

电离层电子总含量不同时间尺度的预报模型研究*

盛 峰^{1)2)†}

1) (解放军理工大学气象海洋学院, 南京 211101)

2) (中国科学院空间天气学国家重点实验室, 北京 100190)

(2012年3月29日收到; 2012年5月16日收到修改稿)

电离层对无线电通信、卫星导航有重要的影响, 因此对电离层电子总含量 (total electron content, TEC) 的预报研究十分重要, 而目前国际上的各种经验电离层预报模型的精度只有 60% 左右, 不能满足实际需求. 本文提出一种新的 TEC 预报模型: 利用经验正交函数对 TEC 数据进行时空分解, 利用遗传算法结合混沌预测的思想对时间场系数进行非线性时间序列预测, 从而达到对 TEC 数据预报的目的. 实验结果表明, 此方法可较好地对 TEC 数据进行不同时间尺度的预测, 提前 1, 2, 4, 7 d 的预报精度分别达到 0.32, 0.48, 0.68, 0.94 TECU.

关键词: 电离层电子总含量, 经验正交函数分解, 时间序列, 遗传算法

PACS: 94.20.Cf, 02.60.Ed

1 引言

在整个日地空间环境中, 电离层上连磁层、下接中层大气, 并通过各种形式与上下层紧密地耦合在一起. 剧烈的太阳活动爆发如太阳耀斑、日冕物质抛射和地震、台风及灾难性天气等都会造成电离层的扰动. 同时电离层还是人类空间活动的主要区域, 其间所发生的各种过程, 特别是能量、动量和质量的传输与剧烈扰动过程, 可能对人类的军事活动及日常生活产生重要作用^[1]. 电子浓度总含量 (total electron content, TEC) 是描述电离层形态和结构的重要参量, 电离层 TEC 的测量和预报就显得特别重要^[2-4]. 目前世界上有几个电离层电子含量监测中心, 为满足实际需要, 他们几乎每天公布地面上空的电子总含量, 同时根据已测得的电子总含量实现对未来电子总含量的预报. 电离层的预报模型有 Klobuchar 模型、Bent 模型等, 现报模型有 IRI 模型等. 在 GPS 单频观测时, 一般采用 Klobuchar 模型, 它预报出电离层电子总含量

的精度一般在 50%—60%, 最乐观的情形下也不超过 75%, 且对夜间的预报具有明显的不合理性. 因此, 有必要对电离层电子总含量的精确预报模型进行研究.

在实际应用中, 电离层预报是对未来时刻地面上空一定高度网格点的电离层电子总含量的预报. 目前国际上通常是每两小时给出经度方向 180° W 到 180° E, 间隔 5°, 纬度方向从纬度 87.5° S 到 87.5° N, 间隔 2.5° 的电离层电子总含量, 这样每两小时全球共有 5183 (73 × 71) 个网格点. 对这 5183 个网格点上的电子总含量直接进行预报在理论上是可以考虑的一种选择, 但是若直接进行预报, 由于没有考虑 TEC 数据变化的物理特性, 其预报误差将较大. 基于以上问题, 首先将 5183 个网格点上的电子总含量直接观测值从地固坐标系转换到日固坐标系下, 以适应 TEC 数据分布的物理场特性, 然后在日固坐标系下将 TEC 数据进行经验正交函数 (EOF) 分解, 利用遗传算法结合混沌预测的思想对时间场系数进行非线性时间序列预测^[5,6], 从而达到对 TEC 数据间接预报的目的. 实

* 国家自然科学基金 (批准号: 41105013)、江苏省自然科学基金 (批准号: BK2011122) 和空间天气学国家重点实验室开放课题 (批准号: 201120FSIC-03) 资助的课题.

† E-mail: 19994035@sina.com

验结果表明预报精度优于传统的直接预报方法和线性时间序列预测方法.

2 预报模型的原理

预报模型的实现过程主要有三个步骤, 第一步首先对 TEC 数据进行 EOF 分解, 第二步利用遗传算法结合混沌预测的思想对时间场系数进行非线性时间序列预测^[7-9], 第三步利用确定的最优重构阶数对 TEC 数据 EOF 分解后的空间场和预报后的时间场进行重构, 以达到预测和降噪的目的^[10].

2.1 EOF 分解和重构去噪

设 \mathbf{X} 是原始的观测数据矩阵, 其中 \mathbf{X} 的维数为 $M \times N (M > N)$. M 为观测数据 \mathbf{X} 空间上的维数, N 为观测数据 \mathbf{X} 时间上的维数. \mathbf{X} 矩阵的每一列对应一次观测数据, 其数据点大小为 M , N 代表共有 N 次观测数据. 对 \mathbf{X} 的奇异值分解 SVD 可表示如下: $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$. \mathbf{U} 的维数为 $M \times r$, 代表 \mathbf{X} 的空间分解量; \mathbf{V} 的维数为 $r \times N$, 代表 \mathbf{X} 的时间分解量; \mathbf{D} 为特征值组成的对角矩阵, 维数为 $r \times r (r \leq \min(M, N))$. 在知道 $\mathbf{U}, \mathbf{D}, \mathbf{V}$ 的值后, 可通过 \mathbf{X} 的重构公式来得到 \mathbf{X} 的重构值 $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T = \sum_{i=1}^k \rho_i \mathbf{U}_i \mathbf{V}_i^T$. ρ_i 代表第 i 阶特征值, k 为重构数据时所需的阶数, $k \leq r$.

EOF 重构数据时所取的阶数为 n , 但 n 具体取为何值为最优, 是一个极为关键的问题. n 值取得太小, 所重构的数据不足以反映数据内在的物理变化特征; n 取得过大所引入的误差信息会加大, 同时会加大计算量^[11,12]. 对于需要重构的 TEC 数据, n 取值标准为前面 n 阶 EOF 分解的特征值方差贡献率达到整个特征值的 95%, 这样的取值标准既可反映数据内在的物理变化特征, 又可避免 n 取得过大所引入的误差信息加大, 从而达到数据去噪的目的.

2.2 时间场系数的非线性预测方法

第二步的工作就是利用遗传算法结合混沌预测的思想对时间场系数进行非线性时间序列预测, 给定一时间序列 $\{A_n(t)\}, t = 1, \dots, T, n = 1, \dots, N_r$, 首先采用从时间序列来计算关联维数以判断系统混沌特征的 G-P (Grassberger-Procaccia) 算法, 计算出关联维 d , 确定嵌入维数 m

($m \geq 2d + 1$), 对于给定的时间序列 $\bar{A}_n(t) = Y_n[A_n(t-\tau), A_n(t-2\tau), \dots, A_n(t-m\tau)]$, 重构 m 维相空间表示为

$$\bar{A}_n(t) = Y_n[A_n(t-\tau), A_n(t-2\tau), \dots, A_n(t-m\tau)], \quad (1)$$

其中 $n = 1, \dots, N_r$, 表示变量的个数, m 表示嵌入维数, 如果能求出 Y_n 的表达式, 那么就可利用 (1) 式对时间序列 $A_n(t)$ 进行预测.

Y_n 的表达式采用遗传算法来确定, 首先在 $[-1, 1]$ 之间区间随机产生 $[A_j(t-\tau), A_j(t-2\tau), \dots, A_n(t-m\tau)]$ 系数, 系数之间的运算符用 $+, -, \times, \div$ 来进行描绘, 和系数一样, $+, -, \times, \div$ 的运算符也是随机产生的, 对于给定的时间序列 $\{A_n(t)\}, t = 1, \dots, T, n = 1, \dots, N_r$, 系数和运算符在遗传算法的运算过程中获取最优值, 在确定最优的系数和运算符后可得出 Y_n 的表达式, 从而实现对时间序列 $A_n(t)$ 的不同时间尺度的预测.

整个预报模型的预报流程见图 1.

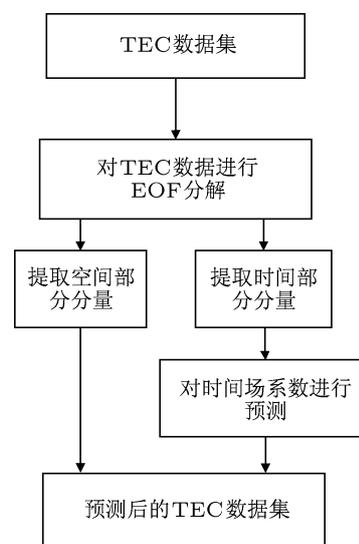


图1 TEC 数据的预报流程

3 实验所采用的数据及坐标系的转换

实验所用的数据为全球电离层 TEC 数据, 将 TEC 数据分布的坐标系从地固坐标系转换到日固坐标系, 以更符合 TEC 数据分布的物理场特性^[13].

3.1 实验所用的 TEC 数据

实验所用的数据采用中国科学院上海天文台提供的 2007 年全球电离层 TEC 数据, 数据为 2 h 一次的全球 TEC 数据, 全年一共 4380 次观测数据. 区域范围为经度 180 °W 到 180 °E, 纬度 87.5 °S 到 87.5 °N, 观测数据像素点大小为 $5183 \times 365 \times 12$, 分辨率大小为经度方向 5°, 纬度方向 2.5°.

3.2 地固坐标系与日固坐标系的转换

TEC 地图绘制的具体方法如下: 根据电离层与太阳辐射的密切相关性, 采用所谓“日固坐标系”, 将通常的 3 维分布 TEC (Lat, Lon, UT) 缩并为 2 维分布 TEC (Lat, LT), 其中

$$LT = UT + Lon/15, \quad (2)$$

这里, Lat, Lon 分别为地理纬度和经度 (以度为单位), UT, LT 分别为世界时和当地时 (以小时为单位). 采用日固坐标系的主要优点是: 1) 通过降维简化了计算, 同时提高了计算结果的稳定性; 2) 充分利用 GPS 观测中的高时间分辨率弥补空间分辨率的不足, 从而可能提高格点 TEC 的估算精度.

4 实验预报结果分析

预报实验中研究数据采用整年 365 个时间样本, 用前面 300 个时间样本建模, 用后 65 个做预报

检验^[14,15]. 在建模之前, 按前面第三节所述对每一组时间样本进行相应的坐标转换处理.

4.1 直接预报方法的精度分析

采用直接预报方法时, 没有考虑 TEC 数据变化的空间特性和时间特性, 只是将同一格点上的 TEC 看成一时间序列进行预测. 此时的预报模型为时间序列预测中经常采用的 ARIMA (auto regressive integrated moving average) 模型, 即自回归求和滑动平均模型, 是由美国统计学家在 20 世纪 70 年代提出的时间序列模型, 基本思想是某些时间序列是依赖于时间 t 的一组变量, 构成该时间序列的单个序列值虽然有不稳定性, 但整个时间序列的变化却有一定的规律可循, 可以采用对应的数学模型来描述. ARIMA 是目前应用最为广泛时间序列预测模型, 适用于处理非平稳时间序列.

图 2(a)—(d) 分别列出预测时间尺度为 1, 2, 4, 7 d 的预测结果误差分布图, TEC 值用色表表示, 单位为 TECU. 图 2(a)—(d) 对应的起始预测时间为 2007 年第 300 d 的 TEC 数据. 在直接采用 ARIMA 模型对 TEC 时间序列进行预报的情况下, 由于没有充分利用各格点 TEC 值内在的相关性, 只是简单地将各格点的 TEC 值看成数学概念上的时间序列, 预测的精度不高, 且随着预测时间尺度的增大, 预测误差迅速增加. 从图 2 中可以看

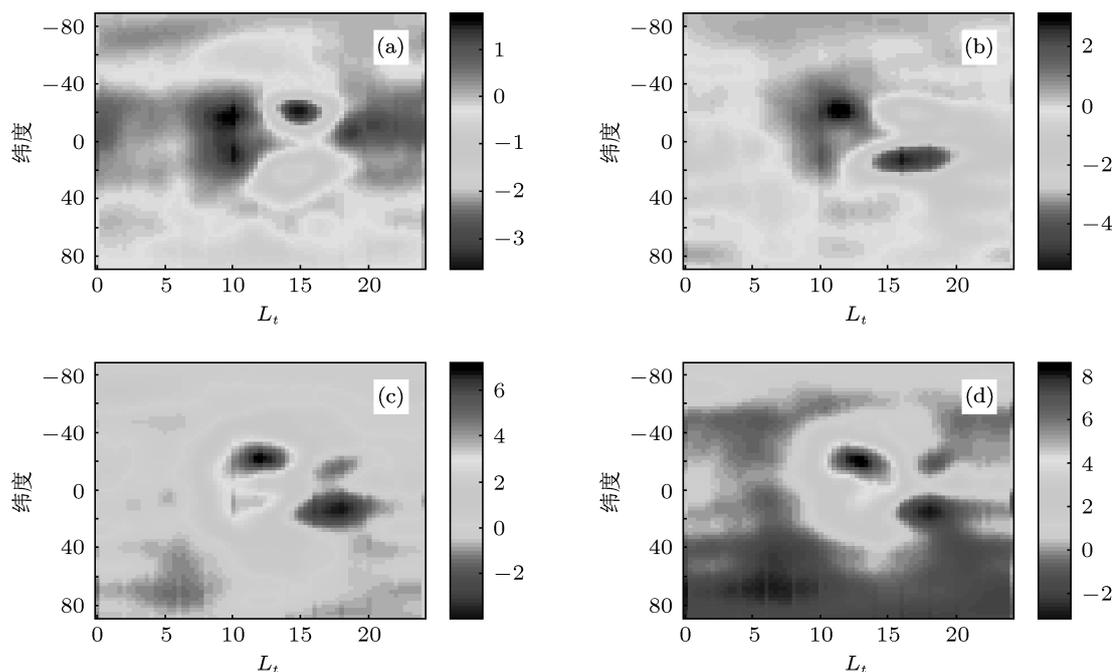


图 2 直接预报方法的误差分布图 (a), (b), (c), (d) 分别表示预测时间尺度为 1, 2, 4, 7 d

出误差较大的区域对应电离层不均匀结构及电离层异常的区域,这和赤道两边容易出现电离不均匀结构,出现时间都集中在电离层密度最大的 10 时到 17 时之间的物理现象完全符合.

直接预报方法对不同时间尺度的详细预测精度见表 1. 表 1 中预报时长的单位为天,用 d 表示; Δ 表示预报值与实测值之差,单位均为 TECU;表中分别统计 Δ 在不同误差区间的百分比统计情况 [16].

表 1 直接预报方法的详细误差分布情况

预报时长 /d	平均偏差 /TECU	$\Delta < 1$ /TECU	$1 \leq \Delta < 2$	$2 \leq \Delta < 3$	$\Delta \geq 3$
1	0.41	91.12%	7.79%	0.71%	0.37%
2	0.57	83.83%	11.04%	2.97%	2.16%
3	0.52	87.27%	8.41%	2.55%	1.78%
4	0.76	81.11%	10.11%	4.44%	4.34%
5	0.98	70.36%	19.47%	3.97%	6.19%
6	1.12	61.62%	25.83%	4.94%	7.60%
7	1.16	56.51%	29.02%	9.13%	5.34%

4.2 时间场系数采用线性时间序列预测方法时的精度分析

对时间场系数采用线性时间序列预测方法时,通过对时间场 V 的系数进行预测,达到对整个 TEC 物理场预测的目的,此时对 V 系数的预报模型依然采用 ARIMA 模型. 图 3(a)—(d) 分别列出预测时间尺度为 1, 2, 4, 7 d 的预测结果误差分布, TEC 值用色表值表示,单位为 TECU. 同图 2 一样,图 3 对应的起始预测时间为 2007 年第 300 d 的 TEC 数据.

在采用 ARIMA 模型对时间场系数时间序列进行预报的情况下,此时通过 EOF 分解和重构,预测后的数据能较好地考虑各格点 TEC 值内在的相关性,整个 TEC 数据的预测误差相对于图 2 减少很多. 对比图 2 和图 3 中的各个子图可以看出,尽管图 3 相对于图 2 预测精度有较好的提高,但是预测误差较大的区域所对应的地理位置和形态结构大致相同,说明误差较大的区域总是对应电离层不均匀结构及电离层异常的区域 [17,18],和预测方法的选取没有关系.

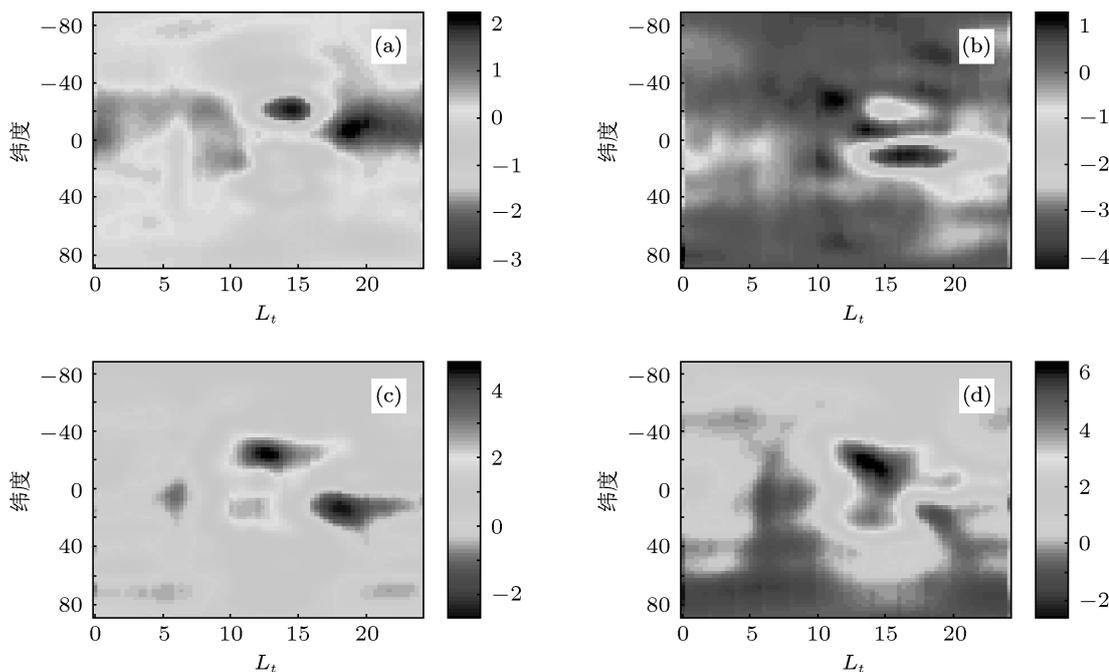


图 3 时间场系数采用线性时间序列预测方法的误差分布图 (a), (b), (c), (d) 分别表示预测时间尺度为 1, 2, 4, 7 d

时间场系数采用线性时间序列预测方法对不同时间尺度的详细预测精度见表 2. 表 2 中预报时长的单位为天, 用 d 表示; Δ 表示预报值与实测值之差, 单位均为 TECU; 表中分别统计 Δ 在不同误差区间的百分比统计情况.

4.3 时间场系数采用非线性预测方法时的精度分析

对时间场系数采用非线性预测方法时, 此时的预报模型采用遗传算法结合混沌预测以实现对时间场 V 系数的非线性预测. 图 4(a)—(d) 分别列出预测时间尺度为 1, 2, 4, 7 d 的预测结果误差分布图, TEC 值用色表值表示, 单位为 TECU. 同前面图 2 和图 3 一样, 图 4 对应的起始预测时间为 2007

年第 300 d 的 TEC 数据. 在采用非线性时间序列预测方法时, 相比时间场系数采用线性时间序列预测方法预测后的数据精度有进一步提高, 误差也相应地减少, 特别是较长时间尺度的预报精度相比前两种方法都有较大的提高. 提前一天预报精度由 0.35 TECU 提高为 0.32 TECU, 提前七天预报精度由 1.13 TECU 提高为 0.94 TECU, 改进的效果在 10%—20%左右. 通过具体的程序分析, 改进效果比较明显的地方时间集中在 10:00—20:00 h 之间, 而纬度方向的整体改进效果分布比较平均. 误差的分布情况与前面所提的两种预测方法相似, 误差较大的区域也是对应电离层不均匀结构及电离层异常的区域. 今后可从研究电离层异常结构的物理机制和分析相应的观测数据入手, 以求更好地对电离层异常结构进行模拟.

表 2 时间场系数采用线性时间序列预测方法的详细误差分布情况

预报时长/d	平均偏差/TECU	$\Delta < 1$ /TECU	$1 \leq \Delta < 2$	$2 \leq \Delta < 3$	$\Delta \geq 3$
1	0.35	92.40%	6.50%	1.02%	0.08%
2	0.52	92.05%	4.96%	1.60%	1.39%
3	0.51	90.37%	8.18%	1.41%	0.04%
4	0.72	78.49%	15.92%	4.21%	1.39%
5	0.95	61.86%	26.64%	7.27%	4.23%
6	1.07	59.44%	28.84%	5.79%	5.92%
7	1.13	58.07%	28.84%	5.44%	7.64%

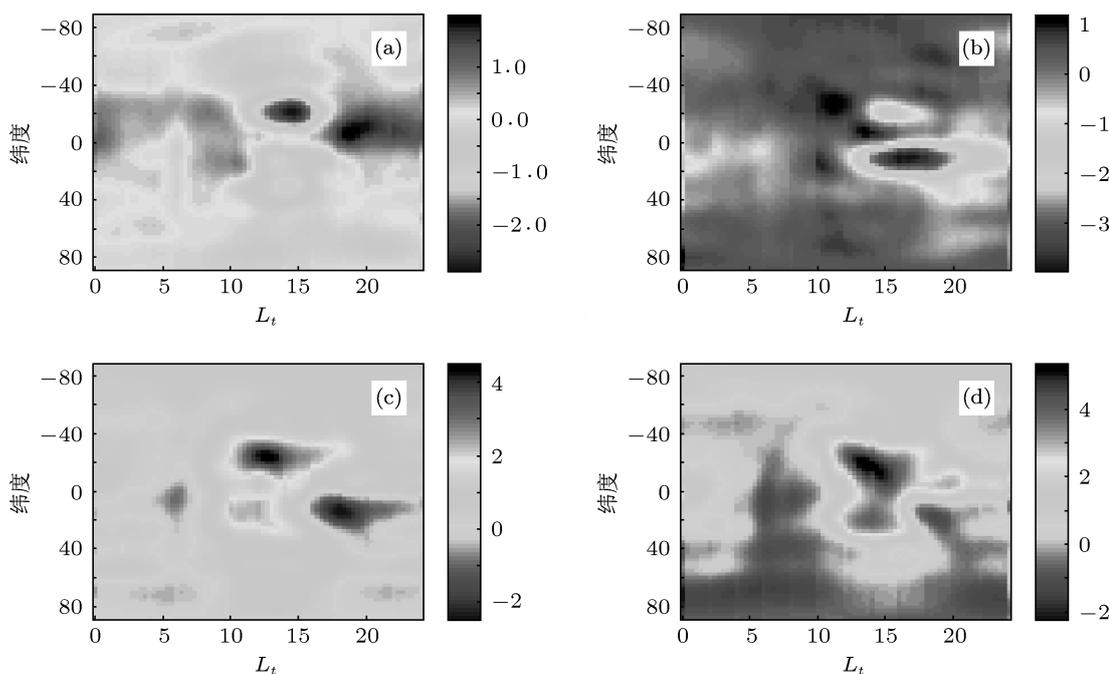


图 4 时间场系数采用非线性时间序列预测方法的误差分布图 (a), (b), (c), (d) 分别表示预测时间尺度为 1, 2, 4, 7 d

时间场系数采用非线性时间序列预测方法对不同时间尺度的详细预测精度见表 3. 表 3 中预报时长的单位为天, 用 d 表示; Δ 表示预报值与实测值之差, 单位均为 TECU; 表中分别统计 Δ 在不同误差区间的百分比统计情况. 综合对比表 1、

表 2、表 3, 当时间场系数采用非线性时间序列预测方法时, 不论是平均偏差还是误差的分布情况, 相对前两种方法都有较好的改进; 对比表 2 和表 3, 可以看出随着预报时长的增加, 改进的效果更为明显.

表 3 时间场系数采用非线性时间序列预测方法的详细误差分布情况

预报时长 /d	平均偏差 /TECU	$\Delta < 1$ /TECU	$1 \leq \Delta < 2$	$2 \leq \Delta < 3$	$\Delta \geq 3$
1	0.32	93.94%	5.38%	0.68%	0
2	0.48	92.55%	4.84%	1.54%	1.06%
3	0.46	93.34%	5.85%	0.81%	0
4	0.68	80.47%	14.74%	3.59%	1.20%
5	0.82	70.11%	21.80%	5.40%	2.68%
6	0.90	67.18%	23.98%	5.21%	3.63%
7	0.94	65.43%	24.23%	5.36%	4.98%

5 结论

本文提出基于 EOF 分解和非线性时间序列预测的 TEC 不同时间尺度预报模型, 并将其与传统的直接预报方法和线性时间序列预测方法进行对比, 改进效果较为明显, 特别是预报时间尺度较大时, 改进的效果更为明显. 从预测精度统计表中可看出, 表 2 相对于表 1, 整体改进效果在 10% 左右; 而表 3 相对于表 2, 整体改进效果又在 10% 左右. 说明相对于传统的算法, 本文所提的算法改进效果在 20% 左右. 此算法使用简单, 并且不需要任何先验的误差统计信息, 在实际使用中也证明了此算法的可靠性与有效性. 此算法充分考虑 TEC 数据时空分解后

时间场系数的混沌特性, 构建混沌预报方程, 通过遗传算法求出最优的预报方程, 从而实现对 TEC 数据的预报, 可较好地体现 TEC 数据的物理场特性, 克服当前时间序列预报方法中没有考虑 TEC 数据物理背景的缺陷. 在下一步研究工作中, 如需要处理更大型的数据矩阵时, 还需要做一些附加的优化工作, 例如对 EOF 分解重构的迭代收敛准则进行改进. 在以后研究工作中还要考虑利用此算法对其他一些电离层遥感物理量进行预报, 并可考虑同时对电离层几种遥感物理量进行预报^[19], 以充分利用电离层不同物理量之间的相关性, 对预报结果进行改善, 进一步提高整体的预报精度.

- [1] Yan H J 1996 *The Astronomical Journal* **112** 1312
- [2] Liu R Y, Liu G H, Wu J 2008 *Chin. J. Geophys.* **51** 300 (in Chinese) [刘瑞源, 刘国华, 吴健 2008 地球物理学报 **51** 300]
- [3] Li Z G, Cheng Z Y, Feng C G 2007 *Chin. J. Geophys.* **50** 327 (in Chinese) [李志刚, 程宗颐, 冯初刚 2007 地球物理学报 **50** 327]
- [4] Yue X A, Wan W X, Liu L B 2010 *Chin. J. Geophys.* **53** 787 (in Chinese) [乐新安, 万卫星, 刘立波 2010 地球物理学报 **53** 787]
- [5] Everson R, Cornillon P, Sirovich L 2002 *J. Phys. Ocean.* **27** 468
- [6] Bennett A 2002 *Inverse Modeling of the Ocean and the Atmosphere* (Cambridge: Cambridge University Press)
- [7] Beckers J, Rixen M 2003 *J. Atmos. Ocean Technol.* **20** 1839
- [8] Toumazou V, Cretaux J 2001 *Mon. Weath. Rev.* **125** 1243
- [9] Alvarez A 2003 *J. Atmos. Ocean Technol.* **20** 717
- [10] Sheng Z, Shi H Q, Ding Y Z 2009 *Adv. Ocean. Sci.* **27** 243 (in Chinese) [盛峥, 石汉青, 丁又专 2009 海洋科学进展 **27** 243]
- [11] Alvarez A, Barth A, Rixen M 2005 *Ocean Model.* **17** 325
- [12] Beckers J, Barth A, Alvarez A 2006 *Ocean Sci.* **16** 183
- [13] Guo P, Xu X, Zhang G X 2011 *J. Atmos. Terr. Phys.* **69** 947
- [14] Sheng Z 2011 *Acta Phys. Sin.* **60** 119301 (in Chinese) [盛峥 2011 物理学报 **60** 119301]
- [15] Sheng Z, Chen J Q, Xu R H 2012 *Acta Phys. Sin.* **61** 069301 (in Chinese) [盛峥, 陈加清, 徐如海 2012 物理学报 **61** 069301]
- [16] Sheng Z, Fang H X 2012 *Chin. Phys. B* **21** DOI: 10.1088/1674-1056/21/2/02
- [17] Wu Y Y, Hong Z J, Guo P, Zheng J 2010 *Chin. J. Geophys.* **53** 639

[18] Guo P, Kuo Y H, Sokolovskiy S V, Lenschow D H 2011 *J. Atmos.*

Sci. DOI: 10.1175/2011JAS3612.1

[19] Wei M, Qin X, Wang X H, Jiang W M 2007 *Trans. Atmos. Sci.*

30 736 (in Chinese) [魏鸣, 秦学, 王啸华, 蒋维楣 2007 南京气象学院学报 30 736]

Research on different time-scale prediction models for the total electron content*

Sheng Zheng^{1)2)†}

1) (*Institute of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 211101, China*)

2) (*State Key Laboratory of Space Weather, Chinese Academy of Sciences, Beijing 100190, China*)

(Received 29 March 2012; revised manuscript received 16 May 2012)

Abstract

In the solar-terrestrial space environment, the ionosphere couples tightly with the upper magnetic layer as well as the lower middle atmosphere in various forms. Meanwhile, the ionosphere can affect radio-communication and satellite navigations, so the research on ionosphere prediction model is very important. Now, the accuracy of statistic prediction mode is about 60%, but cannot meet the practical requirements. In order to solve the problem, the prediction model of total electron content (TEC) data is achieved in three major phases: decomposition of the spatiotemporal variability of the TEC data, noise reduction of the encoded space, and time variability and the prediction, by a nonlinear forecasting technique of the time variability. Experiments show that the new prediction model is better than traditional prediction model. The prediction data shows realistic features and a reliable physical distribution, and the relative accuracies of prediction for 1, 2, 4, and 7 d obtained by our method is 0.32, 0.48, 0.68 and 0.94 TECU.

Keywords: total electron content, empirical orthogonal functions decomposition, time series, genetic algorithm

PACS: 94.20.Cf, 02.60.Ed

* Project supported by the National Natural Science Foundation of China (Grant No. 41105013), the Natural Science Foundation of Jiangsu, China (Grant No. BK2011122), and the Specialized Research Fund for State Key Laboratories, China (Grant No. 201120FSIC-03).

† E-mail: 19994035@sina.com