

基于偏最小二乘回归的赤道电离异常区
峰值 TEC 预测*

徐振中^{1,2)} 王伟民¹⁾ 邓亮³⁾
(1)解放军理工大学气象学院,南京 211101
(2)解放军 66011 部队,北京 102600
(3)广州军区气象水文中心,广州 510083

摘 要 基于 IGS 提供的 120°E 上空 1999—2009 年 IONEX 格式电离层电子浓度总含量(TEC)资料,使用偏最小二乘回归方法进行拟合训练,构建了赤道电离异常南、北驼峰峰值 TEC 的预测模型,并以 2009 年的 TEC 数据进行预测试验。结果表明,该预测模型能较准确地描述南北驼峰峰值 TEC 的变化特征,但对北驼峰峰值的预测效果要好于对南驼峰的预测;对南、北驼峰峰值 TEC 预测的平均绝对误差分别为 4.441 0、2.915 1 TECU,平均相对误差分别为 23.26%、10.78%,预测值与真实值的相关系数分别为 0.712 2、0.785 9。

关键词 赤道电离异常;驼峰;电子浓度总含量(TEC);偏最小二乘回归;预测
中图分类号:P352 **文献标识码**:A

PREDICTION OF PEAK IONOSPHERIC TOTAL ELECTRON
CONTENT OF EQUATORIAL IONIZATION ANOMALY BASED
ON PARTIAL LEAST SQUARE REGRESSION

Xu Zhenzhong^{1,2)}, Wang Weimin¹⁾ and Deng Liang³⁾

(1) Meteorology Institute, PLA Science and Technology University, Nanjing 211101
(2) 66011 Troops of PLA, Beijing 102600
(3) Meteorological and Hydrological Center of Guangzhou Military Region, Guangzhou 510083

Abstract On the basis of the 1999—2009 IONEX format total electron content (TEC) data over 120°E from IGS, the method of Partial Least Square Regression is used to establish the prediction model of EIA(equatorial ionization anomaly) index. The TEC data from 1999 to 2008 were used to fit the prediction model and the all-year TEC data of 2009 was used to test the model. The results of the forecasting experimentations of the model show that this model can accurately describe the variation of EIA index Inc (TEC of north crest) and Isc (TEC of south crest), but the prediction effect of Inc is better than the Isc. The mean absolute error of Isc and Inc is about 4.4410TECU and 2.915 1 TECU, the mean relative error of Isc and Inc is about 23.26% and 10.78%. The correlation coefficient between the actual values and predicted values of Isc and Inc is 0.712 2 and 0.785 9 respectively.

Key words: equatorial ionization anomaly (EIA); crest; total election content(TEC); partial least square regression; prediction

* 收稿日期:2012-01-08
基金项目:总装预研项目
作者简介:徐振中,男,1985 年生,研究方向为电离层物理与预报技术. E-mail: xuzhenzhong821@yahoo.com.cn

1 引言

赤道电离异常(EIA)是指白天在磁赤道上空电离层电子浓度总含量(TEC)出现极小值,而在磁赤道南北两边 $15^{\circ} \sim 20^{\circ}$ 的区域出现两个极大值,分别被称为南驼峰和北驼峰^[1-4]。近些年,赤道电离异常(EIA)现象一直是电离层研究领域的热点。如Huang和Cheng^[5]在1996年利用处于赤道异常北驼峰区内单站电离层资料,分析得出赤道电离异常区的强度和太阳黑子数之间存在线性正相关,但异常区出现的时间和纬度以及太阳活动之间没有相关性,同时还指出冬半球驼峰区出现早于夏半球,夏季驼峰区出现纬度低于其他季节等现象。Wu等^[6]使用台湾区域9个电离层测站资料,分析得出在太阳活动低年EIA强度指数均值与地磁Dst指数有较好的相关性,而与太阳F10.7射电流量相关性不好,只有在较长时间尺度上才与太阳F10.7射电流量有较好相关性。Shweta等^[7]使用处于赤道电离异常北驼峰区的印度Bhopal站2005—2006年TEC资料,验证了赤道电离异常区TEC的“半年异常”、“冬季异常”以及“日落后增强”等现象。Zhang等^[8]在2009年通过分析1998—2004年 120°E 经线附近TEC-CTR指数,得到该指数呈明显的半年变化周期和在南驼峰区该指数变化没有出现“半年异常”等结论。由于太阳天顶角、磁场作用、越赤道的中性风以及化学成分的改变等原因,赤道电离异常呈现出日变化、季节变化,年变化等特征。

然而,在大范围内收集获取赤道电离异常资料绝非易事,现有的研究工作大多是采用赤道低纬地区单站或者局部区域几个站的短期资料进行分析,而且也主要是对赤道电离异常的变化特征和影响机制进行分析,赤道异常指数预测方面的工作几乎没有,对我国东南沿海地区所处的与赤道异常北驼峰区域的研究工作也相对较少。因此我们选取IGS发布的 120°E 上空1999—2009年IONEX格式TEC资料,以及同时期的太阳、地磁活动等资料,利用偏最小二乘回归方法对南、北驼峰的峰值TEC进行预测。

2 研究资料

根据文献[5,9]研究,存在一些典型的表征赤道异常的特征指数:北驼峰(南驼峰)峰值TEC,即Inc(Isc);南、北驼峰峰值TEC出现的纬度Lnc(Lsc)和时间Tnc(Tsc);It1300表示LT1300的赤道槽区TEC值。本文使用从IGS下载的1999—2009年的IONEX数据,该资料自1998年起开始发布,在

2002年第307天之前,每天包括12副VTEC地图,起始时间为01:00UT,终止时间为23:00UT,时间间隔为2小时。之后的数据起始时间为00:00UT,终止时间为24:00UT,时间间隔仍为2小时。地理范围为 $-87.5^{\circ} \sim 87.5^{\circ}\text{N}$, $-180^{\circ} \sim 180^{\circ}\text{E}$,纬度间隔为 2.5° ,经度间隔为 5° 。我们选取 120°E 线上的数据制作时间和纬度的TEC地图,再从中分别选取磁赤道两边最大等值线内的区域作为赤道电离异常驼峰区,求得这一区域的平均TEC即为南、北驼峰区峰值Isc和Inc。

太阳活动资料包括太阳F10.7射电流量和太阳黑子数SSN,地磁活动资料包括Kp、Ap指数和地磁Dst指数。

3 偏最小二乘回归方法

偏最小二乘回归(PLS, Partial Least Square Regression)方法作为一种新型的多元统计方法,可以基于多因变量对多自变量的回归建模,能较好地解决许多以往用普通多元回归无法解决的自变量之间的多重相关性问题。因变量预测计算原理如下:

对于 p 个自变量 $\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$ 和单

个因变量 $\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, 分别取 n 组样本点, 然后对原始

数据进行标准化处理。记 \boldsymbol{F}_0 为因变量的标准化向量, \boldsymbol{E}_0 为自变量集合 \boldsymbol{X} 的标准化矩阵。从 \boldsymbol{X} 中提取综合变量:

$$t_1 = \boldsymbol{X} \boldsymbol{w}_1 = w_{11}x_1 + w_{12}x_2 + \cdots + w_{1p}x_p \tag{1}$$
其中:

$$\begin{cases} w_1 = \frac{1}{\sqrt{\sum_{j=1}^p \text{cov}^2(x_j, y)}} \begin{vmatrix} \text{cov}(x_1, y) \\ \text{cov}(x_2, y) \\ \vdots \\ \text{cov}(x_p, y) \end{vmatrix} \\ w_{1j} = \frac{\text{cov}(x_j, y)}{\sqrt{\sum_{j=1}^p \text{cov}^2(x_j, y)}} \end{cases} \tag{2}$$

$$t_1 = \boldsymbol{E}_0 \boldsymbol{w}_1 = \frac{1}{\sqrt{\sum_{j=1}^p \text{cov}^2(x_j, y)}} [\text{cov}(x_1, y) \boldsymbol{E}_{01} + \text{cov}(x_2, y) \boldsymbol{E}_{02} + \cdots + \text{cov}(x_p, y) \boldsymbol{E}_{0p}] \tag{3}$$

将式(3)实施 \boldsymbol{E}_0 在 t_1 上的回归以及 \boldsymbol{F}_0 在 t_1 上的回归,即:

$$\boldsymbol{E}_0 = t_1 p_1 + \boldsymbol{E}_1, \boldsymbol{F}_0 = t_1 r_1 + \boldsymbol{F}_1 \quad (4)$$

其中, \boldsymbol{E}_1 、 \boldsymbol{F}_1 分别是 \boldsymbol{E}_0 和 \boldsymbol{F}_0 的残差矩阵, p_1 、 r_1 是回归系数:

$$p_1 = \frac{\boldsymbol{E}_0' t_1}{\|t_1\|^2}, r_1 = \frac{\boldsymbol{F}_0' t_1}{\|t_1\|^2} \quad (5)$$

进一步将式(4)以残差矩阵 \boldsymbol{E}_1 和 \boldsymbol{F}_1 代替 \boldsymbol{E}_0 和 \boldsymbol{F}_0 , 采用与提取 t_1 同样的方法, 进行第二个综合变量 t_2 的提取并施以回归。如此循环往复, 直到提取了 h 个综合变量 t_1, t_2, \dots, t_h , 并利用交叉有效性检验 $Q_h^2 \geq 0.0975$ 来确定 h 的值^[10]。

由于 t_1, t_2, \dots, t_h 均可以表示成 $E_{01}, E_{02}, \dots, E_{0p}$ 的线性组合, 因此最终可以得出 PLS 模型为:

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + y_h \quad (6)$$

4 基于偏最小二乘回归的预测模型

4.1 预测因子的选取和模型的构建

用太阳黑子数 SSN 和太阳 10.7 cm 射电流量 F10.7 表征太阳活动对电离层的影响, 用全球三小时磁情指数 Kp、磁扰指数 Ap 和环电流指数 Dst 来表征磁层活动的作用, 还有前一天北驼峰的峰值 TEC。

另外, 考虑到 Inc 和 Isc 变化呈现出明显的季节性特点, 因此选择天数 DAY 为年变化因子 ($1 \leq DAY \leq 365$), 且将 2000、2004 和 2008 年的 2 月 29 日数据舍弃, 但 DAY 值可能发生跳变, 为了克服跳变问题, 将其转换为两个正交分量:

$$DAYS = \sin\left(\frac{2\pi DAY}{365}\right) \quad (7)$$

$$DAYC = \cos\left(\frac{2\pi DAY}{365}\right) \quad (8)$$

所以输入数据包括观测数据输入和时间输入两类。观测数据包括 SSN_T、F10.7_T、Kp_T、Ap_T、Dst_T、Inc_T 或 Isc_T, 时间输入包括 DAYS_T 和 DAYC_T, 输出数据则为 $T+1$ 时刻的北驼峰峰值 Inc 即 Inc_{T+1}。使用 1999—2008 年 120°E 赤道电离异常历史资料进行拟合构建预测模型, 并以 2009 年数据作为测试数据进行预测试验。经对北驼峰历史资料计算, 提取第一个 PLS 分量交叉有效性检验 Q_h^2 值为 0.844 2, 第二个分量为 0.295 6, 第三个分量为 0.324 2, 第四个分量为 0.328 5, 第五个分量为 0.000 3 (小于临界值 0.097 5), 因此选取四个分量构建模型:

$$\begin{aligned} Y = & 0.856\,776X_1 + 0.093\,378X_2 - 0.010\,345\,9X_3 + \\ & 1.375\,67X_4 - 0.100\,576X_5 - 0.062\,479\,6X_6 + \\ & 0.671\,55X_7 + 0.647\,337X_8 - 2.748\,41 \end{aligned} \quad (9)$$

式中, Y 为预测值 Inc_{T+1}, X_1, X_2, \dots, X_8 分别对应

Inc_T、F10.7_T、SSN_T、Kp_T、Ap_T、Dst_T、DAYS_T、DAYC_T。

经对南驼峰历史资料的计算, 提取第一个 PLS 分量交叉有效性检验 Q_h^2 值为 0.780 7, 第二个分量为 0.217 9, 第三个分量为 0.130 9, 第四个分量为 0.113 2, 第五个分量为 -0.000 8 (小于临界值 0.097 5), 因此选取四个分量来构建预测模型:

$$\begin{aligned} Y = & 0.725\,311X_1 + 0.112\,246X_2 - 0.029\,321\,5X_3 + \\ & 0.670\,623X_4 - 0.192\,995X_5 - 0.105\,595X_6 + \\ & 1.652\,64X_7 + 2.396\,82X_8 + 0.450\,123 \end{aligned} \quad (10)$$

4.2 预测结果和误差分析

首先对带入预测模型的数据进行归一化处理, 然后再进行预测试验, 试验结果与 2009 年的实测值进行对比, 并用绝对误差和相对误差来表征预测精度。

$$\text{绝对误差 } |Error| = |I_{\text{pre}} - I_{\text{act}}| \quad (11)$$

$$\text{平均绝对误差 } Ea = \frac{1}{N} \sum_{i=1}^N |I_{\text{pre}}(i) - I_{\text{act}}(i)| \quad (12)$$

$$\text{平均相对误差 } Er = \frac{1}{N} \sum_{i=1}^N \frac{|I_{\text{pre}}(i) - I_{\text{act}}(i)|}{I_{\text{act}}(i)} \quad (13)$$

式中, I_{pre} 、 I_{act} 分别表示 Inc 的预测值和实际值, N 为预测总样本数。

Inc 和 Isc 预测值与实际值对比以及绝对误差分布情况如图 1 和图 2 所示。可见模型的预测效果很好, 对 Inc 预测的平均绝对误差为 2.915 1 TECU, 平均相对误差为 10.78%, 预测值与真实值相关系数为 0.785 9。对 Isc 预测的平均绝对误差为 4.441 0 TECU, 平均相对误差为 23.26%, 预测值与真实值相关系数为 0.7122。图 1 和图 2 分别给出对 Inc 和 Isc 预测的逐月平均绝对误差和平均相对

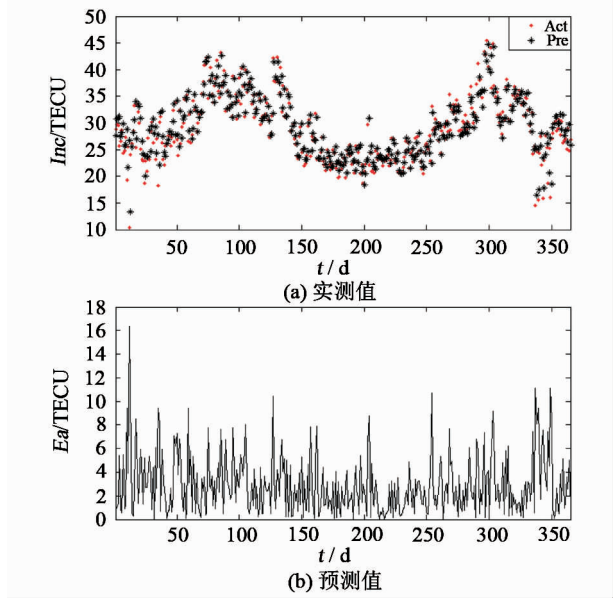


图 1 赤道电离异常北驼峰峰值 Inc 预测效果

Fig. 1 Forecasted results of magnitude of TEC at the northern anomaly crest

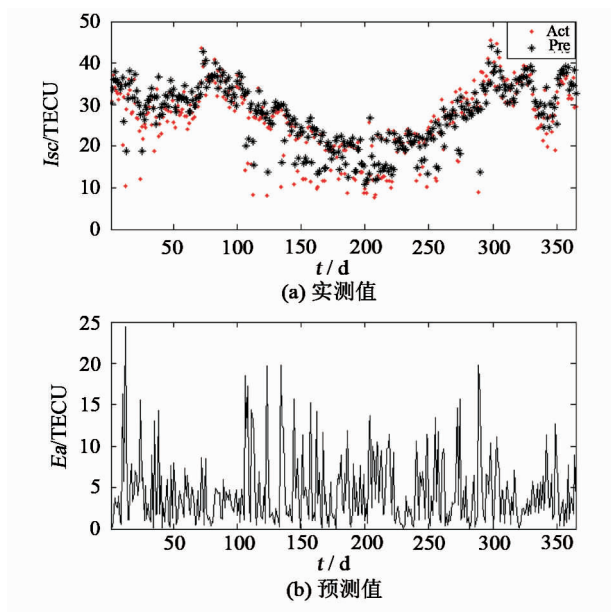


图2 赤道电离异常南驼峰峰值 I_{sc} 预测效果

Fig. 2 Forecasted results of magnitude of TEC at the southern anomaly crest

误差统计情况,从统计结果来看,对 I_{nc} 预测在1月和12月份误差较大,没有呈现典型的季节分布特征;但对北驼峰峰值 I_{nc} 的预测精度要高于南驼峰峰值 I_{sc} 的预测精度,可能的原因是地理纬度和地磁纬度的差异,南驼峰几乎处于地理赤道附近,电动力学和越赤道风场的变化机理更为复杂,使得预测准确度较低。总的来说,基于偏最小二乘回归方法能够较准确地实现南、北驼峰峰值 TEC 的预测。

5 结论与讨论

使用 $120^{\circ}E$ 上空 IGS 发布的 1999—2009 年 IONEX 格式 TEC 格点资料,引入包括观测数据和时间输入的 SSN、F10.7、Kp、Ap、Dst、Inc 或 I_{sc} 、DAY5 和 DAYC 八个预测因子,基于偏最小二乘回归方法,构建了赤道电离异常南、北驼峰峰值 I_{sc} 和 I_{nc} 的预测模型,并以 2009 年全年的数据作为测试数据来检验预测效果。结果表明,基于偏最小二乘回归方法构建的预测模型能够较好地实现赤道电离异常驼峰峰值的预测,对 I_{nc} 预测的平均相对误差为 10.78%,对 I_{sc} 预测的平均相对误差为 23.26%。本文对赤道电离异常特征指数预测工作只是初步尝试,尚有许多不完善需要改进之处:

1) 受限于格点资料精度的局限性,只选择了 $120^{\circ}E$ 经线上数据进行研究。

2) 影响赤道电离异常的因子众多,其变化过程

非常复杂,所选取的预测因子无法涵盖所有的影响因素和完整的作用过程信息,如发生剧烈的太阳活动如耀斑、日冕物质抛射或者行星际磁场发生南向翻转,太阳风能量注入磁层并向低纬地区渗透引发电离层暴等,都会对 EIA 特征产生影响,预测精度也随之降低,因此预测也就需要引入导致突变的影响因子,再根据赤道电离异常对变化的响应将预测结果加以修正,得到更准确的结果,这也是预测模型下一步改进的方向。

致谢 感谢 IGS 提供 IONEX 资料、世界数据中心提供地磁指数资料!

参 考 文 献

- 1 Appleton E V. Two anomalies in the ionosphere[J]. Nature, 1946, 157: 691–693.
- 2 Bramley E N and Peart M. Effect of ionization transport on the equatorial F region[J]. Nature, 1965, 206: 705–706.
- 3 Wu C C, et al. Variation of ionospheric total electron content in Taiwan region of the equatorial anomaly from 1994–2003[J]. Advances in Space Research, 2008, 41: 611–616.
- 4 Duncan R A. The equatorial F region of the ionosphere[J]. J Atmos Terr Phys., 1960, 18(2): 89–100.
- 5 Huang Y N and Cheng K. Solar cycle variation of equatorial ionospheric anomaly in total electron content in the Asian region[J]. Journal of Geophysical Research, 1996, 101: 24513–24520.
- 6 Wu C C, et al. Annual TEC variation in the equatorial anomaly region during solar minimum[J]. J Atmos Terr Phys., 2004, 66: 199–207.
- 7 Shweta M, et al. Seasonal variation of total electron content at crest of equatorial anomaly station during low solar activity condition[J]. Advances in Space Research, 2010, 46: 291–295.
- 8 Man Lian Zhang, et al. Variability study of the crest-to-trough TEC ratio of the equatorial ionization anomaly around $120^{\circ}E$ longitude[J]. Advances in Space Research, 2009, 43: 1762–1769.
- 9 Liu H X, et al. Solar activity dependence of the electron density in the equatorial anomaly regions observed by CHAMP[J]. Journal of Geophysical Research, 2007, 112, A11311, doi:10.1029/2007JA012616.
- 10 王慧文. 偏最小二乘回归方法及其应用[M]. 北京: 国防工业出版社, 1999. (Wang Huiwen. Partial least square regression and application[M]. Beijing: National Defense Industry Press, 1999)