0254-6124/2009/29(2)-195–07 *Chin. J. Space Sci.* 空间科学学报

Feng Jing, Liu Wen, Jiao Peinan, Fan Junmei. Autocorrelation method for interpolation of ionospheric characteristic parameters. *Chin. J. Space Sci.*, 2009, **29**(2): 195∼201

电离层特征参量的自相关原理插值方法*

冯 静 柳 文 焦培南 凡俊梅

(中国电波传播研究所 电波环境特性及模化技术国家重点实验室 青岛 266107)

摘 要 通过选用合适的电离层平稳性参数, 建立相应的正定自相关系数模型, 利用自相关分析原理, 提出了一种针对电离层特征参量历史缺失数据插值处理的新方法. 该方法能够提高 Muhtarov 和 Kutiev 在 1999 年提出的自相关系数法的插值精度, 通常情况下可以把误差降低 1 到 2 个百分点以上, 有时甚至能降低接近 9 个百分点, 在很大程度上改善了对电离层历史缺失数据的插值处理效果. 此外, 本文还对插值误差随季节、太阳活动性和地理纬度等的变化规律进行了分析.

关键词 电离层; 自相关; 插值

中图法分类号 P 353

Autocorrelation Method for Interpolation of Ionospheric Characteristic Parameters

FENG Jing LIU Wen JIAO Peinan FAN Junmei

(*National Key Laboratory of Electromagnetic Environment, China Research Institute of Radiowave Propagation, Qingdao* 266107)

Abstract The autocorrelation method considers the ionospheric characteristic of interest as a realization of a random process, modeled as the sum of a periodical component and a random Gaussian process, assuming stationarity during the period of interest. Then the process is defined by its autocorrelation function. Based on the autocorrelation analysis theory, a proper steady ionospheric parameter is uesd and the corresponding autocorrelation model which is positive definite is also established in this paper. On this basis, an interpolation method for ionospheric missing data is presented. This method improves the interpolation accuracy of the method given by Muhtarov and Kutiev in 1999. It generally decreases the error by 1∼2 percent, and even by nearly nine percent sometimes. Therefore it largely advances the effect of interpolation for ionospheric historical missing data. Furthermore, the variation rules of the interpolation error along with seasons, solar activites and latitudes are analysed.

Key words Ionosphere, Autocorrelation, Interpolation

* 国家自然科学基金项目 (60771049) 和国家基础研究项目及国家重点实验室基金项目 (9140C0801070602) 共同资助

2008-04-29 收到原稿, 2008-12-19 收到修定稿

E-mail: hellen1917@163.com

1 引言

电离层 F_2 层临界频率 f_0F_2 是描述电离层形态最重要的特征参量之一, 广泛应用于电离层理论研究和工程应用中. 目前, f_0F_2 参数主要由基于地面的电离层垂直探测得到, 全球范围内建立了广泛的电离层垂直探测网, 积累了大量历史数据. 但是, 常常由于电离层本身的原因或者其他因素 (如探测设备故障等) 的影响, 使某些时刻或时段 f_0F_2 的历史数据存在缺失现象, 这给后面的研究带来很多不便. 近年来, 针对如何对缺失数据进行插值, 人们进行了一些研究, 其中较为简单实用的是月中值插值法, 即利用对应时刻的月中值来代替缺失数据, 但是该方法在精度上常常不能令人满意.

1999 年, Muhtarov 和 Kutiev[1] 提出了以 f_0F_2 为参数的自相关系数插值方法 (下文简称 f_0F_2 插值), 该方法以平稳随机过程为背景, 使用简便, 并且具有相当的精度, 在很大程度上优于月中值插值法. 但是实际的电离层并非是一个理想的平稳随机过程, 而是带有很大的非平稳特性, 这就降低了自相关分析原理的适用性. 刘瑞源等[2] 在电离层短期预报中使用了平稳性电离层参数, 使得预报精度有了较大提高, 但参数的自相关系数模型仍旧沿用文献 [1] 中建立的 f_0F_2 参数的自相关系数模型, 该模型并不能完全适合平稳性电离层参数. 本文针对文献 [2] 中提出的平稳性电离层参数建立了相应的正定自相关系数模型, 形成了新的插值方法, 提高了插值精度, 并利用中国部分电离层观测站的数据给出了该方法插值精度的定量估计.

2 z_{DF} 自相关分析插值方法

2.1 自相关分析插值原理

自相关分析插值方法的基本思想是将电离层特征参量 z 视为平稳时间序列, 假设 t 时刻有数据缺失, 则可以利用其前后时刻的有效观测数据以及它们之间的时间相关关系, 采用线性滤波器方法对其进行插值计算, 插值方法的基本表达式如下:

$$z(t)=\sum_{j=0}^{n-1}\lambda_j z(t_j). \tag{1}$$

上式中, $z(t)$ 为 t 时刻缺失数据的插值结果, $z(t_j)$ $(j=0,1,\cdots,n-1)$ 为选取的一组有效观测值 (通常是在缺失数据前后 2 到 3 天的数据范围内选取自相关系数较高的数据), λ_j 为 $z(t_j)$ 对应的加权系数, 在无偏和最优的条件下, 利用拉格朗日乘数法求解函数极值, 最终可以得到下列方程组[1], 从中可以解出加权系数 λ_j, 有

$$\begin{aligned}&\sum_{j=0}^{n-1}\lambda_j\rho(t_i-t_j)+\mu=\rho(t-t_i),\\ &\qquad i=0,1,\cdots,n-1;\\ &\sum_{j=0}^{n-1}\lambda_j-1=0.\end{aligned} \tag{2}$$

其中, μ 为拉格朗日系数, $\rho(\tau)$ 为函数 $z(t)$ 的自相关系数. 将解得的加权系数 λ_j 代入式 (1), 便可以计算得到 t 时刻的插值结果 $z(t)$.

基于自相关分析原理做插值处理时必须满足以下两个条件:

(1) 时间序列 $z(t)$ 为 (或近似为) 平稳时间序列;

(2) 为了保证 λ_j 解的唯一性, 要求 $\rho(\tau)$ 必须具有正定矩阵形式.

2.2 z_{DF} 序列自相关系数 $\rho(\tau)$ 的数学模型

事实上, 电离层 f_0F_2 序列并非是一个理想的平稳时间序列, 基于这一点, 文献 [2] 提出了两种平稳性电离层参数用以取代 f_0F_2, 分别是

$$\begin{aligned}z_{\mathrm{DF}}&=f_0F_2-\overline{f_0F_2},\\ z_{\mathrm{RDF}}&=(f_0F_2-\overline{f_0F_2})/\overline{f_0F_2},\end{aligned}$$

其中, $\overline{f_0F_2}$ 为相应时刻的月平均值.

从时间序列的平稳性角度考虑, z_{DF} 和 z_{RDF} 更适合自相关分析插值. 在文献 [2] 的基础上, 我们利用历史探测数据对两种参数的插值精度进行了统计分析, 结果表明, 利用参数 z_{DF} 进行插值的效果优于利用参数 z_{RDF} 的效果. 因此在后续工作中, 本文选用 z_{DF} 值作为自相关分析插值的输入参数 z.

实际应用中, 自相关系数一般是利用有限个实测数据根据样本计算公式计算得到, 结果不一定满足正定性要求. 文献 [1] 中根据实测 f_0F_2 的自相关特征, 建立了如下 f_0F_2 的正定自相关系数模型:

$$\rho_1(\tau)=(1-A_{24}-A_{12})\exp\Big(-\frac{|\tau|}{\tau_0}\Big)+$$

$$A_{24}\cos\left(\frac{2\pi}{24}\tau\right)+A_{12}\cos\left(\frac{2\pi}{12}\tau\right). \quad (3)$$

其中, A_{12}, A_{24} 和 τ_0 为待定系数, 可以使用最小二乘法拟合样本计算值来确定.

文献 [2] 利用 $z_{\rm DF}$ 参数进行短期预报时, 仍然沿用了式 (3) 描述的自相关系数模型. 大量数据试验发现该模型并不完全适合参数 $z_{\rm DF}$ 的自相关特征. 为此, 我们建立了另外一种具有正定形式的模型来描述 $z_{\rm DF}$ 序列的自相关系数如下:

$$\rho_2(\tau)=\frac{\sum_{i=1}^{N-\tau}z(t_i)z(t_i+\tau)}{\sum_{i=1}^{N}z^2(t_i)}. \quad (4)$$

其中, $z(t)$ 为 t 时刻对应的 $z_{\rm DF}$ 值, N 为一个月的数据总数. 该模型实际上是具有零均值的 $z_{\rm DF}$ 平稳随机序列的理论自相关系数的一个有偏估计, 可以证明 $\rho_2(\tau)$ 是正定的.

分析一段长度为 N 的 $z_{\rm DF}$ 时间序列 $\{z(t)\}$, $t\in\{1,2,\cdots,N\}$, 事实上, 当 $t>N$ 或 $t\leqslant 0$ 时, 有 $z(t)=0$. 对于任意 m 个不全为零的实数 $\alpha_1,\alpha_2,\cdots,\alpha_m$ 有

$$\begin{aligned}
&\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\rho_2(\tau)=\\
&\quad\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\left[\frac{\sum_{t=1}^{N-|j-i|}z(t)z(t+|j-i|)}{\sum_{t=1}^{N}z^2(t)}\right]=\\
&\quad\frac{1}{\sum_{t=1}^{N}z^2(t)}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\sum_{t=1}^{N-|j-i|}z(t)z(t+|j-i|)=\\
&\quad\frac{1}{\sum_{t=1}^{N}z^2(t)}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\sum_{t=-\infty}^{\infty}z(t)z(t+|j-i|)=\\
&\quad\frac{1}{\sum_{t=1}^{N}z^2(t)}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\sum_{t=-\infty}^{\infty}z(t)z(t+j-i)=\\
&\quad\frac{1}{\sum_{t=1}^{N}z^2(t)}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\sum_{t=-\infty}^{\infty}z(t+i)z(t+j)=\\
&\quad\frac{1}{\sum_{t=1}^{N}z^2(t)}\sum_{t=-\infty}^{\infty}\Big[\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j z(t+i)z(t+j)\Big]=\\
&\quad\frac{1}{\sum_{t=1}^{N}z^2(t)}\sum_{t=-\infty}^{\infty}\Big[\sum_{i=1}^{m}\alpha_i z(t+i)\Big]^2>0.
\end{aligned}$$

以上证明了 $\rho_2(\tau)$ 具有正定矩阵形式.

$\rho_2(\tau)$ 可以通过样本数据直接计算得到, 计算简便, 而且能够改善 $\rho_1(\tau)$ 对一些数据的 $z_{\rm DF}$ 自相关系数拟合不好的情况. 图 1 和图 2 所示为使用长春站 (43.90°N, 125.40°E) 1990 年 1 月和兰州站 (36.03°N, 103.90°E) 1996 年 12 月的 $z_{\rm DF}$ 数据, 利用模型 $\rho_1(\tau)$, $\rho_2(\tau)$ 和样本计算公式计算得到的自相关系数 (样本计算公式采用文献 [1] 中的 (12) 式). 从图 1 中可以看出, 模型 $\rho_2(\tau)$ 比模型 $\rho_1(\tau)$ 更贴近自相关系数的样本计算值.

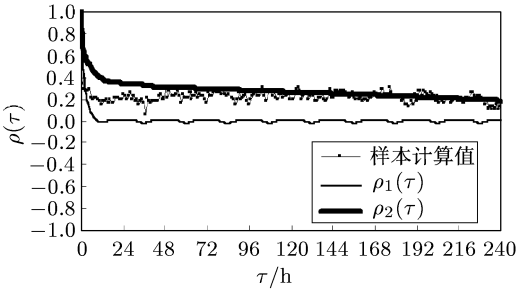


图 1　长春站 1990 年 1 月 $z_{\rm DF}$ 自相关系数模型比较

Fig.1　Comparison of the autocorrelation coefficients models of $z_{\rm DF}$ in January 1990 from Changchun

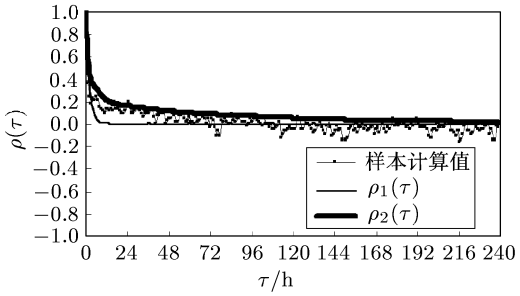


图 2　兰州站 1996 年 12 月 $z_{\rm DF}$ 自相关系数模型比较

Fig.2　Comparison of the autocorrelation coefficients models of $z_{\rm DF}$ in December 1996 from Lanzhou

2.3　模型 $\rho_2(\tau)$ 插值效果验证

利用 $z_{\rm DF}$ 自相关分析插值方法, 分别使用模型 $\rho_1(\tau)$ 和 $\rho_2(\tau)$ 对长春站 1990 年 2 月和兰州站 1996 年 12 月的 f_0F_2 数据进行插值分析. 为了使计算结果具有统计意义上的说服力, 也为了了解不同的插值长度对插值精度的影响, 我们对一个月的数据 (缺失数据用对应时刻的月中值进行了替换) 人为地制造连续的缺失数据, 长度分别为 1 h, 2 h, $\cdots$, 24 h. 将一段固定的长度从月初逐渐移动到月末, 移动的步进等于该固定长度, 每次都对应着同等长度的连续缺失数据, 将每个缺失数据的插值计算值与实测值的偏差都记录下来, 以待最后统计该固定长度对应的插值误差. 以上操作

针对 24 种插值长度将重复 24 次, 最终计算出不同插值长度的误差. 定义插值绝对误差为插值计算值相对于实测值的均方根误差, 相对误差为每点实测值归一化的平均均方根误差, 表达式分别如下:

$$E_{\mathrm{RMSE}} = \sqrt{\frac{1}{N-1}\sum_{j=0}^{N-1}(P_i - O_i)^2}. \tag{5}$$

$$E_{\mathrm{RM}} = \sqrt{\frac{1}{N}\sum_{j=0}^{N-1}\left(\frac{P_i - O_i}{O_i}\right)^2}. \tag{6}$$

其中, P_i 表示某时刻的插值计算值, O_i 表示某时刻的实测值. 表 1 和表 2 给出了使用两种不同模型的插值相对误差, 为方便比较, 同时给出文献 [1] 中的 f_0F_2 插值方法对应不同插值长度的相对误差.

从表 1 和表 2 可知, 模型 $\rho_2(\tau)$ 能够提高模型 $\rho_1(\tau)$ 的插值精度, 有时可把误差降低 2 个百分点以上, 而比起 f_0F_2 插值优势更加明显. 从表中还可看出, 各种插值方法的插值精度都随着插值长度的增加而下降, 并逐渐趋于平稳.

2.4 两种模型综合使用

利用 z_{DF} 自相关分析法进行插值时, 发现有的数据适合使用模型 $\rho_1(\tau)$, 而有的数据用模型 $\rho_2(\tau)$ 插值效果更好. 很显然, 模型值拟合样本计算值的效果越好, 插值精度将越高. 我们用模型值与样本计算值之间的相关系数和绝对误差来评价模型的适用性. 相关系数反映了模型曲线和样本计算曲线之间的形状相似 (相关) 程度, 取值在 +1 和 −1 之间, 绝对值越大说明相关性越好, 值为正表示正相关, 值为负表示负相关; 绝对误差反映的是模型值对样本计算值的绝对偏离, 拟合无误差时等于 0. 将模型值与样本计算值之间的相关系数和绝对偏差分别记为 E_{PA} 和 E_{abs}, 定义如下:

$$E_{\mathrm{PA}} = \frac{\sum_{i=1}^{N}[(P_i - P_m)(O_i - O_m)]}{(N-1)S_{\mathrm{P}}S_{\mathrm{O}}}. \tag{7}$$

$$E_{\mathrm{abs}} = \frac{1}{N}\sum_{i=1}^{N}|P_i - O_i|. \tag{8}$$

其中, N 为数据点个数, P_i 为模型值, P_{m} 为模型值的平均值, O_i 为样本计算值, O_{m} 为样本计算值的平均值, S_{P} 为模型值的标准差, S_{O} 为样本计算值的标准差. 经过对实测数据的插值处理分析, 我们得出下列模型选择准则.

如果 $\rho_1(\tau)$ 和 $\rho_2(\tau)$ 中的某一个与另外一个相比:

表 1 长春站 1990 年 2 月 f_0F_2 数据使用模型 $\rho_1(\tau)$ 和 $\rho_2(\tau)$ 的 z_{DF} 自相关分析插值相对误差 (%)

Table 1 Relative errors of z_{DF} interpolation methods respectively using $\rho_1(\tau)$ and $\rho_2(\tau)$ in February 1990 from Changchun (%)

模型	插值长度/h																
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	22	24
$\rho_1(\tau)$	11.2	11.5	11.7	11.8	11.9	11.9	11.9	11.9	11.9	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0
$\rho_2(\tau)$	9.0	9.5	9.9	10.1	10.4	10.6	10.7	10.8	11.0	11.1	11.2	11.3	11.3	114	11.5	11.5	11.5
f_0F_2 插值	11.3	11.6	11.8	11.9	11.9	11.9	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0

表 2 兰州站 1996 年 12 月 f_0F_2 数据使用模型 $\rho_1(\tau)$ 和 $\rho_2(\tau)$ 的 z_{DF} 自相关分析插值相对误差 (%)

Table 2 Relative errors of z_{DF} interpolation methods respectively using $\rho_1(\tau)$ and $\rho_2(\tau)$ in December 1996 from Lanzhou (%)

模型	插值长度/h																
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	22	24
$\rho_1(\tau)$	10.2	10.3	10.5	10.6	10.7	10.7	10.8	10.9	11.0	11.1	11.2	11.2	11.3	11.3	11.4	11.4	11.4
$\rho_2(\tau)$	10.0	10.2	10.3	10.3	10.4	10.5	10.6	10.7	10.8	11.0	11.0	11.1	11.1	112	11.3	11.3	11.3
f_0F_2 插值	11.7	11.9	11.9	11.9	11.9	11.9	11.9	11.9	11.9	11.9	12.0	12.0	12.0	12.0	12.0	12.0	12.0

(a) $E_{\rm PA}$ 值偏大, 同时 $E_{\rm abs}$ 偏小, 则选择该模型; 若两个模型都不满足此条件, 转为准则 (b);

(b) $E_{\rm abs}$ 值偏大, 并且差值大于门限 T_1, 则选择另外一个模型; 若两个模型都不满足此条件, 转为准则 (c);

(c) $E_{\rm PA}$ 值偏大, 并且差值大于门限 T_2, 则选择该模型; 若两个模型都不满足此条件, 转为准则 (d);

(d) $E_{\rm PA}$ 值偏大, 则选择该模型.

以上表述中, T_1 和 T_2 是门限值, 需要对大量实测数据进行统计得出. 这里 T_1 取值 10, T_2 取值 0.1.

将长春站 1990 年 11 月和兰州站 1996 年 12 月 f_0F_2 数据对应的模型 $\rho_1(\tau)$ 和 $\rho_2(\tau)$ 的 $E_{\rm PA}$ 及 $E_{\rm abs}$ 值分别列于表 3, 与表 1 和表 2 对照后发现, 上述模型选择准则完全适用于这两组数据. 长春站利用准则 (b), 兰州站利用准则 (d), 都能得出选择模型 $\rho_2(\tau)$ 的结论, 而事实上使用模型 $\rho_2(\tau)$ 确实比使用 $\rho_1(\tau)$ 的插值误差小.

表 3　长春站 1990 年 11 月和兰州站 1996 年 12 月模型 $\rho_1(\tau)$ 和 $\rho_2(\tau)$ 的 $E_{\rm PA}$ 及 $E_{\rm abs}$ 值 (单位: MHz)

Table 3　Values of $E_{\rm PA}$ and $E_{\rm abs}$ for $\rho_1(\tau)$ and $\rho_2(\tau)$ in November 1990 from Changchun and December 1996 from Lanzhou (unit: MHz)

模型	长春站 1990 年 11 月		兰州站 1996 年 12 月	
	$E_{\rm PA}$	$E_{\rm abs}$ (×0.1)	$E_{\rm PA}$	$E_{\rm abs}$ (×0.1)
$\rho_1(\tau)$	0.7792	39.3	0.8032	8.2
$\rho_2(\tau)$	0.5719	9.5	0.9128	8.6

3　插值精度的评估

为了评估模型选择准则下 $z_{\rm DF}$ 自相关分析插值法 (以下简称为 $z_{\rm DF}$ 插值) 的插值精度, 我们利用北京站 1996 年 10 月 (中纬地区太阳活动低年) 和 2000 年 10 月 (中纬地区太阳活动高年) 以及海口站 1996 年 10 月 (低纬地区太阳活动低年) 和 2000 年 10 月 (低纬地区太阳活动高年) 的 f_0F_2 数据, 计算出 $z_{\rm DF}$ 插值与文献 [1] 中的 f_0F_2 插值的相对误差, 并将结果分别列于表 4~7. 可以看出, 无论在中纬还是低纬, 太阳活动高年还是低年, $z_{\rm DF}$ 插值的精度都明显高于 f_0F_2 插值, 尤其是在插值长度小于 4 h 的情况下 $z_{\rm DF}$ 插值优势更为明显; 一般情况下, 前者比后者可以把误差降低 1 到 2 个百分点以上, 在某种情况下甚至可以降低接近 9 个百分点 (例如海口站 1996 年 10 月缺失长度为 1 h 情况下误差降低了 8.8 个百分点).

作为一个总体的评估, 表 8 和表 9 分别列出了长春站和海口站 (20.00°N, 110.30°E) 在 1990 年和 1996 年 $z_{\rm DF}$ 插值全年的均方根误差统计计算结果, 同时还给出了文献 [1] 的 f_0F_2 插值和月中值插值的均方根误差. 表中数据表明, 在插值长度小于 10~12 h 的情况下, $z_{\rm DF}$ 插值较 f_0F_2 插值效果有明显的改善, 比起月中值插值改善更加明显; 对于更长的插值长度, 其插值误差趋向饱和, 但仍优于 f_0F_2 插值法和月中值插值.

表 4　北京站 1996 年 10 月 $z_{\rm DF}$ 插值的相对误差

Table 4　Relative error of $z_{\rm DF}$ interpolation method in October 1996 from Beijing

方法	插值长度/h																
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	22	24
$z_{\rm DF}$ 插值	8.8	9.4	9.8	9.9	10.1	10.2	10.3	10.4	10.5	10.5	10.6	10.7	10.8	10.8	10.8	10.8	10.9
f_0F_2 插值	10.7	11.4	11.7	11.9	12.0	12.0	12.1	12.1	12.1	12.1	12.2	12.2	12.2	12.2	12.2	12.2	12.2

表 5　北京站 2000 年 10 月 $z_{\rm DF}$ 插值的相对误差

Table 5　Relative error of $z_{\rm DF}$ interpolation method in October 2000 from Beijing

方法	插值长度/h																
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	22	24
$z_{\rm DF}$ 插值	5.4	6.1	6.8	7.4	7.8	8.0	8.1	8.3	8.3	8.4	8.6	8.7	8.8	8.8	8.9	8.9	8.9
f_0F_2 插值	7.8	8.9	9.4	9.7	9.8	9.9	10.0	10.1	10.1	10.1	10.2	10.2	10.2	10.2	10.2	10.2	10.3

表 6 海口站 1996 年 10 月 z_{DF} 插值的相对误差

Table 6 Relative error of z_{DF} interpolation method in October 1996 from Haikou

方法	插值长度/h																
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	22	24
z_{DF} 插值	12.4	14.8	16.6	17.8	18.6	19.2	19.7	20.2	20.6	21.1	21.8	22.4	22.8	23.2	23.4	23.6	23.8
f_0F_2 插值	21.2	22.5	23.8	24.5	24.9	25.1	25.3	25.4	25.5	25.6	25.7	25.8	25.9	26.0	26.0	26.1	26.1

表 7 海口站 2000 年 10 月 z_{DF} 插值的相对误差

Table 7 Relative error of z_{DF} interpolation method in October 2000 from Haikou

方法	插值长度/h																
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	22	24
z_{DF} 插值	8.1	9.6	10.7	11.5	12.0	12.4	12.7	13.0	13.1	13.2	13.4	13.5	13.6	13.7	13.8	13.8	13.9
f_0F_2 插值	12.8	14.6	15.5	15.9	16.1	16.3	16.4	16.4	16.5	16.6	16.6	16.7	16.7	16.8	16.8	16.8	16.9

表 8 长春站 f_0F_2 数据 z_{DF} 插值的均方根误差 (年平均值, 单位 0.1 MHz)

Table 8 Root mean square errors of z_{DF} interpolation method for the hourly values of f_0F_2 from Changchun (yearly average, unit: 0.1 MHz)

年份	方法	插值长度/h																	月中值
		1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	22	24	插值
1990	z_{DF} 插值	5.7	6.0	6.2	6.4	6.5	6.6	6.8	6.8	6.9	7.0	7.1	7.2	7.3	7.3	7.4	7.4	7.4	8.1
1990	f_0F_2 插值	6.2	6.6	7.0	7.1	7.3	7.4	7.4	7.5	7.6	7.6	7.6	7.7	7.7	7.7	7.7	7.7	7.7	8.1
1996	z_{DF} 插值	3.9	4.1	4.3	4.4	4.5	4.6	4.6	4.6	4.6	4.7	4.7	4.7	4.7	4.8	4.8	4.8	4.8	4.9
1996	f_0F_2 插值	4.1	4.5	4.6	4.7	4.8	4.8	4.9	4.9	4.9	4.9	4.9	4.9	4.9	4.9	4.9	5.0	5.0	4.9

表 9 海口站 f_0F_2 数据 z_{DF} 插值的均方根误差 (年平均值, 单位 0.1 MHz)

Table 9 Root mean square errors of z_{DF} interpolation method for the hourly values of f_0F_2 from Haikou (yearly average, unit: 0.1MHz)

年份	方法	插值长度/h																	月中值
		1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	22	24	插值
1990	z_{DF} 插值	9.3	10.6	11.4	11.9	12.3	12.7	12.9	13.1	13.3	13.4	13.6	13.7	13.8	13.9	14.0	14.0	14.0	14.4
1990	f_0F_2 插值	9.6	10.9	11.7	12.3	12.6	13.0	13.2	13.3	13.5	13.6	13.9	14.0	14.1	14.2	14.2	14.2	14.2	14.4
1996	z_{DF} 插值	6.5	7.7	8.5	8.9	9.2	9.5	9.6	9.8	9.9	10.0	10.1	10.2	10.2	10.3	10.3	10.3	10.3	10.6
1996	f_0F_2 插值	7.4	8.7	9.3	9.6	9.8	9.9	9.9	10.0	10.0	10.1	10.2	10.3	10.4	10.4	10.4	10.4	10.4	10.6

电离层特征参量历史缺失数据的插值误差与季节、太阳活动性和地理纬度等都有关系, 但是均方根误差和相对误差的变化规律却有所不同. 一般说来, 均方根误差在太阳活动高年, 4 月和 5 月出现最大值的概率比较大, 而最小值经常出现在 6 月和 7 月, 在太阳活动低年, 最大值一般出现在 3 月和 10 月, 6 月、7 月和 11 月则常常出现最小值. 另外分析其随地理纬度的变化, 发现在中国低纬地区的均方根误差普遍大于高纬地区, 而且均方根误差随太阳活动性的增强而增大. 相对误差的变化曲线, 在季节变化上呈现出规则的双峰结构 (一年中出现两个极大值), 随着地理纬度的变化, 双峰结构出现的月份不同, 由高纬到低纬, 双峰结构的位置由冬季向分季过渡, 且波峰 (最大值) 和波谷 (最小值) 逐渐趋于平展, 其中夏季一直取最小值; 相对误差一年中最大值与最小值的差值在

太阳活动低年比高年要小; 与绝对误差无论是在太阳活动高年还是低年低纬地区都大于高纬地区的特征不同, 相对误差在太阳活动高年, 最大值呈现随地理纬度的减小而逐渐减小的趋势, 但最小值却随之增大, 而在太阳活动低年, 最大值和最小值 (从而反映整体上) 均随着地理纬度的减小而逐渐增大.

4　结论

本文选取了合适的电离层参数, 将非平稳的电离层 f_0F_2 序列转变成平稳的 $z_{\rm DF}$ 序列, 并重新建立其正定的自相关系数模型, 在自相关分析原理的基础上形成了一个电离层历史缺失数据的插值处理新方法——$z_{\rm DF}$ 自相关分析插值方法. 该方法比起文献 [1] 中的 f_0F_2 插值和月中值插值均有明显的改善. 经过统计, 该方法与 f_0F_2 插值相比, 一般情况下能够把误差降低 1 到 2 个百分点以上, 有时甚至能够降低接近 9 个百分点, 大大提高了缺失数据的插值精度.

对于发生电离层暴情况下缺失数据的插值处理, 我们也进行了相关研究, 发现使用 $z_{\rm DF}$ 的插值结果要优于使用 f_0F_2 插值和月中值插值. 这是因为 $z_{\rm DF}$ 自相关分析插值方法是基于利用缺失数据两边与之相关性较高的数据的加权平均值来插值的思想, 所以只要暴时的数据没有全部丢失或者保留了一定数量, 缺失数据均会以一定的精度得到插值处理, 这也进一步反映了 $z_{\rm DF}$ 自相关分析插值方法的有效性.

致谢　使用的数据取自中国电波传播研究所中心数据库, 为此对各观测站及中心数据库的工作人员表示感谢.

参考文献

[1] Muhtarov G, Kutiev I. Autocorrelation method for temporal interpolation and short-term prediction of ionospheric data. *Radio Sci.*, 1999, **34**(2):459∼464

[2] Liu Ruiyuan *et al.* Application of autocorrelation method on ionosphreic short-term forecasting in China. *Chin. Sci. Bull.*, 2005, **50**(24):2781∼2785. in Chinese (刘瑞源等. 自相关分析法在中国电离层短期预报中的应用. 科学通报, 2005, **50**(24):2781∼2785)

[3] Peter J, Brockwell, Richard A. Davis. Time Series: Theory and Methods. New York: Springer Press, 1991

[4] Wilkinson P J. Predictability of ionospheric variations for quiet and disturbed conditions. *J. Atmosph. Terr. Phys.*, 1995, **57**:1469∼1481

[5] Marin D, Miro G, Mikhailov A V. A method for f_0F_2 short-term prediction. In: 4th COST 251 Workshop proceedings, Madeira, Portugal, 1999, COST251TD (99) 008: 214∼222

[6] Rushm C M. Edwards Jr W R. An authomated mapping technique for representing the hourly behavior of the ionosphere. *Radio Sci.*, 1976, **11**:931∼937