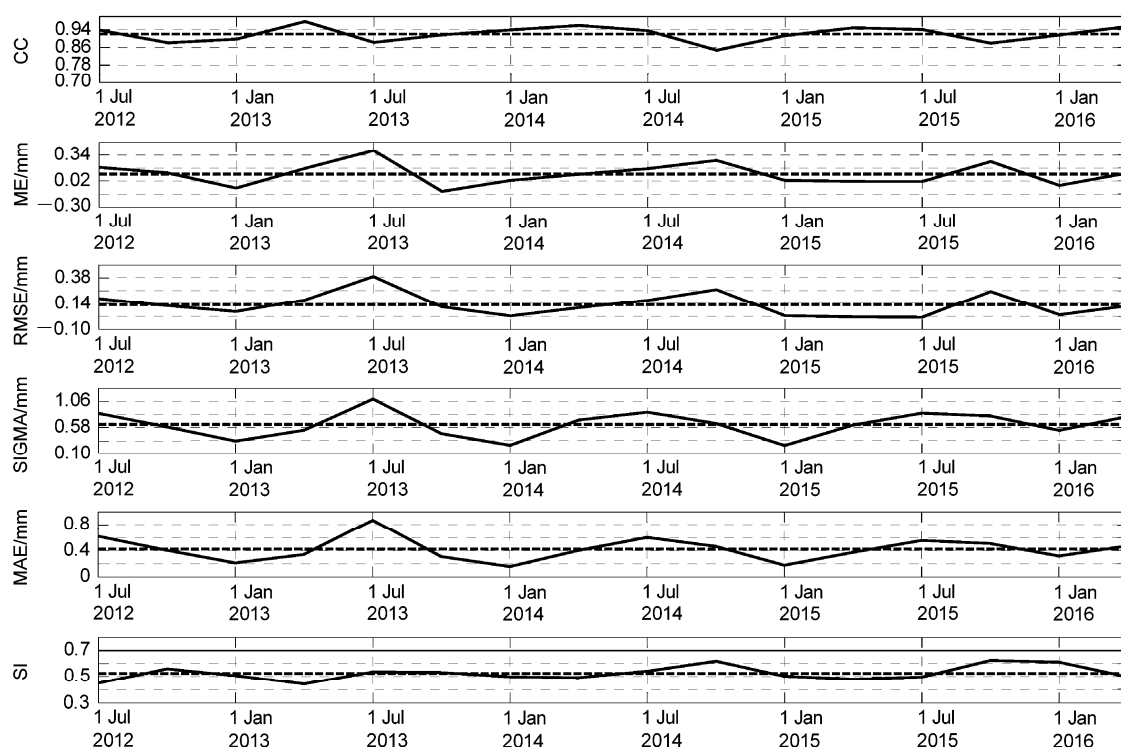


图 9 场回归模型季节尺度降水效果。CC、ME、RMSE、SIGMA、MAE、SI 表示相关系数、平均误差、均方根误差、标准差、平均绝对误差、预报 SI 评分，粗虚线为相应统计量的平均值，细实线为 SI 阈值 0.7，下同

Fig. 9 Seasonal precipitation results based on mod\_e. CC, ME, RMSE, SIGMA, MAE, and SI are correlation coefficient, mean error, root mean square error, standard deviation, mean absolute error, and forecasting SI (skill index) score, respectively, dashed thick lines are the mean values of the corresponding statistics, the threshold of the SI (skill index) score (0.7) is marked with solid thin line, the same below

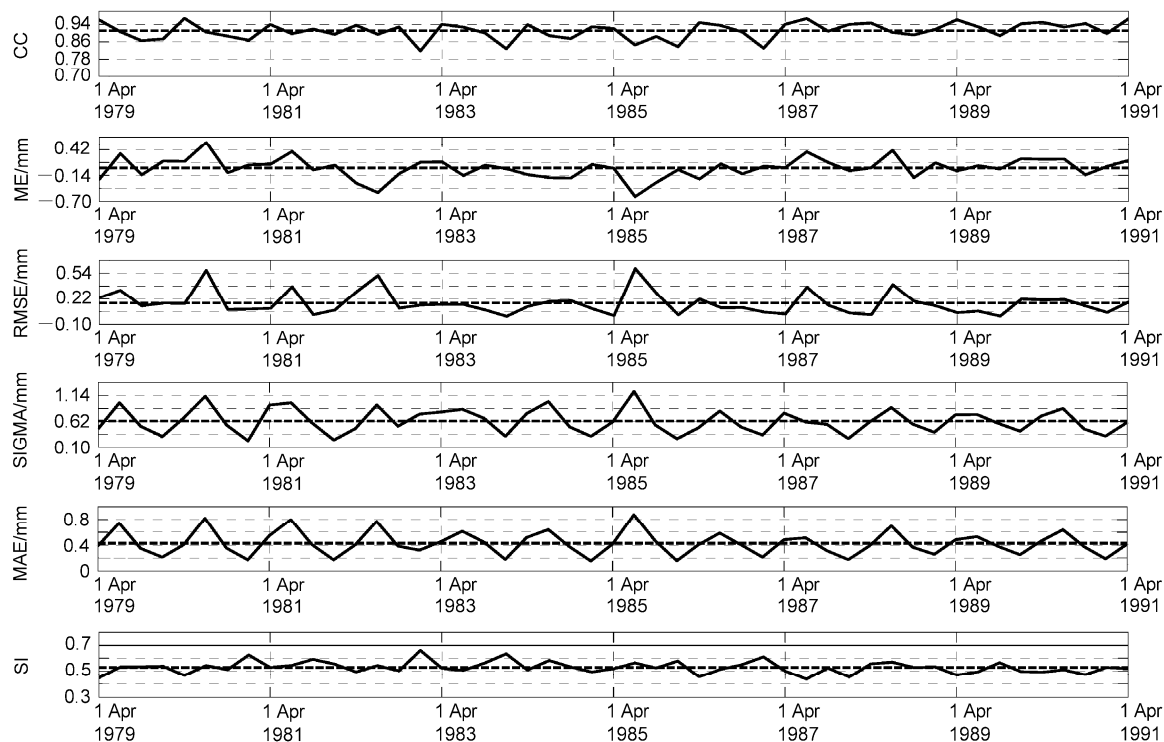


图 10 试验 1 (exp\_1) 的场回归模型季节尺度降水效果

Fig. 10 Seasonal precipitation results based on mod\_e in exp\_1 (the first 150 times data are the test group)

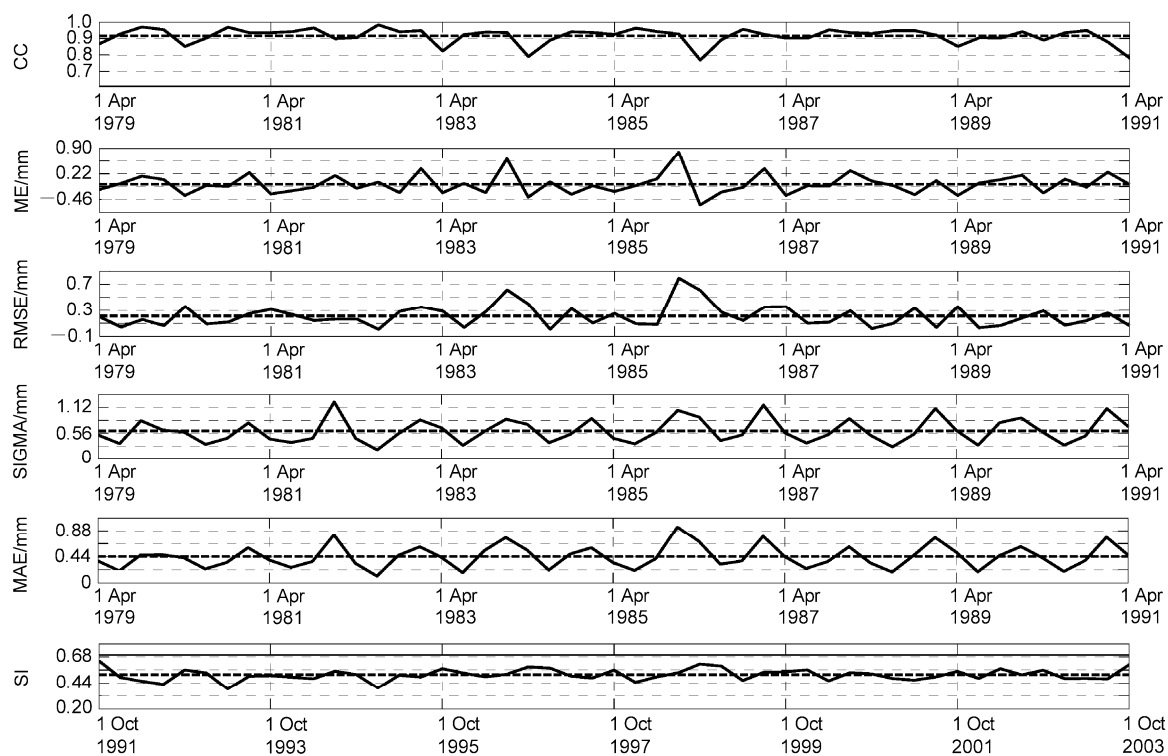


图 11 试验 2 (exp\_2) 的场回归模型季节尺度降水效果

Fig. 11 Seasonal precipitation results based on mod\_e in exp\_2 (the second 150 times data are the test group)

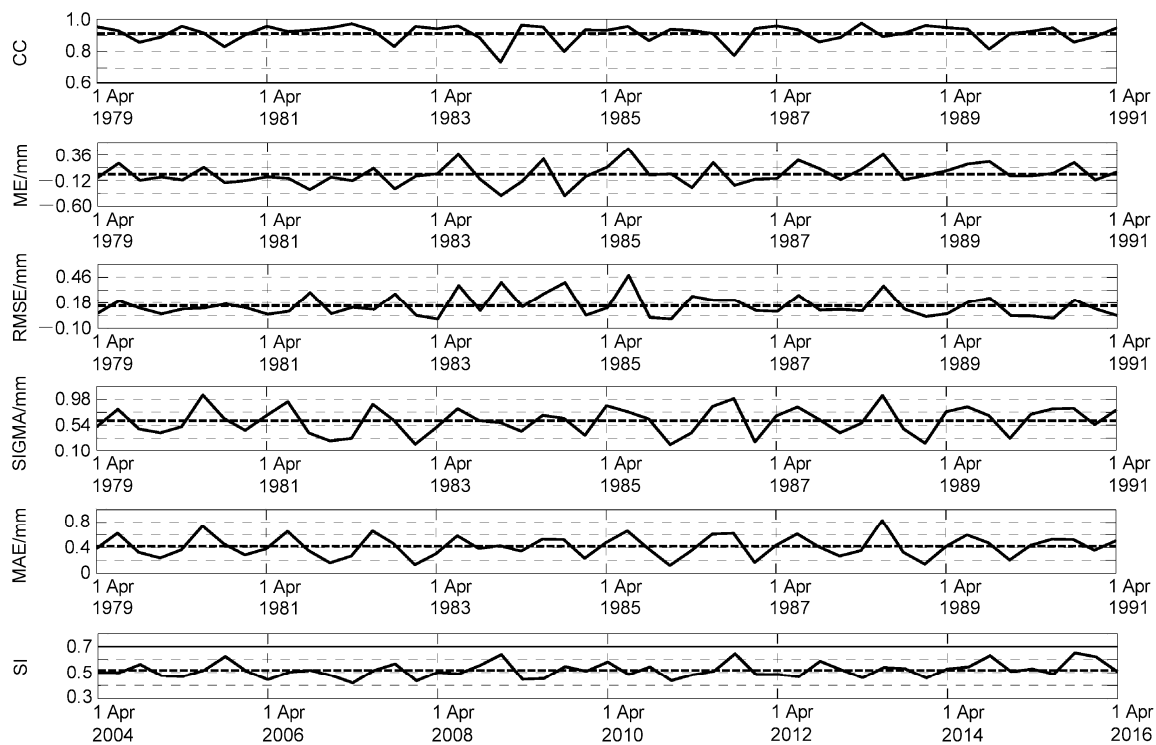


图 12 试验 3 (exp\_3) 的场回归模型季节尺度降水效果

Fig. 12 Seasonal precipitation results based on mod\_e in exp\_3 (the third 150 times data are the test group)

建模所采用的样本量仅为 300 个时次的数据，较图 9 试验的建模数据少 100 个时次。

## 5 结论与讨论

更为广义的线性回归模型是一般线性回归模型的延拓，相比于一般线性回归模型具有乘法不可交换性、可分析性、延拓性、降维特征及容错性等特点，其基本原理使该模型在标量因子及因变量的情形下可转化为一般线性回归，即，具有一般线性回归相等价的性质，因此，将其称之为更为广义的线性回归模型。基于该模型，本文利用图 1 所示的两个分布区域内的高度场和风场及相关副高指数对季节降水模型做了一般性的分析、讨论。从结果来看，在影响因子层面，基于更为广义的线性回归季节降水模型较基于西太副高指数的一般回归模型能更好的把握副高的不同特征及对降水的影响，而副高指数只能从不同侧面描述副高特征，且不具有描绘要素场不同分布特征对降水影响的能力；从分析结果上来看，南海 850 hPa 风场经向分量仅在海南至长江中下游一片有显著影响，以长江中下游地区贡献率最大，500 hPa 高度场则在其余地区有较高贡献率。

## 参考文献 (References)

- 陈丽娟, 李维京, 张培群, 等. 2003. 降尺度技术在月降水预报中的应用 [J]. 应用气象学报, 14 (6): 648–655. Chen Lijuan, Li Weijing, Zhang Peiqun, et al. 2003. Application of a new downscaling model to monthly precipitation forecast [J]. Journal of Applied Meteorological Science (in Chinese), 14 (6): 648–655, doi:10.3969/j.issn.1001-7313.2003.06.002.
- Chen Xuan. 2018. A preliminary study on dimension-reduction algorithm for variational methods in three dimensions [OL]. Preprints, doi:10.20944/preprints201801.0293.v1.
- 陈璇, 郑崇伟, 张伟涛, 等. 2017. 更为广义的线性回归模型及其气象应用 [J]. 解放军理工大学学报 (自然科学版), 18 (2): 144–149. Chen Xuan, Zheng Chongwei, Zhang Weitao, et al. 2017. More generalized linear regression model and its meteorological application [J]. Journal of PLA University of Science and Technology (Natural Science Edition) (in Chinese), 18 (2): 144–149, doi:10.12018/j.issn.1009-3443.20161215002.
- Haykin S O. 2009. Neural Networks and Learning Machines [M]. New Jersey: Prentice Hall, 50–55.
- 黄菲, 王宏, 戴平. 2007. 南海季风型海—气耦合模态的时空特征及其与中国夏季降水的关系 [J]. 中国海洋大学学报 (自然科学版), 37 (3): 351–356. Huang Fei, Wang Hong, Dai Ping. 2007. Spatial-temporal characters of the monsoon-ocean coupled mode over the South China Sea and its relation with summer precipitation of China [J]. Periodical of Ocean University of China (in Chinese), 37 (3): 351–356, doi:10.3969/j.

- issn.1672-5174.2007.03.002.
- 黄嘉佑, 刘舸, 赵昕奕. 2004. 副高、极涡因子对我国夏季降水的影响 [J]. 大气科学, 28 (4): 517–526. Huang Jiayou, Liu Ge, Zhao Xinyi. 2004. The influence of subtropical high indexes and polar vortex indexes on the summertime precipitation in China [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 28 (4): 517–526, doi:10.3878/j.issn.1006-9895.2004.04.04.
- 黄玮, 程建刚, 王学锋, 等. 2013. 基于气候模式的动力—统计降尺度技术在云南降水预测中的应用 [J]. 云南大学学报 (自然科学版), 35 (6): 782–790. Huang Wei, Cheng Jiangang, Wang Xuefeng, et al. 2013. An application of a dynamical-statistical downscaling method based on climate model to precipitation forecast over Yunnan [J]. Journal of Yunnan University (Natural Sciences) (in Chinese), 35 (6): 782–790, doi:10.7540/j.ynu.20120078.
- 靳莉君, 赵平. 2012. 夏季南海季风对长江中下游季风降水影响的观测分析和数值模拟 [J]. 气象学报, 70 (4): 670–680. Jin Lijun, Zhao Ping. 2012. Observationally analyzing and numerically simulating the impacts of the South China Sea summer monsoon on summer monsoon rainfall in the middle-lower reaches of the Yangtze River [J]. Acta Meteorologica Sinica (in Chinese), 70 (4): 670–680, doi:10.11676/qxxb2012.054.
- 王会军, 陈丽娟, 李维京, 等. 2007. 中国区域月平均温度和降水的模式可预报性分析 [J]. 气象学报, 65 (5): 725–732. Wang Huijun, Chen Lijuan, Li Weijing, et al. 2007. Predictability of DERF on monthly mean temperature and precipitation over China [J]. Acta Meteorologica Sinica (in Chinese), 65 (5): 725–732, doi:10.11676/qxxb2007.068.
- 魏凤英. 2007. 现代气候统计诊断与预测技术 (第二版) [M]. 北京: 气象出版社, 182–184. Wei Fengying. 2007. Modern Climate Statistical Diagnosis and Prediction Technology (second edition) (in Chinese) [M]. Beijing: China Meteorological Press, 182–184.
- 魏凤英, 黄嘉佑. 2010. 我国东部夏季降水量统计降尺度的可预测性研究 [J]. 热带气象学报, 26 (4): 483–488. Wei Fengying, Huang Jiayou. 2010. A study of predictability for summer precipitation on East China using downscaling techniques [J]. Journal of Tropical Meteorology (in Chinese), 26 (4): 483–488, doi:10.3969/j.issn.1004-4965.2010.04.013.
- 吴洪宝, 吴蕾. 2010. 气候变率诊断和预测方法 (第 2 版) [M]. 北京: 气象出版社, 112–116. Wu Hongbao, Wu Lei. 2010. Methods for Diagnosing and Forecasting Climate Variability (second edition) (in Chinese) [M]. Beijing: China Meteorological Press, 112–116.
- 吴尚森, 梁建茵, 李春晖. 2003. 南海夏季风强度与我国汛期降水的关系 [J]. 热带气象学报, 19 (Z1): 25–36. Wu Shangsen, Liang Jianyin, Li Chunhui. 2003. Relationship between the intensity of South China Sea summer monsoon and the precipitation in raining seasons in China [J]. Journal of Tropical Meteorology (in Chinese), 19 (Z1): 25–36, doi:10.3969/j.issn.1004-4965.2003.z1.003.
- 严华生, 严小冬, 王会军. 2003. 用 500 hPa 月高度场月降水预报所提供的可预报性时空分布 [J]. 热带气象学报, 19 (4): 389–396. Yan Huasheng, Yan Xiaodong, Wang Huijun. 2003. The predictability spatial and temporal distribution of making monthly rainfall prediction with 500 hPa monthly height field [J]. Journal of Tropical Meteorology (in Chinese), 19 (4): 389–396, doi:10.3969/j.issn.1004-4965.2003.04.006.
- 杨林, 魏应植, 丁一汇, 等. 2007. 东亚地区夏季旱涝分布与南海夏季风

- 爆发时间关系 [J]. 自然灾害学报, 16 (5): 52–58. Yang Lin, Wei Yingzhi, Ding Yihui, et al. 2007. Relationship between summer drought–flood distribution in East Asia region and breaking out time of South China Sea summer monsoon [J]. Journal of Natural Disasters (in Chinese), 16 (5): 52–58.
- 姚愚, 严华生, 程建刚. 2004. 主汛期 (6~8 月) 副高各指数与中国 160 站降雨的关系 [J]. 热带气象学报, 20 (6): 651–661. Yao Yu, Yan Huasheng, Cheng Jiangang. 2004. Relationship between subtropical high indexes at the main raining period with the rainfall of 160 sampling stations in China [J]. Journal of Tropical Meteorology (in Chinese), 20 (6): 651–661, doi:10.3969/j.issn.1004-4965.2004.06.005.
- 姚永红, 钱永甫. 2001. 用湿位涡定义的南海西南季风指数及其与我国区域降水的关系研究 [J]. 南京大学学报 (自然科学), 37 (6): 781–788. Yao Yonghong, Qian Yongfu. 2001. A study on the South China Sea monsoon index and the relationship between the index and regional rainfalls of China [J]. Journal of Nanjing University (Natural Sciences) (in Chinese), 37 (6): 781–788, doi:10.3321/j.issn:0469-5097.2001.06.020.
- 俞亚勋, 王式功, 钱正安, 等. 2013. 夏半年西太副高位置与东亚季风雨带(区)的气候联系 [J]. 高原气象, 32 (5): 1510–1525. Yu Yachu, Wang Shigong, Qian Zheng'an, et al. 2013. Climatic linkages between SHWP position and EASM rainy-belts and -areas in east part of China in summer half year [J]. Plateau Meteorology (in Chinese), 32 (5): 1510–1525, doi:10.7522/j.issn.1000-0534.2013.00033.
- 袁媛, 李崇银. 2009. 热带印度洋海温异常不同模式对南海夏季风爆发的可能影响 [J]. 大气科学, 33 (2): 325–336. Yuan Yuan, Li Chongyin. 2009. Possible impacts of the tropical Indian Ocean SST anomaly modes on the South China Sea summer monsoon onset [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 33 (2): 325–336, doi:10.3878/j.issn.1006-9895.2009.02.11.
- 赵俊虎, 封国林, 王启光, 等. 2011. 2010 年我国夏季降水异常气候成因分析及预测 [J]. 大气科学, 35 (6): 1069–1078. Zhao Junhu, Feng Guolin, Wang Qiguang, et al. 2011. Cause and prediction of summer rainfall anomaly distribution in China in 2010 [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 35 (6): 1069–1078, doi:10.3878/j.issn.1006-9895.2011.06.07.
- 赵俊虎, 支蓉, 申茜, 等. 2014. 2012 年我国夏季降水预测与异常成因分析 [J]. 大气科学, 38 (2): 237–250. Zhao Junhu, Zhi Rong, Shen Qian, et al. 2014. Prediction of the distribution of the 2012 summer rainfall in China and analysis of the cause for anomaly [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 38 (2): 237–250, doi:10.3878/j.issn.1006-9895.2013.12215.
- 郑彬, 施能. 2006. 南海夏季风对华南夏季降水年代际变化的影响 [J]. 南京气象学院学报, 29 (4): 477–483. Zheng Bin, Shi Neng. 2006. Effects of South China Sea summer monsoon on the interdecadal variability of summer rainfall over South China [J]. Journal of Nanjing Institute of Meteorology (in Chinese), 29 (4): 477–483, doi:10.3969/j.issn.1674-7097.2006.04.007.
- 郑崇伟, 高悦, 陈璇. 2017. 巴基斯坦瓜达尔港风能资源的历史变化趋势及预测 [J]. 北京大学学报 (自然科学版), 53 (4): 617–626. Zheng Chongwei, Gao Yue, Chen Xuan. 2017. Climatic long term trend and prediction of the wind energy resource in the Gwadar Port [J]. Acta Scientiarum Naturalium Universitatis Pekinensis (in Chinese), 53 (4): 617–626, doi:10.13209/j.0479-8023.2017.012.
- 郑崇伟. 2018a. 21 世纪海上丝绸之路: 风能的长期变化趋势 [J]. 哈尔滨工程大学学报, 39 (3): 399–405. Zheng Chongwei. 2018a. Wind energy trend in the 21st Century Maritime Silk Road [J]. Journal of Harbin Engineering University (in Chinese), 39 (3): 399–405, doi:10.11990/jheu.201704019.
- 郑崇伟. 2018b. 21 世纪海上丝绸之路: 斯里兰卡海域的波浪能评估及决策建议 [J]. 哈尔滨工程大学学报, 39 (4): 614–621. Zheng Chongwei. 2018b. 21st Century Maritime Silk Road: Wave energy evaluation and decision and proposal of the Sri Lankan waters [J]. Journal of Harbin Engineering University (in Chinese), 39 (4): 614–621, doi:10.11990/jheu.201704020.
- 郑崇伟. 2018c. 21 世纪海上丝绸之路: 关键节点的能源困境及应对 [J]. 太平洋学报, 26 (7): 71–78. Zheng Chongwei. 2018c. Energy predicament and countermeasures on key junctions of the 21st Century Maritime Silk Road [J]. Pacific Journal (in Chinese), 26 (7): 71–78.