

奇异谱分析在太阳 10.7 cm 射电流量 中期预测中的应用^{*}

钟秋珍¹ 刘四清¹ 何卷雄² 龚建村¹

¹(中国科学院空间科学与应用研究中心 北京 100080)

²(中国科学院大气物理研究所)

摘要 首次尝试利用信号处理技术奇异谱分析方法预测太阳活动低年未来 27 天太阳 10.7 cm 射电流量。选取的预报试验时间段是 2004 年 4 月 30 日至 5 月 30 日, 此试验期内太阳活动水平相对较低。在样本时间序列的构建上, 吸取了相似周数据分析思路, 采取的是 23 周实时观测数据与其相似周第 20 周下降年部分数据相结合的方式, 既增加了样本长度又避开了太阳活动的活跃期。这 31 天的预报试验结果表明, 大部分情况下, 预报值基本上体现出 $F_{10.7}$ 的变化趋势, 平均相对误差为 10.5%; 比同时期美国空军预测值的平均相对误差小, 前者为 11.3%, 后者为 14.6%; 除两天外, SSA 每一次 27 天的预报结果的平均相对误差比美国空军 (AAF) 的要小; 对不同的时间提前量而言, AAF 提前 1 天到提前 12 天的预报准确性较奇异谱分析方法要高, 即 AAF 较短期的预报效果更好。

关键词 奇异谱; 10.7 cm 射电流量; 中期预报

中图分类号 P 41

Application of Singular Spectrum Analysis to Solar 10.7 cm Radio Flux 27-day Forecast

ZHONG Qiuzhen¹ LIU Siqing¹ HE Juanxiong² GONG Jiancun¹

¹(Center for Space Science and Applied Research, The Chinese Academy of Sciences, Beijing 100080)

²(Institute of Atmospheric Physics, The Chinese Academy of Sciences)

Abstract In this paper, singular spectrum analysis (SSA) method was applied to 27-day solar 10.7 cm radio flux forecast for the first time. In order to focus on the $F_{10.7}$ prediction during the period of the low level solar activity, the test prediction range was chosen from April 30 to May 30, 2004 within the cycle 23 descendent phase. To ensure there were enough data to construct the sample series, both real time data and partial data in the cycle 20 descendent periods following the rule of similar cycle method were chosen, in this way sampling number increased and at the same time, effects of high solar activity are kept away so to get more precise $F_{10.7}$ forecast during the low solar activity. The forecasted results of these 31 days show that they mainly coincided with the general tendency of $F_{10.7}$ variations. The averaged relative error of each 27-day forecast was about 10%. The total average relative error of the 31 days was 10.5%. The minimum average relative error was 4.8% and the maximum was 16.7% during the test forecast period. Comparing with American

^{*} 国家重点基础研究发展规划项目 (G200078408) 和中国科学院空间环境研究预报中心共同资助

2004-10-22 收到原稿, 2005-03-09 收到修定稿

Air Force (AAF) forecast in the same periods, the total average relative error of SSA is less than AAF, the former was 11.3% and the latter was 14.6%. Furthermore, each average relative error of SSA was less than AAF except two days. But to the short-term forecast within one day to 12 days, the result of AAF was superior to SSA.

Key words Singular spectrum analysis, Solar 10.7 cm radio flux, 27-day forecast

1 引言

太阳活动及其对近地空间环境的影响与人类的生存环境和活动有着广泛而密切的关系,因而对太阳活动的预报已经逐渐发展成为一项非常重要的空间环境预报项目^[1]. 太阳 10.7 cm 射电流量 (以下简称 $F_{10.7}$), 由于它的测量相对容易和便捷, 而且能够较好地反映太阳活动的周期性变化, 因此在很多方面已经作为表征太阳活动性的一个重要指数. 在航天系统轨道预测的大气密度计算和影响通信系统的电离层预报中, $F_{10.7}$ 都是重要的输入参数.

对 $F_{10.7}$ 的预报, 中长期预报做得相对多些, 主要是对 $F_{10.7}$ 月均平滑值的预报. 由于太阳黑子数与 $F_{10.7}$ 较好的相关关系, 现在常用的方法是通过太阳黑子数的预报来间接实现 $F_{10.7}$ 的预报^[2]. 最近, 也有一些研究结果认为, $F_{10.7}$ 虽然与太阳黑子数有相似的变化趋势, 但其仍存在某些自身的变化特性, 直接利用 $F_{10.7}$ 历史数据进行预测, 结果比间接通过黑子数预报更准确^[3]. 对 $F_{10.7}$ 日值的预测过去做的工作还不是很多, 但由于各种大气模式和电离层预报模式的需求, 其已逐渐成为空间环境应用研究的重点课题之一. 目前, $F_{10.7}$ 预报已经成为空间环境业务预报重要内容之一. AAF 逐日发布未来 45 天 $F_{10.7}$ 的日值预报, NOAA 的 SEC (Space Environment Center) 每周发布未来 27 天的 $F_{10.7}$ 的日值预报.

奇异谱分析 (Singular Spectrum Analysis, SSA) 作为信号处理技术可以在事先未知系统物理本质的情况下, 从系统中提取尽可能多的可靠信息, 并根据这些信息建立预报模型, 适合于研究有周期振荡的时间序列. SSA 在地球科学领域有着广泛的应用, 如分析全球平均气温的年际变率^[4], 预测汛期降水, 磁层动力学研究^[5], 补充缺失资料^[6]等. 但尚未应用于空间环境预报领域. 考虑到 $F_{10.7}$ 较为明显的 11 年左右长周期和 27 天左右的短周期变化, 因此, 我们尝试将 SSA 分析方法应用到 $F_{10.7}$ 的中期预报中, 希望通过 SSA 分析方法的应用, 提高对 $F_{10.7}$ 的预报能力.

2 SSA 简介和预报流程

2.1 SSA 简介

SSA 本质上是时间域上的经验正交函数 EOF (Empirical Orthogonal Function), 详细情况可参考文献 [7]、[8], 这里只作简单的介绍.

对于样本量为 N 的标准化时间序列 $x_i = x(t)$, 给定嵌套维数 M , $M < N/2$, 建立时滞矩阵

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_{N-M+1} \\ x_2 & x_3 & \cdots & x_{N-M+2} \\ \cdots & \cdots & \cdots & \cdots \\ x_M & x_{M+1} & \cdots & x_N \end{bmatrix}, \quad (1)$$

(1) 式的滞后自协方差为一个 $M \times M$ 的矩阵,

$$S = \begin{bmatrix} s(0) & s(1) & \cdots & s(M-1) \\ s(1) & s(0) & \cdots & s(M-2) \\ \cdots & \cdots & \cdots & \cdots \\ s(M-1) & s(M-2) & \cdots & s(0) \end{bmatrix}, \quad (2)$$

S 为对称阵且主对角线为同一常数, 称为 Toeplitz 矩阵, 其特征值

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M \geq 0, \quad (3)$$

$$\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \cdots \geq \sqrt{\lambda_M} \geq 0, \quad (4)$$

(4) 式即为序列 $\{x_i\}$ 的奇异谱, 称对 X 的奇异值运算为奇异谱分析.

计算出 S 的特征值 λ_k 和相应的特征向量, 序列的 SSA 展开为

$$x_{i+j} = \sum_{k=1}^M a_{ik} E_{kj}. \quad (5)$$

式中, $i = 1, 2, \cdots, N - M + 1$; $j = 1, 2, \cdots, M$. 由于资料阵是由嵌入时间滞后构成的, E_{kj} 为时间 EOF, 记为 T-EOF, a_{ik} 为时间主分量, 记为 T-PC (Principal Components).

$$a_{ik} = \sum_{j=1}^M x_{i+j} E_{kj}. \quad (6)$$

SSA 最重要的应用是重建成分 RC (Reconstruction Components). 由第 k 个的 T-EOF 和 T-PC 重建 x_i 的成分记为 x_i^k , 公式为

$$x_i^k = \begin{cases} \frac{1}{M} \sum_{j=1}^M a_{ij,k} E_{kj}, & M \leq i \leq N - M + 1; \\ \frac{1}{i} \sum_{j=1}^i a_{ij,k} E_{kj}, & 1 \leq i \leq M - 1; \\ \frac{1}{N - i + 1} \sum_{j=i-n+m}^M a_{ij,k} E_{kj}, & N - M + 2 \leq i \leq N. \end{cases} \quad (7)$$

截取前 k 个贡献大的成分近似表示原序列,

$$\hat{x}_i = \sum_{k=1}^k x_{ik}, \quad i = 1, 2, \dots, N. \quad (8)$$

所有 RC 之和等于原序列

$$x_i = \sum_{k=1}^M x_{ik}, \quad i = 1, 2, \dots, N. \quad (9)$$

分量重建可进一步转化为预报问题.

2.2 预报流程

首先从资料中提取与资料自适应的长期趋势项和周期项, 使其按照方差贡献大小排列, 只保留该序列中的前几位振荡模态, 就能达到降噪效果. 在此基础上, 用多阶自回归方程对时间序列进行外推, 然后按照原序列的线性组合关系, 可得到序列的预测值. 具体的预报流程见图 1.

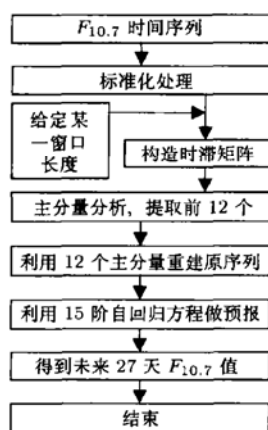


图 1 奇异谱分析预报流程

Fig.1 The forecast process of singular spectrum analysis

3 预报过程

3.1 数据选取

从多年的观测资料来看, $F_{10.7}$ 在太阳活动下降年和太阳活动低年具有较为明显的周期性, 在太阳活动高年, 太阳活动频繁爆发, $F_{10.7}$ 的周期性一定程度上遭到破坏. 因此, 第一步我们准备首先针对太阳活动低年进行 $F_{10.7}$ 的日值预测. 在样本数据选择上, 主要选取下降年和低年中太阳爆发活动较少的时间段的数据. 首先选取的是本次太阳活动周中太阳爆发活动较少的时间段: 2004 年 1 月 1 日至 2004 年 4 月 30 日. 预报试验的时间范围是从 2004 年 4 月 30 日至 5 月 30 日, 共 31 天. 通过分析预报结果发现, 预报的平均相对误差一般都在 14% 左右, 因此, 为提高计算精度, 需要增加样本数目. 由于所选样本时间段之前的一段时间太阳活动一直比较剧烈, 所以改选其他太阳活动周的下降年的数据. 在选择太阳活动周时, 考虑到“相似周”可能具备其他相似特征. 选择的另一段样本数据为本次太阳活动周的相似周 20 周下降年的数据. 数据资料选取的时间范围为 1972 年 11 月 6 日至 1973 年 09 月 22 日. 加上第 23 周 4 个月的数据, 样本长度为 441.

$F_{10.7}$ 观测资料可以在网上得到*. 为了对预报结果进行对比检验, 我们将 SSA 预报结果与 SEC 发布美国空军预报值进行比较**.

3.2 序列重建

经过多次试验, 确定窗口长度 $M = 80$, 主分量的个数是 12, 用的是 15 阶自回归方程. 预报过程是动态建模, 即每次预报都有新的实测值加入, 但仍保持样本长度为 441.

我们取前 12 个主分量, 是因为这 12 个特征向量表征了时间序列中最主要的周期模态, 总的拟合率为 93.1%. 将前 12 个特征向量分量合成拟合序列并与标准化后实况序列对比分析, 由图 2 可知, 拟合序列已基本体现实况序列的变化, 两者的相关系数达 0.97.

4 试预报效果

在选定样本数据和奇异谱分析方法参数的基础

*<http://spidr.sepc.ac.cn/spidr/>

**<http://www.sec.noaa.gov/Data/index.html>

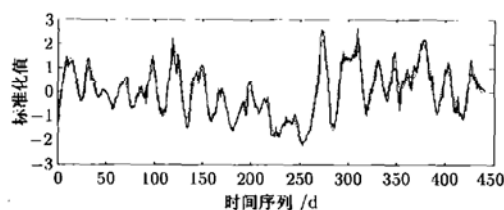


图2 标准化后的 $F_{10.7}$ 实测值序列 (实线)、拟合序列 (虚线) 变化图

Fig.2 The varieties of standardization solar $F_{10.7}$ observation (solid) and fitting value (dashed)

上, 针对 2004 年 4 月 30 日至 5 月 30 日这 31 天的时间开展了预报试验. 大部分情况下, 预报结果

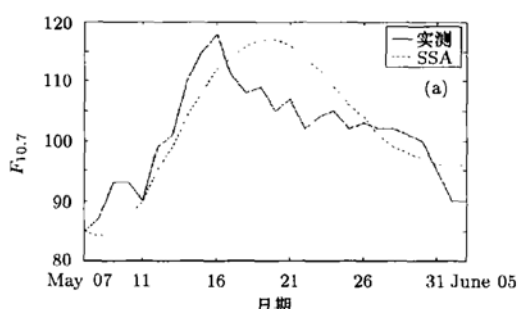
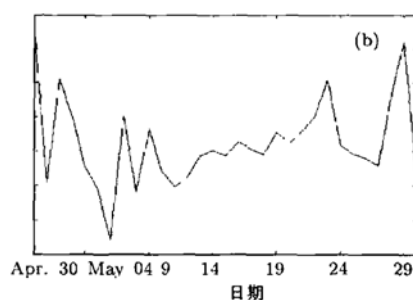


图3 (a) 5 月 6 日预测的未来 27 天 $F_{10.7}$ 变化与实测值的比较, (b) 4 月 30 日—5 月 30 日 SSA 分析预报的平均相对误差
Fig.3 (a) The comparison between solar $F_{10.7}$ observation and forecast of May 6, 2004, (b) the average relative error of SSA forecast from April 30 to May 30, 2004

些 $F_{10.7}$ 预报方法的预报能力, 我们将预报结果与目前国际上空间环境业务预报能力最强的美国空间环境中心所发布的 AAF 的 $F_{10.7}$ 预报结果进行对比分析. 对比的时间范围为 5 月 17—30 日, 共 14 天. 图 4(a) 显示的是 SSA 和 AAF 两种方法预报结果的平均相对误差. 从图中可以看出, 除了 5 月 28、29 两天外, SSA 的平均相对误差都小于 AAF. 这 14 天 SSA 的平均相对误差为 11.3%, AAF 为 14.6%.

图 4(b) 为 5 月 27 日 SSA、AAF 的预测和实际 $F_{10.7}$ 的对比图. 27 日为这 14 天中 SSA 方法平均相对误差较小的一天, 其值为 9.1%, AAF 为 13.2%. 从图中可以看出, SSA 方法预测的位相较 AAF 接近于观测, 但前 7 天 AAF 的预测值要好于 SSA. 29 日是这 14 天中 AAF 方法平均相对误差较小的一天. 但是尽管 AAF 预测的平均相对误差较 SSA 要小一些, 但是变化位相和实测也是相反的 (图略), 只是前几天的预报结果与实际观测很接近. 由此看来, AAF 前几天的预测结果要优于 SSA, 为了更全面地检测预报结果, 我们对不同时间提前量的预报平均相对误差进行了分析比较.

能基本反映 $F_{10.7}$ 的变化, 平均相对误差为 10.5%. 图 3(a) 显示的是 5 月 6 日预报结果与实测结果的比较, 这天是这 31 天预报结果中预报效果最好的一天. 从图 3(a) 中可以看到, 中期预报效果与短期预报效果基本持平. 其平均相对误差为 4.8%. 但大部分时间的预报平均相对误差都在 10% 附近. 图 3(b) 为这 31 天每天的平均相对误差, 这期间平均相对误差最大为 16.7%. 造成这种预报效果的差异的原因与预测的 $F_{10.7}$ 位相和实际位相的符合程度密切相关.

为了比较 SSA 分析方法与目前国际现有的一

图 4(c) 为 SSA 和 AAF 提前 1 天到提前 27 天预报平均相对误差. 从图中可以看到, 对于不同的时间提前量, SSA 方法的平均相对误差变化不大, 最小为 8.2%, 最大为 14.1%, 在 10% 上下摆动; AAF 的变化幅度较大, 最小为 4.4%, 最大为 24.3%. 但是提前 1—12 天 AAF 的平均相对误差要小于 SSA 方法, 即 AAF 的前面一段时间的预报结果的可信度更高. 由此可见, SSA 方法对 $F_{10.7}$ 的周期性变化预测得较好, 所以对于不同的时间提前量其平均相对误差的变化不大, 而 AAF 显然对其前期的预报效果较好.

5 结论和讨论

针对太阳活动低年 $F_{10.7}$ 日值的中期预报, 这里初步尝试将 SSA 分析作为一种预报方法. 在样本数据的选取上既利用了实时观测数据, 又选择一部分历史数据, 即第 23 太阳活动周的相似周第 20 周下降年的一部分数据. 从预报试验的结果可以得到

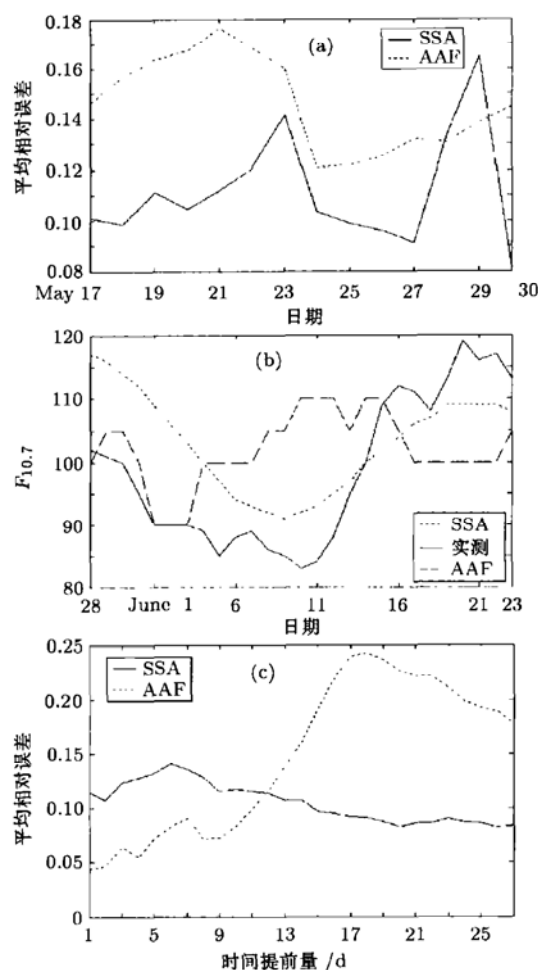


图 4 (a) 5 月 17—30 日 SSA 方法和 AAF 的预报平均相对误差, (b) 5 月 28—6 月 2 日 $F_{10.7}$ 日值的变化, (c) 提前 1—27 天预报的平均相对误差

Fig.4 (a) The comparison of the forecast average relative errors between SSA and AAF from 17 May to 30 May, (b) the solar $F_{10.7}$ varieties from 28 May to 2 June, (c) the comparison of the forecast average relative errors between SSA and AAF within 1 day and 27 days

如下结论.

(1) 31 天预报试验的结果, 大部分情况下都能基本反映 $F_{10.7}$ 未来 27 天的变化趋势, 每天预报的平均相对误差都在 10% 左右, 这 31 天总的平均相对误差为 10.5%. (2) 除个别日期外, 奇异谱分析方法每天的平均相对误差较同时期美国空军的小, 而且总的平均相对误差也较小. 但是提前 1 天到提前 12 天的平均相对误差较美国空军的大. 这说明 SSA 在 27 天这样时间长短的中期预报中具有一定的优势, 但在提前几天的短期预报方面还有待提高. (3)

按照相似周方法增加了 20 周下降段样本数据后, 预报准确性增加了. 这说明利用不同时间段构建时间序列的方法是可行的, 但其中的物理基础, 还有待于进一步考证. (4) 奇异谱分析方法对 $F_{10.7}$ 的周期性变化预测得较好, 而美国空军对其前期的预报效果较好. 可见两种方法具有不同的优势, 综合两种方法的优势, 应该能够提高 $F_{10.7}$ 中期预报的实际预报能力.

总之, 初步将 SSA 方法应用到 $F_{10.7}$ 预报中, 虽然取得了一些不错结果, 但仍然存在需要进一步研究和提高之处. 首先, 目前预报试验的时间较短, 还需要更长的时间段来检测奇异谱分析方法在太阳 10.7 cm 射电流量中期预测中的能力; 其次还要试验选取其他活动周的数据来增加样本数据, 包括 23 周的相似周以及非相似周, 并检验预报结果; 在 10 天以内的预报精度与国际水平还有一定差距, 有待于进一步完善和提高.

参考文献

- [1] Wang Jialong et al. On solar activity predictions. *Prog. Astron.*, 1990, 8(2):89—95. in Chinese (王家龙等. 太阳活动预报简论. 天文学进展, 1990, 8(2):89—95)
- [2] Zhang Guiqing, Wang Jialong. Extrapolation of solar 10.7 cm fluxes. *Astrophys. Rep.*, 1995, 27:61—68
- [3] Miao Juan et al. Primary research on prediction method of 10.7 cm solar radio flux. *Chin. J. Space Sci.*, 2003, 23(1):50—54. in Chinese (苗娟等. 太阳 10.7 cm 射电流量预报方法初探. 空间科学学报, 2003, 23(1):50—54)
- [4] Ding Yuguo, Yu Jinghua et al. Quasi-biennial oscillation variability in interannual variance of the global surface temperature during the last 100-year period. *Chin. J. Atmos. Sci.*, 2001, 25(1):89—102. in Chinese (丁裕国, 余锦华等. 近百年全球平均气温年际变率中的 QBO 长期变化特征. 大气科学, 2001, 25(1):89—102)
- [5] Sharma A S, Vassiliadis D, Papadopoulos K. Reconstruction of low-dimensional magnetospheric dynamics by singular spectrum analysis. *Geophys. Res. Lett.*, 1993, 20(5):335—338
- [6] David H. Schoellhamer. Singular spectrum analysis for time series with missing data. *Geophys. Res. Lett.*, 2001, 28(16):3187—3190
- [7] Vautard R, Yiou P, Ghil M. Singular spectrum analysis: A toolkit for short noisy chaotic signals. *Physica D*, 1992, 58(1):95—126
- [8] Jiang Zhihong, Ding Yuguo. Generality and applied features for singular spectrum analysis. *ACTA Meteor. Sinica*, 1995, 56(6):736—744. in Chinese (江志红, 丁裕国. 奇异谱分析的广义性及其应用特色. 气象学报, 1995, 56(6):736—744)