0254-6124/2009/29(2)-208–05 *Chin. J. Space Sci.* 空间科学学报

Kong Qingyan, Liu Wen, Jiao Peinan, Fan Junmei. One-step ahead prediction of f_0F_2 based on phase space reconstruction. *Chin. J. Space Sci.*, 2009, **29**(2): 208∼212

基于相空间重构的 f_0F_2 一步预测算法研究*

孔庆颜 柳 文 焦培南 凡俊梅

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

摘 要 利用低纬度地区某垂测站 2005 年 3 月和 4 月两个月 f_0F_2 参数共 5856 个数据样本, 对提前 15 min 的一步预测算法进行了研究. 基于混沌时间序列的相空间重构方法, 以相近邻点轨迹具有相似性为预测理论基础, 采用 k 最近邻点的方法对下一时刻的 f_0F_2 进行预测. 对邻点个数的选取采用了训练法和自适应选择法, 对选出的邻点采用平均法和自回归两种方法进行处理, 并对几种不同方法的预测结果进行了比较. 结果证明, 基于相空间重构的一步预测算法预测精度较高, 并且容易实现, 运算速度高, 适用于电离层参数准实时预报.

关键词 相空间重构; f_0F_2; 嵌入维数; 短期预报

中图法分类号 P 353

One-step Ahead Prediction of f_0F_2 Based on Phase Space Reconstruction

KONG Qingyan LIU Wen JIAO Peinan FAN Junmei

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

Abstract One-step ahead prediction of critical frequency f_0F_2 is presented, using samples of about 5856 observations of f_0F_2 with 15-minute time resolution taken from a station in the low latitude, from March 2005 to April 2005. Based on phase space reconstruction of chaotic time series, and according to the theory that near trajectories have similar characteristics, the k-nearest neighbor method is taken to forecast f_0F_2 15 minutes ahead. Training method and self-adaptive-select method are attempted to define the neighbor numbers used for prediction, and for these neighbors, averaging method and regression ones are taken to treat them. The forecasting errors for those methods are compared. It is shown that one-step ahead forecasting method based on phase space reconstruction can reach high accuracy. In addition, the method has the advantage of easy to implementing with high processing speed. So it is quite suitable for quasi real-time prediction.

Key words Phase space reconstruction, f_0F_2, Embedding dimension, Short-term prediction

---

* 国家自然科学基金项目 (60771049), 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
2008-03-31 收到原稿, 2008-10-11 收到修定稿
E-mail: kqy_127@163.com

1 引言

在电离层诸多特征参数中, 临界频率 f_0F_2 是一个十分重要的电离层参量, 其对于很多短波工程应用系统的工作性能起了决定性作用, 因此, 提前预知该参数的大小, 适当调整相关系统的工作参数, 才可以使系统达到最佳工作状态. 基于这一需求, f_0F_2 参数的短期预报成为备受瞩目的研究热点之一. 目前, 国际上对该参数的预测方法不断创新, 主要预报方法有神经网络法、自相关预测法、多元线性回归法等, 有对月中值进行的长期预测, 例如柳文等[5] 采用神经网络法对 f_0F_2 月中值进行预测, 平均误差小于 0.3 MHz. 也有提前 1 h 和提前 1 天进行的短期预测, 如文献 [6] 利用欧洲中纬 1982 年及 1991 年数据, 以时间信息, 当前 f_0F_2 值及其一阶差分、二阶差分、相对差分和 Kp 指数作为神经网络输入参数, 实现 f_0F_2 参数提前 1 h 预测, 而文献 [7] 利用前一天 24 h 的 f_0F_2 值, 通过主成分分析法变换为 5 个系数并结合地磁指数 Ap 和太阳活动指数 $F_{10.7}$ 作为神经网络的输入, 对提前 24 h 的 f_0F_2 预测进行了研究.

目前国内外对于准实时预报研究较少, 但对于一些特殊短波系统, 需要精确预知准实时电离层参数, 提前几小时或几天预测的结果及精度无法满足需求. 因此, 本文对提前 15 min 的一步预测算法进行了研究. Koutroumbas[4] 曾利用欧洲的数据, 对提前 15 min 一步预测算法进行了研究, 并对基于相空间重构的线性预测方法和神经网络法进行了比较, 指出对于提前 15 min 的一步预测, 两种方法预测精度差别不大, 加之线性方法易于实现的优点, 基于相空间重构的线性预测方法具有很大的优势, 文献 [8] 也对相空间重构方法进行了研究. 本文在文献 [4] 的研究基础之上, 对 k 最近邻点法进行了改进, 并提出了回归预测算法, 取得了较高的预测精度.

2 相空间重构

所谓相空间重构[1],[2] 就是把具有混沌特性的时间序列重建为一种低阶非线性动力学系统. 对于时间序列 $\{x_i\colon\ i=1,2,3,\cdots,n\}$, n 是序列长度, 根据 Takens 提出的嵌入定理, 可以将该序列延拓成如下 m 维相空间的一个相型分布:

$$\begin{cases} D_1: & x(t_1) & x(t_2) & \cdots & x(t_i) & \cdots & x[t_n-(m-1)\tau], \\ D_2: & x(t_1+\tau) & x(t_2+\tau) & \cdots & x(t_i+\tau) & \cdots & x[t_n-(m-2)\tau], \\ D_3: & x(t_1+2\tau) & x(t_2+2\tau) & \cdots & x(t_i+2\tau) & \cdots & x[t_n-(m-3)\tau], \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots, \\ D_m: & x[t_1+(m-1)\tau] & x[t_2+(m-1)\tau] & \cdots & x[t_i+(m-1)\tau] & \cdots & x(t_n). \end{cases} \tag{1}$$

$$y(t_1)\quad y(t_2)\quad y(t_i)\qquad \cdots\quad y(t_n-(m-1)\tau). \tag{2}$$

上式组成的相型分布中有 m 个分量, 式 (1) 中每一列就是一个相点, 即式 (2) 中的 $y(t_i)$, 记为:

$$y(t_i)=\{x(t_i),x(t_i+\tau),\cdots,x[t_i+(m-1)\tau]\},$$

其中,

$$x(t_i),x(t_i+\tau),x(t_i+2\tau),\cdots,x[t_i+(m-1)\tau]$$

分别称为 $y(t_i)$ 的第一、第二、第三 …… 及第 m 个分量. 这 $n-(m-1)\tau$ 个相点间的连线描述了系统在 m 维相空间的演化轨迹. 其中 m 称为嵌入维数, τ 称为延迟时间. 根据 Takens 定理, 只要嵌入维数 m 和时间延迟量 τ 选择合适, 就可以在拓扑等价的意义下恢复吸引子的动力学特性. 在文献 [2] 中也证实, 在重构后的相空间中, 相邻相轨道的演变具有相似性, 邻近的多条轨道遵从类似变化, 这就为下面要讨论的预报方法奠定了理论基础.

由于本文研究的是一步预测, 因此 τ 取值为 1. 这时, 我们所研究的 m 维相空间分布可以演变为式(3).

剩下的关键问题就是如何求解嵌入维数, 本文中利用虚假邻点法[3],[4] 求解嵌入维数. 所谓虚假邻点法的基本思想是当嵌入维数从 m 变成 $m+1$ 时, 考察轨线的邻点中哪些是真实的邻点, 哪些是虚假的邻点, 当没有虚假的邻点时, 可以认为几何结构被完全展开.

$$\begin{cases} D_1: & x(t_1) & x(t_2) & \cdots & x(t_i) & \cdots & x[t_n-(m-1)], \\ D_2: & x(t_1+1) & x(t_2+1) & \cdots & x(t_i+1) & \cdots & x[t_n-(m-2)], \\ D_3: & x(t_1+2) & x(t_2+2) & \cdots & x(t_i+2) & \cdots & x[t_n-(m-3)], \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots, \\ D_m: & x[t_1+(m-1)] & x[t_2+(m-1)] & \cdots & x[t_i+(m-1)] & \cdots & x(t_n). \end{cases} \tag{3}$$

同文献 [4], 本文仍采用 L_2 范数定义两矢量间的欧式距离, 设 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 是相空间中任意两个 m 维矢量, 则这两个相点之间的距离为

$$\|\boldsymbol{u}-\boldsymbol{v}\| = \sqrt{\sum_{i=1}^{m}(u_i-v_i)^2}. \tag{4}$$

首先固定嵌入维数 m, 对于式 (3) 中每个相点 $y(t_i)$, 利用式 (4) 求其与重构的相空间中其他全部相点之间的距离. 距离最小的相点 $y(t_j)$ 就称为相点 $y(t_i)$ 的最近邻点, 其距离为 $\|y(t_i)-y(t_j)\|^m$. 下面再来判断此邻点是否为虚假邻点. 将维数增加到 $m+1$, 重新求解这两个相点之间的距离 $\|y(t_i)-y(t_j)\|^{m+1}$, 若 $\|y(t_i)-y(t_j)\|^{m+1}$ 比 $\|y(t_i)-y(t_j)\|^m$ 大很多, 则认为该最近邻点是虚假邻点, 即若

$$\frac{\|y(t_i)-y(t_j)\|^{m+1}-\|y(t_i)-y(t_j)\|^m}{\|y(t_i)-y(t_j)\|^m} > R_{\mathrm{T}}, \tag{5}$$

则 $y(t_j)$ 是 $y(t_i)$ 的虚假最近邻点. 门限值 R_{T} 可以在 [10, 50] 之间选取. 对于无限长精确数据, 上述准则可以给出很好的结果, 但由于实测数据往往含有噪声, 或者可用的时间序列长度有限, 所以补充以下准则:

$$\frac{\|y(t_i)-y(t_j)\|^{m+1}}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_n-\bar{x})^2}} > 2. \tag{6}$$

其中, $\bar{x}=\dfrac{1}{N}\sum_{n=1}^{N}x_n$ 是时间序列的均值. 满足式 (6) 的 $y(t_j)$ 也是 $y(t_i)$ 的虚假最近邻点. 实际操作过程如下: 首先确定 m 为 2, 选择 $R_{\mathrm{T}}=30$[3], 计算全部虚假最近邻点的比例 (最近虚假邻点的判断依据上面给出的两个准则), 若虚假最近邻点的比例小于全部相点的 1%[4], 或虚假最近邻点不再随着 m 的增加而减少时, 此时可以认为相空间完全打开, 此时的 m 就是求得的嵌入维数; 若条件不满足, 则增加 m, 直至条件满足为止.

3 最近邻点个数确定及预报方法原理分析

在确定嵌入维数 m 后, 由式 (3) 确定的相空间分布就唯一确定了. 由于前面已经论述过, 相邻相轨道的演变具有相似性, 邻近的多条轨道遵从类似变化, 因此, 本文采用的 k 最近邻点的方法进行预测, 具体方法如下.

给定相点 $y(t_i)=\{x(t_i), x(t_i+1), x(t_i+2), \cdots, x[t_i+(m-1)]\}$, 其 m 个分量为最近实测值 (已知量), 那么其下一时刻的值即为我们要预测的值 $x(t_i+m)$. 利用时间序列中在 $y(t_i)$ 之前的点求解 $y(t_i)$ 的 k 个最近邻点 $y'_1(t_i), y'_2(t_i), y'_3(t_i), \cdots, y'_k(t_i)$, 对于这 k 个邻点, 其下一时刻值分别为 $x'_1(t_i+m)$, $x'_2(t_i+m)$, $x'_3(t_i+m)$, $\cdots, x'_k(t_i+m)$, 这些都是已有的实测值, 介于邻点轨道的类似变化, 可以通过这 k 个值对 $x(t_i+m)$ 进行预测.

首先来看一下如何确定 k. 本文中对于 k 的选取采用了样本训练法和自适应选择两种方法. 所谓样本训练法就是利用样本数据进行训练[4], 使得对于全部训练样本预测结果的均方根误差最小的 k 就是我们选择的邻点数, 然后使用该值对测试样本进行预测. 文中利用 3 月份的后半个月的数据作为训练样本, k从 2 取到 30, 并计算每次训练样本预测结果的均方根误差, 选择使得均方根误差最小的 k 对所有的测试样本进行预测; 而自适应选择法则是在每次预测之前都将待预测点之前一定量的数据 (本文采用 100 个数据) 作为训练样本, 通过上述的训练法选择 k, 并利用该 k 值对待预测点进行预测. 确定了 k 之后, 本文采用以下几种方法对测试样本进行预测:

(1) 平均法

平均法[4] 很容易理解, 也易于操作, 即将 k 个最近邻点的下一时刻值直接取平均来作为待预测点的值, 有

$$x(t_i+m)=\frac{1}{k}[x'_1(t_i+m)+x'_2(t_i+m)+\cdots+x'_k(t_i+m)]. \quad (7)$$

(2) 回归法

回归法是通过对选择的 k 个最近邻点进行回归运算, 确定一组回归系数, 然后代入待预测相点, 得到最终的预测值. 具体方法如下.

我们把相点 $y(t_i)$ 的 k 个最近邻点 $y'_1(t_i)$, $y'_2(t_i)$, $y'_3(t_i),\cdots,y'_k(t_i)$ 的下一时刻值 $x'_1(t_i+m), x'_2(t_i+m), x'_3(t_i+m),\cdots,x'_k(t_i+m)$ 分别作为 k 个邻点的函数值, 而将 k 个邻点的各个分量作为变量进行回归计算, 即

$$\begin{aligned}
x'_1(t_i+m)&=a_1x'_1(t_i)+a_2x'_1(t_i+1)+\cdots+a_mx'_1[t_i+(m-1)],\\
x'_2(t_i+m)&=a_1x'_2(t_i)+a_2x'_2(t_i+1)+\cdots+a_mx'_2[t_i+(m-1)],\\
&\cdots\quad\cdots\quad\cdots\quad\cdots\quad\cdots\\
x'_k(t_i+m)&=a_1x'_k(t_i)+a_2x'_k(t_i+1)+\cdots+a_mx'_k[t_i+(m-1)].
\end{aligned}$$

这样, 通过均方误差最小准则可以确定一组回归系数 a_i $(i=1,2,\cdots,m)$, 该组系数反映了预测点与相点元素之间的相关关系, 将其代入相点 $y(t_i)$ 可以计算得到预测值 $x(t_i+m)$, 即

$$x(t_i+m)=a_1x_1(t_i)+a_2x_2(t_i+1)+\cdots+a_mx_k[t_i+(m-1)]. \quad (8)$$

4 预测结果分析

下面对测试数据的预测结果进行比较分析. 分析中采用平均绝对误差、均方根误差及相对误差进行衡量. 对于各时刻及全部样本误差分析都采用以下公式计算:

$$E_{\mathrm{AA}}=\frac{1}{N}\sum_{i=1}^{N}|m(i)-f(i)|, \quad (9)$$

$$E_{\mathrm{RMS}}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}[m(i)-f(i)]^2}, \quad (10)$$

$$E_{\mathrm{AR}}=\frac{1}{N}\sum_{i=1}^{N}\frac{|m(i)-f(i)|}{m(i)}\times 100. \quad (11)$$

其中, $m(i)$ 表示第 i 个预测点对应的实测值, 而 $f(i)$ 表示第 i 个点的预测值. 我们以某垂测站 4 月份的数据为例, 各种预测算法的预测误差如图 1 所示.

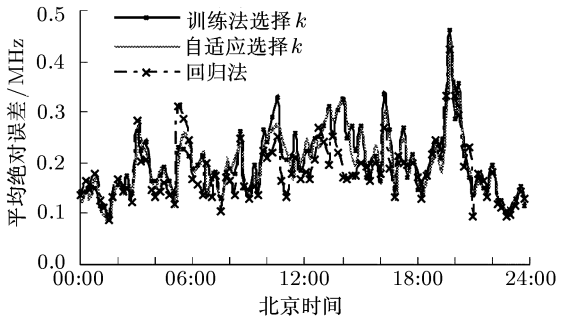


图 1 平均绝对预测误差

Fig.1 Average absolute error

从图 1 可以清楚地看出, 基于相空间重构的 k-最近邻点法进行一步预测精度较高, 表 1 给出了各种预测方法的全月绝对误差平均值、均方根误差及相对误差平均值.

表 1 不同方法预测的误差

Table 1 Prediction errors for each method

预测方法	平均绝对误差/MHz	均方根误差/MHz	平均相对误差/(%)
平均法 (训练法选择 k)	0.2022	0.2923	3.78
平均法 (自适应选择 k)	0.195	0.2888	3.64
回归法 (训练法选择 k)	0.1816	0.2676	3.48

表 1 给出的三种基于相空间重构预测方法中, 平均法 (训练选择 k) 是文献 [4] 中提到的方法, 平均法 (自适应选择 k) 是对该方法的改进, 从预测结果可以看出, 预测精度有了一定提高. 回归法是本文提出的新方法, 在几种预测方法中预测精度最高. 但总体来讲, 基于相空间重构的三种预测算法对于凌晨 (北京时间 03:00—06:00) 及日落 (北京时间 20:00 左右) 预测误差相对较大, 从平均相对误差分布曲线 (见图 2) 可以更清楚地看到这点. 针对这一误差特征, 我们分析了各时刻预测误差与对应时刻 f_0F_2 参数一阶差分绝对平均值之间的关系, 见图 3.

从图 3 中可以看出, 一阶差分绝对平均值与预测误差绝对平均值基本成正相关关系. 通过公式 (12) 计算两者之间的相关系数为 0.9271, 这更

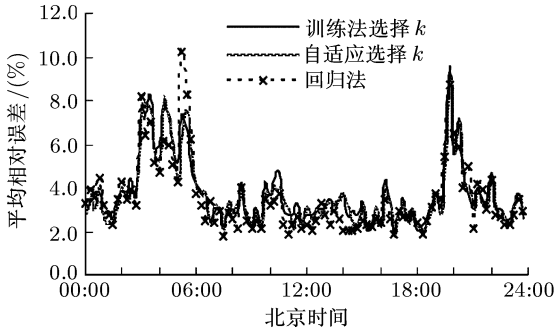


图 2 平均相对误差

Fig.2 Average relative error

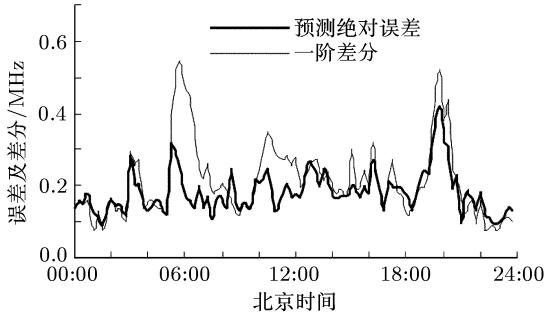


图 3 绝对误差与一阶差分比较

Fig.3 Comparison of absolute error and first difference

说明了两者之间有很强的线性相关性, 在今后研究中将进一步考虑一阶差分与 f_0F_2 参数之间的关系, 有望提高预测精度,

$$\rho_{XY} = E\Big\{\frac{X-E(X)}{\sqrt{D(X)}}\frac{Y-E(Y)}{\sqrt{D(Y)}}\Big\}. \qquad (12)$$

5 结论

本文提出了基于相空间重构的 k 最近邻点法进行 f_0F_2 参数一步预测, 对 k 值的选择采用了训练法和自适应选择法, 对最近邻点的处理采用了平均法和回归法, 预测结果表明, 自适应选择 k 优于训练法, 回归法优于平均法. 总体来讲, 几种基于相空间重构的一步预测算法均简单实用, 且预测精度很高, 标准差小于 0.3 MHz, 平均相对误差小于 4%, 但对于参数变化较快的时段预测误差相对偏大, 反映出线性预测法对于非线性参数变化规律预测的局限性.

参考文献

[1] Jiang Herong, Yang Zhengyu, Yang Meihua. Chaotic characteristics of the ionospheric parameters and the predict table time scale. *Chin. J. Space Sci.*, 1994, **14**(3):221∼226. in Chinese (蒋和荣, 杨征宇, 杨美华. 电离层参数的混沌特性及可预报的时间尺度. 空间科学学报, 1994, **14**(3):221∼226)

[2] Jiang Herong, Yang Zhengyu, Yang Meihua. Representations of the phase trajectory on the behaviour of ionospheric parameters and the new prediction model. *Chin. J. Space Sci.*, 1995, **15**(2):90∼107. in Chinese (蒋和荣, 杨征宇, 杨美华. 电离层参数的相轨道表示及新的预测模型. 空间科学学报, 1995, **15**(2):99∼107)

[3] Wang Haiyan, Sheng Zhaohan. Choice of the parameters for the phase space reconstruction of chaotic time series. *J. Southeast Univ.*, 2000, **30**(5):113∼117. in Chinese (王海燕, 盛招瀚. 混沌时间序列相空间参数的选取方法. 东南大学学报, 2000, **30**(5):113∼117)

[4] Koutroumbas K, Belehaki A. One-step ahead prediction of f_0F_2 using time series forecasting techniques. *Ann. Geophys.*, 2005, **23**:3035∼3042

[5] Liu Wen, Jiao Peinan. Prediction the monthly mean values of f_0F_2 by using the artificial neural network. *Chin. J. Space Sci.*, 2000, **20**(4):310∼317. in Chinese (柳文, 焦培南. 利用人工神经网络预测电离层 f_0F_2 月中值. 空间科学学报, 2000, **20**(4):310∼317)

[6] Altinay O, Tulunay Y. Forecasting of ionospheric critical frequency using neural networks. *Geophys. Res. Lett.*, 1997, **24**(12):1467∼1470

[7] Wintoft P, Cander Lj R. Twenty-four hour predictions of f_0F_2 using time delay neural networks. *Radio Sci.*, 2000, **35**(2):395∼408

[8] Ying Yemei, Wang Xiaodong. Phase Space Prediction of Chaotic Time Series with Nu-Support Vector Machine Regression. *Commun. Theor. Phys.*, 2005, **43**(1):102∼106