中国科学: 物理学 力学 天文学 2016年 第46卷 第11期: 119601
SCIENTIA SINICA Physica, Mechanica & Astronomica physcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS

论 文

太阳黑子数平滑月均值的混合预测模型

田中大①*, 李树江①, 王艳红①, 王向东①, 沙毅②

① 沈阳工业大学信息科学与工程学院, 沈阳 110870;
② 东北大学计算机科学与工程学院, 沈阳 110819
*联系人, E-mail: tianzhongda@126.com

收稿日期: 2016-04-05; 接受日期: 2016-06-30; 网络出版日期: 2016-09-26
国家自然科学基金(编号: 11273001)和辽宁省博士启动基金(编号: 20141070)资助项目

摘要 提出一种太阳黑子数平滑月均值的混合预测模型. 通过最大Lyapunov指数得到太阳黑子数平滑月均值时间序列的最大可预测周期, 结果表明太阳黑子数平滑月均值序列的最大可预测周期为42个月. 太阳黑子数平滑月均值时间序列中包含着线性与非线性的成分, 利用自回归滑动平均模型对线性成分进行预测, 将太阳黑子数平滑月均值的实际值与自回归滑动平均模型的预测值作差值得到仅含有非线性成分的残差序列, 利用具有良好非线性预测能力的回声状态网络预测残差序列, 并通过人工蜂群算法来确定回声状态网络预测模型的最佳参数. 将自回归滑动平均模型预测值与回声状态网络预测的残差相加, 得到太阳黑子数平滑月均值的最终预测值. 通过第23太阳活动周的太阳黑子数平滑月均值的预测表明提出的预测模型具有较高的预测精度. 同时, 对第24太阳活动周的太阳黑子数平滑月均值进行了预测, 结果表明第24太阳活动周将在2020年2月结束.

关键词 太阳黑子数平滑月均值, 混合预测, 自回归滑动平均, 回声状态网络

PACS: 96.50.Wx, 96.60.-j, 96.60.Q-, 96.60.qd

1 引言

太阳黑子数是描述太阳活动水平的最基本的参量, 由于太阳黑子数与太阳耀斑及日冕物质抛射等爆发现象密切相关[1,2], 而这些爆发现象又引起地磁扰动. 准确的太阳黑子数预测可反映磁层、电离层和中高层大气扰动水平的状态, 从而可以为导航、定位、通信和低轨卫星轨道衰减的预报等提供重要的参考信息[3]. 因此, 太阳黑子数的精确预测具有非常重要的研究意义[4,5]. 在讨论和分析太阳黑子数的长期变化以及太阳活动水平时, 使用更为广泛的是太阳黑子数月均值的平滑值. 该值是国际上界定太阳活动周(即太阳黑子活动周)的位相和黑子数极值的参量, 因此也是大多数太阳活动周预测研究的主要预测量[6,7].

目前诸多学者对于太阳黑子数平滑月均值的预测问题进行了深入的研究. 其中基于相似周[8,9]与发动机模型[10]的两种物理预报模型取得了良好的预测效果. 而将太阳黑子数平滑月均值看作一个时间序列, 利用长期记录所积累的历史数据构造时间序列, 通过某种预测算法进行预测是另一个重要的研究方向. 近

引用格式: 田中大, 李树江, 王艳红, 等. 太阳黑子数平滑月均值的混合预测模型. 中国科学: 物理学 力学 天文学, 2016, 46: 119601
Tian Z D, Li S J, Wang Y H, et al. A hybrid prediction model of smoothed monthly mean sunspot number (in Chinese). Sci Sin-Phys Mech Astron, 2016, 46: 119601, doi: 10.1360/SSPMA2016-00191

© 2016 《中国科学》杂志社

几十年来, 以神经网络为代表的非线性预测模型因其广泛的适应能力和学习能力, 在非线性系统的预测方面得到了广泛的应用. 在太阳黑子数平滑月均值预测中应用最广泛的是径向基(RBF)神经网络[3,11–13]、回声状态网络[14]以及最小二乘支持向量机[15,16]等为代表的非线性预测模型. 虽然这些非线性预测模型能够以任意精度逼近非线性函数, 只需较少的统计学知识即可获得理想的预测效果. 但是RBF神经网络等都是典型的非线性预测模型, 而实际包括太阳黑子数平滑月均值在内的时间序列里面同时包含着线性成分与非线性成分. 也即是说, 如果能对太阳黑子数平滑月均值时间序列的线性与非线性成分分别进行预测, 将会极大地提高预测的精度.

基于以上讨论, 本文提出一种基于自回归滑动平均(ARMA)与回声状态网络(ESN)相结合的太阳黑子数平滑月均值混合预测模型. 首先利用ARMA模型预测太阳黑子数平滑月均值时间序列的线性成分. 将太阳黑子数平滑月均值实际值与ARMA模型预测值相减得到仅含有非线性成分的残差序列, 在此基础上, 利用具有良好非线性预测能力的回声状态网络预测该残差序列, 并通过人工蜂群算法确定回声状态网络的最佳预测参数. 将残差序列预测值与ARMA模型预测值相加, 得到最终的太阳黑子数平滑月均值的预测值. 通过第23太阳活动周的太阳黑子数平滑月均值的预测表明本文的预测方法具有较高的预测精度. 同时, 本文对第24太阳活动周的太阳黑子数平滑月均值进行了预测, 结果表明第24太阳活动周将在2020年2月结束.

2 太阳黑子数平滑月均值时间序列可预测性及预测尺度

本文的太阳黑子数平滑月均值时间序列来源于比利时皇家天文台太阳作用数据分析中心(http://sidc.oma.be/silso/datafiles), 该网站记录了从1749年至今的太阳黑子数据(2015年7月1日该数据分析中心对太阳黑子数进行了大幅度的修正, 本文的太阳黑子数平滑月均值数据来源于修正后的数据集), 太阳黑子的活动周期约为11.2年[17]. 在开始的4年左右时间里, 黑子不断产生, 越来越多, 活动加剧, 在黑子数达到极大的那一年, 称为太阳活动峰年. 在随后的7年左右时间里, 黑子活动逐渐减弱, 黑子也越来越少, 黑子数极小的那一年, 称为太阳活动谷年. 国际上规定, 从1755年起算的黑子周期为第1周, 然后顺序排列, 2009年开始为第24太阳活动周. 太阳黑子数平滑月均值是当前月的前6个月、后6个月以及当前月太阳黑子数月均值共计13个月的均值(–6月与+6月的权重为0.5, 其他11个月的权重为1, 13个月太阳黑子数月均值总和/12), 图1为1755年1月至2015年10月共计23周多的太阳黑子数平滑月均值与太阳黑子数月均值曲线图. 从图1中可看出, 太阳黑子数平滑月均值序列曲线的波形起伏性以及尖峰数量等都要少于太阳黑子数月均值序列, 从曲线上看, 太阳黑子数的平滑月均值具有更好的可预测性.

为了比较太阳黑子数月均值序列与太阳黑子数平滑月均值序列是否具有可预测性以及可预测性能的区别, 利用R/S类方法对2个时间序列分别计算了其Hurst指数, 计算结果如下图2所示. 结果显示太阳黑子数月均值序列的Hurst指数为0.7016, 太阳黑子数平滑月均值序列的Hurst指数为0.7374, 二者的Hurst指数都大于0.5, 说明两个时间序列在一定时段内具有一定的相关性, 即当前的数据会保持历史数据的变化趋势, 因此两个时间序列都具有可预测性. 同时太阳黑子数平滑月均值序列的Hurst指数大于太阳黑子数月均值序列的Hurst指数, 因此说明前者的可预测性是强于后者的.

太阳黑子数平滑月均值时间序列是一个典型的混沌系统[18], 对其进行长期预报是不可能的, 因此可预测的时间尺度是一个重要的参数, 只有在其可预测

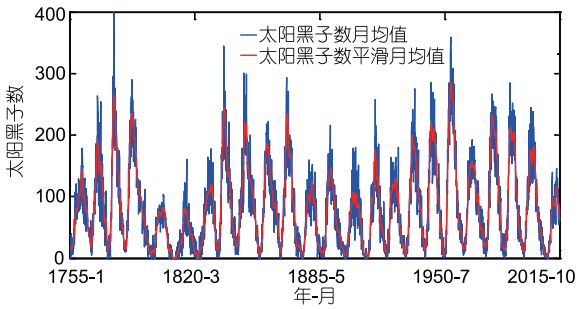

图 1 (网络版彩图)太阳黑子数平滑月均值与太阳黑子数月均值时间序列

Figure 1 (Color online) Smoothed monthly mean sunspot number and monthly mean sunspot number time series.

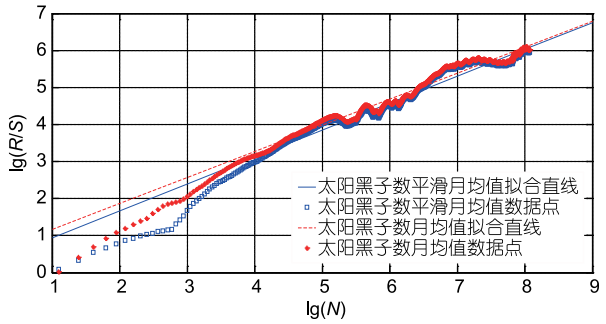

图 2 (网络版彩图)太阳黑子数月均值与太阳黑子数平滑月均值序列Hurst指数

Figure 2 (Color online) Hurst exponent of smoothed monthly mean sunspot number and monthly mean sunspot number time series.

时间尺度之内进行预测才是可靠的. 混沌系统的Lyapunov指数是反映系统混沌水平的一个变量, 通过最大Lyapunov指数可确定时间序列的最大预测步数.

计算最大Lyapunov指数的前提是对时间序列进行相空间重构, 相空间重构用原始系统中某个变量的延迟坐标来重构相空间, 重构的相空间的轨迹也反映系统状态的演化规律. 在相空间重构中, 延迟时间τ与嵌入维数m的选取十分重要, 本文采用C-C方法进行τ与m的选取, 该方法利用统计结果获得, 因此运算量少, 实现简单. C-C方法计算过程如下[19]:

取m=2, 3, 4, 5, $r_i=i\times0.5\sigma$, i=1, 2, 3, 4, 计算下列公式:

$$\bar{S}(t)=\frac{1}{16}\sum_{m=2}^{5}\sum_{i=1}^{4}S(m,r_i,t), \tag{1}$$

$$\Delta\bar{S}(t)=\frac{1}{4}\sum_{m=2}^{5}\Delta S(m,t), \tag{2}$$

$$S_{\rm cor}(t)=\Delta\bar{S}(t)+|\bar{S}(t)|, \tag{3}$$

式中, σ为时间序列标准差, t为时间延迟, $\bar{S}(t)$代表检验统计量均值, $\Delta\bar{S}(t)$为最大与最小检验统计量差值的均值, $S_{\rm cor}(t)$为定义的指标. 其中的$\bar{S}(t)$与$\Delta\bar{S}(t)$均可反映时间序列的自相关特性, 寻找$\bar{S}(t)$的第一次过零点或者$\Delta\bar{S}(t)$的第一个局部极小值所对应的t即为最佳时延τ, 同时综合考虑$\bar{S}(t)$与$\Delta\bar{S}(t)$, 寻找$S_{\rm cor}(t)$的全局极小值对应的t来获得最佳嵌入窗$\tau_{\rm w}$, 依据下式(4)来求取最佳的嵌入维数m:

$$\tau_{\rm w}=(m-1)\tau_d. \tag{4}$$

利用C-C方法计算可得到太阳黑子数平滑月均值时间序列$\bar{S}(t)$, $\Delta\bar{S}(t)$与$S_{\rm cor}(t)$的变化曲线如图3所示. 从图中可看出$\Delta\bar{S}(t)$的第一个局部最小值发生在t=1, 因此确定太阳黑子数平滑月均值时间序列的时延τ为1. $S_{\rm cor}(t)$的全局最小值位于t=49, 因此最佳嵌入窗$\tau_{\rm w}$为49, 根据公式(4)可得到嵌入维数m为50.

在确定相空间重构参数后, 本文采用Wolf法[20]计算最大Lyapunov指数, 方法如下: 相空间内两个相邻相点$Y(i)$与$Y(j)$的n次迭代后, 计算相点$Y(i+n)$与$Y(j+n)$距离分离的平均指数:

$$|Y(i+n)-Y(j+n)|=d(0)\cdot {\rm e}^{n\lambda_1}, \tag{5}$$

式中, $d(0)$为其初始距离, 令t_0为可预测的临界时间, 将上式转换为

$$\frac{|Y(i+n)-Y(j+n)|}{d(0)}={\rm e}^{\lambda_1 t_0}, \tag{6}$$

假设上式(6)超过临界值C时即可认为序列轨道发散无法预测.

$$C=\frac{|Y(i+n)-Y(j+n)|}{d(0)}={\rm e}^{\lambda_1 t_0}, \tag{7}$$

因此有$t_0=(1/\lambda_1)\ln C$, 取C=e, 则时间序列最大预测步数$t_0=(1/\lambda_1)$, 则λ_1为最大Lyapunov指数. λ_1越大, 则说明可预测时间越短, 蝴蝶效应越强.

通过计算得到的时延τ与嵌入维数m对太阳黑子数平滑月均值序列进行计算, 可得到最大Lyapunov指数λ_1为0.024, 则太阳黑子数平滑月均值序列的最大可预测周期为42个月. 从以上分析可知由于太阳黑子数

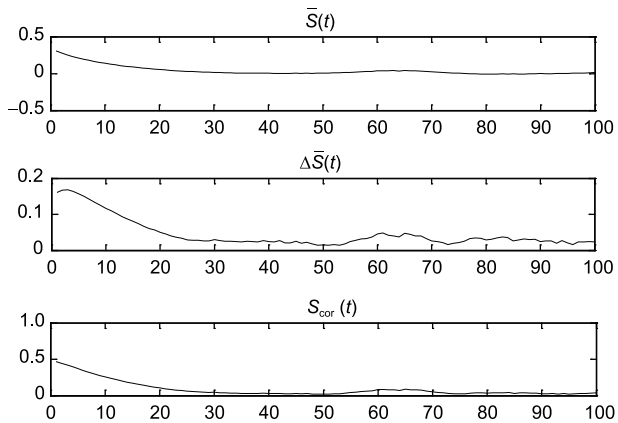

图 3 $\bar{S}(t)$, $\Delta\bar{S}(t)$与$S_{\rm cor}(t)$的变化曲线

Figure 3 Change curve of $\bar{S}(t)$, $\Delta\bar{S}(t)$and $S_{\rm cor}(t)$.

平滑月均值时间序列具有混沌特性, 对其进行长期预测是不准确的, 但是可进行短期预测.

3 ARMA模型

ARMA模型是目前最常见的一种平稳序列预测模型[21], 表示为ARMA(p, q), 具体表达式为[22]

$$\Phi(B)x_t = \Theta(B)\varepsilon_t, \tag{8}$$

式中, $\Phi(B)$为p阶自回归系数多项式, 可表示为$\Phi(B)=1-\phi B-\cdots-\phi_p B^p$; $\Theta(B)$为q阶移动平均系数多项式, 可表示为$\Theta(B)=1-\theta_1 B-\cdots-\theta_q B^q$; $\{x_t\}(t=1, 2, \cdots)$为时间序列; $\{\varepsilon_t\}$为均值为0、方差为σ^2_ε的正态白噪声; $\phi_i(i=1, 2, \cdots, p)$与$\theta_j(j=1, 2, \cdots, q)$为模型待估系数; B为后移差分算子. ARMA模型的定阶一般利用Akaike给出的AIC准则. 定义AIC函数为

$$\mathrm{AIC} = -2\ln L + 2n, \tag{9}$$

其中L为模型的极大似然参数, n为模型的独立参数格式. 利用AIC准则来定阶是指在p, q的一定变化范围内寻求使得统计量$A(s)$达到最小的点$(\hat{p}, \hat{q})$, 用来作为(p, q)的估计.

4 人工蜂群算法优化的回声状态网络

4.1 回声状态网络

回声状态网络(Echo State Network, ESN)的网络隐层是一个具有很多神经元组成的稀疏网络, 通过网络内部权值的调整来实现记忆数据的功能. 其内部的动态储备池(Dynamic Reservoir, DR)包含了大量稀疏连接的神经元, 蕴含系统的运行状态, 并具有短期记忆功能, 而非线性系统的动态特性即由DR产生. 回声状态网络一经提出便成为学术界的研究热点, 并应用到包括时间序列预测[23–25]等各种不同领域之中.

设系统具有K个输入单元, N个内部处理单元, 同时具有L个输出单元, 输入单元、内部状态、输出单元在时刻k的值分别为

$$\begin{aligned}\boldsymbol{u}(k) &= [u_1(k), u_2(k), ..., u_K(k)]^{\mathrm{T}},\\ \boldsymbol{x}(k) &= [x_1(k), x_2(k), ..., x_N(k)]^{\mathrm{T}},\\ \boldsymbol{y}(k) &= [y_1(k), y_2(k), ..., y_L(k)]^{\mathrm{T}},\end{aligned} \tag{10}$$

则ESN的学习步骤为

$$\boldsymbol{x}(k+1) = f(\boldsymbol{W}_{\mathrm{in}} \times \boldsymbol{u}(k+1) + \boldsymbol{W}\boldsymbol{x}(k) + W_{\mathrm{back}}\boldsymbol{y}(k)), \tag{11}$$

$$\boldsymbol{y}(k+1) = f_{\mathrm{out}} \times (\boldsymbol{W}_{\mathrm{out}} \times (\boldsymbol{u}(k+1), \boldsymbol{x}(k+1))), \tag{12}$$

其中, f, f_{out}分别为DR内部神经元、输出层神经元激活函数. $\boldsymbol{W}_{\mathrm{in}}$, $\boldsymbol{W}$, $\boldsymbol{W}_{\mathrm{out}}$, $\boldsymbol{W}_{\mathrm{back}}$分别为输入层到储备池的$N\times K$阶, 储备池的$N\times N$阶, 储备池到输出层的$L\times(K+N+L)$阶, 输出层反馈到储备池的$N\times L$连接权值矩阵. 利用网络实际输出$\hat{\boldsymbol{y}}(k)$, 逼近期望输出$\boldsymbol{y}(k)$:

$$\boldsymbol{y}(k) \approx \hat{\boldsymbol{y}}(k) = \sum_{i=1}^{L} \boldsymbol{w}_i^{\mathrm{out}} x_i(k). \tag{13}$$

也就是希望计算权值矩阵满足系统均方误差最小, 即求解如下目标:

$$\min \frac{1}{P-m+1} \sum_{k=m}^{P} \left(\boldsymbol{y}(k) - \sum_{i=1}^{L} \boldsymbol{w}_i^{\mathrm{out}} x_i(k) \right)^2. \tag{14}$$

可归结为如下形式:

$$\boldsymbol{W}_{\mathrm{out}} = (\boldsymbol{M}^{-1} \times \boldsymbol{T})^{\mathrm{T}}, \tag{15}$$

$\boldsymbol{M}$为输入$x_1(i)$, $x_2(i)$, $\cdots$, $x_N(i)$, $i=m$, $m+1$, $\cdots$, P, 构成的$(P-m+1)\times N$的矩阵, $\boldsymbol{T}$为输出$\boldsymbol{y}(k)$构成的$(P-m+1)\times 1$的列矩阵.

ESN的性能由储备池的4个参数决定的[26]. 分别是储备池内部连接权谱半径SR, 储备池规模N, 储备池输入单元尺度IS, 储备池稀疏程度SD. 这4个参数极大地影响了ESN模型的预测能力[27]. 目前的一些文献中只给出了ESN算法4个参数取值的定性分析, 并没有指出如何能够确定最佳的预测参数, 本文利用人工蜂群算法(Artificial Bee Colony, ABC)进行ESN储备池参数的优化.

4.2 人工蜂群算法

人工蜂群(Artificial Bee Colony, ABC)算法来源于蜜蜂采蜜的行为而提出的一种智能优化算法[28]. ABC算法相对于遗传算法、差分进化算法以及粒子群优化算法具有极强的竞争力[29]. 在很多实际问题的优化中得到了广泛的应用[30–32]. 基于此, 本文采用ABC算法对ESN的参数进行优化, 并用得到的优化参数构建最优的ESN预测模型.

ABC算法中蜜源的位置抽象为解空间内的点, 蜜源($i=1, 2, \cdots, SN$)的质量为解的适应度. 设求解问题的维数为D, 对于第t次迭代蜜源i的位置可表示为

$\boldsymbol{X}_i^t = [x_{i1}^t\ x_{i2}^t\ x_{iD}^t]$, $x_{iD} \in (L_d, U_d)$, 其中L_d为搜索空间下限, U_d为搜索空间上限, 蜜源的位置按式(16)在搜索空间随机产生:

$$x_{id} = L_d + \mathrm{rand}(0,1)(U_d - L_d), \tag{16}$$

搜索开始阶段, 引领蜂在蜜源周围根据式(17)搜索而产生一个新的蜜源:

$$v_{id} = x_{id} + \varphi(x_{id} - x_{jd}), \tag{17}$$

式中, $j \in \{1, 2, \cdots, SN\}$, $j \neq i$, 代表在蜜源中随机选择一个不同于i的蜜源; $\varphi \in [-1, 1]$. 当新的蜜源$\boldsymbol{V}_i=[v_{i1}\ v_{i2} \cdots v_{iD}]$的适应度优于$\boldsymbol{X}_i$时, 利用贪婪算法将$\boldsymbol{V}_i$代替$\boldsymbol{X}_i$, 否则保留$\boldsymbol{X}_i$. 然后跟随蜂根据所有引领蜂的蜜源分享信息, 按下式概率进行跟随:

$$p_i = \frac{\mathrm{fit}_i}{\sum_{i=1}^{SN} \mathrm{fit}_i}, \tag{18}$$

也即跟随蜂产生一个[0, 1]间的随机数并与p_i进行比较, 如果该随机数小于p_i则按式(17)产生一个新蜜源. 搜索中, 蜜源$\boldsymbol{X}_i$经过若干次迭代到达阈值而没有找到更好的蜜源, 则$\boldsymbol{X}_i$将被抛弃, 与之对应的引领蜂转变为侦查蜂, 并按式(19)随机产生一个新的蜜源以替代$\boldsymbol{X}_i$:

$$\boldsymbol{X}_i^{t+1} = \begin{cases} L_d + \mathrm{rand}(0,1)(U_d - L_d), \mathrm{trail} \geq \mathrm{limit}, \\ x_i^t, \mathrm{trail} < \mathrm{limit}. \end{cases} \tag{19}$$

为了获取最小的均方根误差(RMSE)以及和ABC算法保持一致, 采用下式(20)作为ABC优化ESN参数的适应度函数:

$$\mathrm{fit}_i = \frac{1}{1 + f_i}, \tag{20}$$

其中f_i为实际值与预测值的均方根误差, 可表示为

$$f_i = \sqrt{\frac{1}{SN}\sum_{j=1}^{SN}(y_j - \hat{y}_j)^2}, \tag{21}$$

式(21)中, y_j为实际值, $\hat{y}_j$为预测值.

4.3 人工蜂群算法优化的回声状态网络

ABC算法优化回声状态网络预测模型的步骤如下:

(1) 生成训练数据样本集, 同时参数初始化: 包括蜜源的个数SN, 最大迭代次数MCN, 蜜源最大开采次数为limit, 待优化的参数SR, N, IS, SD的取值范围, 令t=1;

(2) 为蜜源分配一只引领蜂, 按照式(17)进行搜索, 产生新蜜源$\boldsymbol{V}_i$;

(3) 代入样本数据, 根据高斯过程回归模型计算输出预测值, 并根据式(20)计算适应度值, 根据贪婪算法保留蜜源;

(4) 按式(18)计算蜜源被跟随的概率, 跟随蜂搜索, 根据贪婪算法保留蜜源;

(5) 判断蜜源$\boldsymbol{X}_i$是否应抛弃, 若是, 则引领蜂转变为侦查蜂, 否则转(7);

(6) 侦查蜂按式(19)产生新的蜜源;

(7) t=t+1, 如果满足终止条件, 输出最优参数, 否则转到(2)继续执行.

5 基于ARMA与ESN的太阳黑子数平滑月均值混合预测模型

本文通过ARMA(p, q)模型来进行太阳黑子数平滑月均值时间序列的线性部分的预测, 利用ABC算法优化的ESN完成太阳黑子数平滑月均值非线性残差时间序列的预测, 二者的预测值进行叠加获得最终的预测值, 达到提高预测精度的目的. 假设当前时刻为k, 太阳黑子数平滑月均值时间序列的实际值为$S(k)$, 线性部分可表示为$L(k)$, 非线性残差部分表示为$N(k)$, 那么$S(k)$可表示为$L(k)$与$N(k)$的叠加, 即

$$S(k) = L(k) + N(k). \tag{22}$$

首先, 利用ARMA模型进行$L(k)$的预测, 设其预测值为$\hat{L}(k)$, 有

$$\begin{aligned} \hat{L}(k) = {} & \theta_0 + \varphi_1 S(k-1) + \varphi_2 S(k-2) \\ & + \cdots + \varphi_p S(k-p) + \varepsilon_t - \theta_1 \varepsilon_{k-1} \\ & - \theta_2 \varepsilon_{k-2} - \cdots \theta_q \varepsilon_{k-q}. \end{aligned} \tag{23}$$

利用ARMA模型进行多步预测时, 将$\hat{L}(k)$作为实际值再次代入公式(23)即可完成$\hat{L}(k+1)$的预测, 以此类推迭代循环实现多步预测功能.

同时, 在k时刻的预测残差为

$$N(k) \approx E(k) = S(k) - \hat{L}(k). \tag{24}$$

利用ESN对$N(k)$进行预测, 有

$$[\widehat{N}(k),\widehat{N}(k+1),...,\widehat{N}(k+L)]^{\mathrm{T}} = W_{\mathrm{out}}[E(k-m),E(k-m+1),...,E(k-1)]^{\mathrm{T}}. \tag{25}$$

式中, k为采样时刻, m为嵌入维数, 也就是ESN的输入神经元个数, L为ESN的预测步长. 综上所述, 本文的基于ARMA与ESN的太阳黑子数平滑月均值混合预测模型可如图4所示.

综述, 本文的预测方法可描述为如下步骤:

(1) 采集太阳黑子数平滑月均值数据, 将数据分为训练集以及测试集, 训练集用于ARMA与ESN预测模型的建立, 测试集用于预测模型预测精度的验证. 同时需要将太阳黑子数平滑月均值序列进行归一化, 预测结果则需要反归一化处理, 恢复真实预测值;

(2) 利用训练样本序列进行ARMA模型的训练, 首先对太阳黑子数平滑月均值序列结合AIC准则确定最佳的预测模型参数p, q, 完成ARMA预测模型的建立;

(3) 将原始样本序列减去ARMA模型预测值得到残差序列, 利用该残差序列进行ESN预测模型的训练, 通过ABC算法进行ESN预测模型预测参数的优化;

(4) 当ARMA与ESN预测模型确定后, 利用ARMA模型预测未来的太阳黑子数平滑月均值, 利用ESN模型预测太阳黑子数平滑月均值的残差, 将二者相加得到最终的预测值. 为了减少预测步长超出最大可预测步长带来的误差积累效应, 本文在预测时每次对未来连续42个月的太阳黑子数平滑月均值进行预测, 然后采用滑动窗口机制将实际的太阳黑子数平滑月均值放入输入集合, 再进行连续42个月的太阳黑子数平滑月均值多步预测, 以此达到提高预测精度、减少预测误差积累的目的.

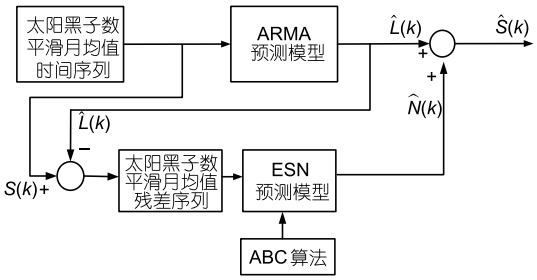

图 4 太阳黑子数平滑月均值混合预测模型

Figure 4 Hybrid prediction model of smoothed monthly mean sunspot number.

6 仿真

由于1818年开始才有了比较常规的太阳黑子数观测, 同时1848年瑞士天文学家沃尔夫首次提出太阳黑子相对数的概念, 因此本文对于训练集采用的是1848年之后的数据. 将1855年12月开始的第10太阳黑子活动周到1986年9月结束的第21太阳黑子活动周的太阳黑子数平滑月均值数据作为训练输入集合, 第22太阳黑子活动周(1986年9月到1996年5月)的太阳黑子数平滑月均值数据作为输出集合完成ARMA与ESN预测模型的建立.

首先进行ARMA模型的训练, 通过输入输出数据, 按照第3节中介绍的AIC准则可得到ARMA模型参数为p=9, q=10, 对应的AIC为0.3524. 利用ARMA(9, 10)模型对训练集进行预测, 得到太阳黑子数平滑月均值预测残差序列如图5所示.

得到太阳黑子数平滑月均值预测残差序列之后, 通过ABC算法确定ESN的预测模型参数. 参数如下: ABC算法蜜源的个数SN为20, 最大迭代次数MCN为100, 蜜源最大开采次数为limit为50, 由于待优化的参数为4个, 则最优解是4维向量. 待优化的储备池参数的变化范围为$SR \in [0.01, 1)$, $N \in [10, 200]$, $IS \in [0.01, 1)$, $SD \in [0.01, 1)$. 按照上文步骤利用ABC算法进行参数优化. 优化之后得到的参数SR为0.7336, N为158, IS为0.4671, SD为0.6582.

利用第11太阳活动周(始于1867年3月)至第22太阳活动周(止于1996年5月)的太阳黑子数平滑月均值数据作为输入测试集, 对第23太阳活动周(1996年5月到2008年1月)太阳黑子数平滑月均值进行预测, 检验

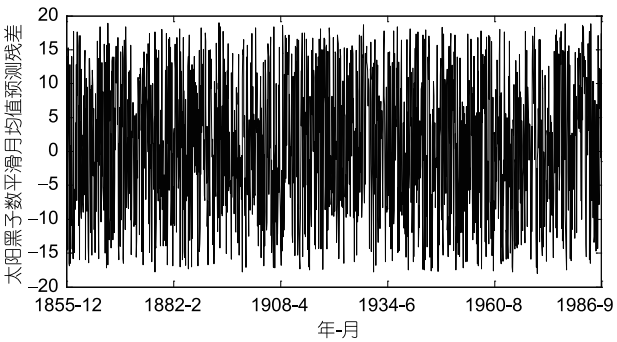

图 5 太阳黑子数平滑月均值预测残差序列

Figure 5 Prediction error series of smoothed monthly mean sunspot number.

本文预测方法的有效性. 首先利用ARMA模型对第23太阳活动周太阳黑子数平滑月均值的线性成分进行预测. 利用ESN模型对ARMA模型的非线性预测残差进行预测. ARMA与ESN两种预测模型的预测结果如图6所示.

将ARMA模型预测的太阳黑子数平滑月均值与ESN预测的非线性预测残差进行叠加, 得到最终的预测结果如图7所示. 为了对比本文预测方法的效果, 选择了文献[13]中的RBF神经网络(输入参数p=10, q=5, 隐层节点k=10, 高斯径向基函数参数λ=1)、文献[14]中的ESN神经网络(通过粒子群算法优化得到SR=0.6239, N=82, IS=0.4131, SD=0.3561)以及文献[15]中的多核LSSVM(优化参数为ρ=0.91, γ=5.93, σ^2=7.25, q=0.93) 3种预测方法作为对比. 为了对比方便, 预测结果也显示在图7中.

图8为本文方法与其他3种方法的绝对误差与相对误差对比分布图. 从图7的预测值与实际值曲线的符合度以及图8的绝对误差与相对误差分布图可看出本文的预测方法在预测精度以及预测误差方面都优于其他3种方法.

为了进一步进行预测效果的对比, 本文引入如下4种性能指标:

(1) RMSE(均方根误差)

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{k=1}^{N}(S(k)-\hat{S}(k))^2}{N}}, \tag{26}$$

(2) MAE(平均绝对误差)

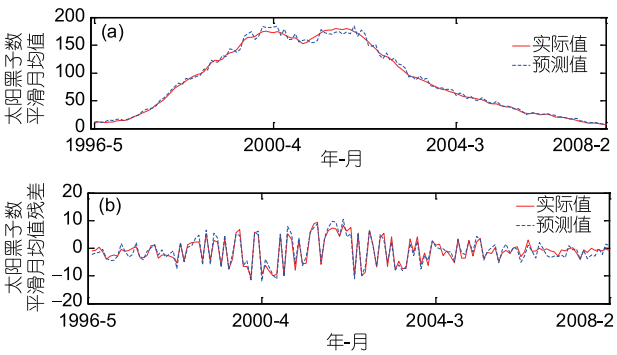

图 6 (网络版彩图) ARMA(a)与ESN(b)模型预测值与实际值对比曲线

Figure 6 (Color online) Contrast curve of prediction and actual value. (a) ARMA; (b) ESN.

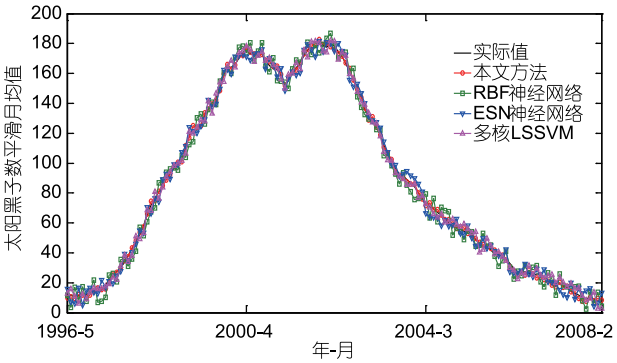

图 7 (网络版彩图)不同预测方法太阳黑子数平滑月均值预测值与实际值对比曲线

Figure 7 (Color online) Contrast curve of prediction and actual value with several prediction methods.

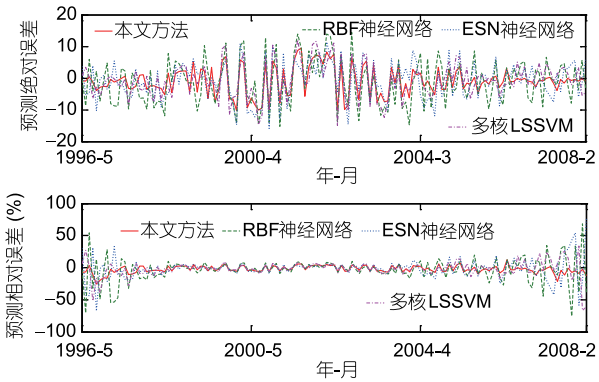

图 8 (网络版彩图)不同预测方法太阳黑子数平滑月均值绝对误差与相对误差分布

Figure 8 (Color online) Absolute and relative error distribution of several prediction methods.

$$\mathrm{MAE}=\frac{1}{N}\sum_{k=1}^{N}|S(k)-\hat{S}(k)|, \tag{27}$$

(3) MAPE(平均绝对百分误差)

$$\mathrm{MAPE}=\frac{1}{N}\sum_{k=1}^{N}|S(k)-\hat{S}(k)|\times 100/S(k), \tag{28}$$

(4) 可靠性

$$R^{(1-a)}=\left[\frac{\xi^{(1-a)}}{N}-(1-a)\right]\times 100\%, \tag{29}$$

式中, N为样本数量, $S(k)$为实际值, $\hat{S}(k)$为预测模型的预测值, $\xi^{(1-\alpha)}$为在置信度1–α下实际值落入预测置信区间的个数.

表1为本文方法与其他3种方法的预测性能指标

的对比.从表1中可观察到本文的预测方法在RMSE, MAE与MAPE这3种性能指标上都要优于其他的3种预测方法.

图9为本文方法与其他3种方法的可靠性的对比, 从实际值落入预测置信区间的个数上也可看出本文的混合预测方法要优于其他的3种预测方法.

本文预测方法预测精度提高的主要原因有两点. 首先采用混合预测模型, 也即利用ARMA模型预测太阳黑子平滑月均值数时间序列中的线性成分, 使得预测残差仅包含非线性成分, 利用ABC算法优化的ESN来进行具有非线性特征的太阳黑子数平滑月均值残差序列的预测. 由于混合预测模型利用了两个预测模型自身各自的优点, 因此使得预测误差得到了进一步的降低. 其次是预测步长没有超出太阳黑子数平滑月均值时间序列的最大可预测长度, 利用滑动窗口机制, 将陈旧的样本数据剔除出去, 而将实际数据放入样本队列中. 这两点共同保证了预测的可靠性与精确性.

目前是第24太阳活动周, 起始于2008年1月, NASA预测结束时间是2019年10月. 本文对第24太阳活动周的太阳黑子数平滑月均值数据进行了外推预测(文中对第24太阳活动周预测与第23太阳活动周预测存在一点区别. 由于第23太阳活动周的数据是完备的, 也就是存在实际的平滑月均值, 因而可以通过滑动窗口机制将42个实际值放入输入样本集合的过程. 但是在进行第24太阳活动周预测时, 2015年11月之后太阳黑子平滑月均值实际值还未获得, 因而预测时是将42个预测值作为实际值放入输入序列之中, 再进行未来42个月的预测, 连续进行, 一直到预测结束). 训练输入集合为第12太阳活动周(始于1878年12月)至第23太阳活动周(止于2008年1月)的太阳黑子数平滑月均值数据, 预测曲线如下图10所示.

从图中10可看出, 对于2015年10月前的数据, 预测值与实际值符合得较好, 2015年11月之后的外推预测显示太阳黑子数平滑月均值在2020年2月达到最小值2.2, 因此本文认为第24太阳活动周结束时间为2020年2月, 本次太阳活动周时间跨度约为12.1年. 文献[33]中预测第24太阳活动周结束于2020年6月, 本文预测结果介于NASA与文献[33]之间. 由于本文中太阳黑子数据采用的是2015年7月之后更新的新数据, 因此本文预测的太阳黑子数平滑月均值数较文献[33]预测的要大, 预报的准确性需要今后的实际数据去检验. 同时, 由于2015年11月之后缺少实际数据的监督作用, 同时预测步长超过最大可预测步长, 这就导致预测误差会逐渐累积进入预测模型中, 也就是过去的误差会造成未来误差的增大, 但是预测序列的整体演变还是符合太阳黑子数平滑月均值时间序列的变化趋势.

表 1 预测方法性能指标对比

Table 1 Comparison of performance index of prediction methods

预测方法	RMSE	MAE	MAPE
本文方法	3.1236	2.9132	0.1367
RBF神经网络	7.2434	5.0393	0.2923
ESN神经网络	6.0173	4.9047	0.2185
多核LSSVM	5.6295	4.4224	0.2147

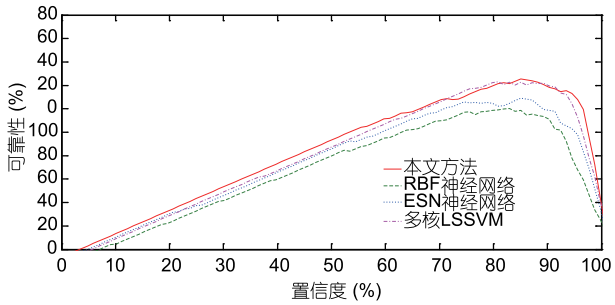

图 9 (网络版彩图)预测方法的可靠性对比曲线

Figure 9 (Color online) Reliability contrast curve of prediction methods.

7 结论

本文针对太阳黑子数平滑月均值时间序列的预测问题而提出一种基于自回归滑动平均模型与回声状态

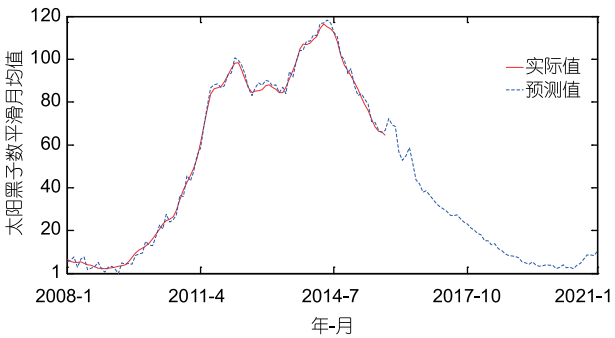

图 10 (网络版彩图)第24太阳活动周太阳黑子数平滑月均值预测

Figure 10 (Color online) Prediction of smoothed monthly mean sunspot number for cycle 24.

网络的混合预测模型. 首先利用自回归滑动平均模型预测太阳黑子数平滑月均值时间序列中的线性趋势分量, 然后得到具有非线性特征的太阳黑子数平滑月均值残差序列, 通过人工蜂群算法优化的回声状态网络进行残差的预测. 最后线性的自回归滑动平均模型预测值与非线性的回声状态网络预测残差值进行叠加而得到最终的预测值. 通过仿真对比表明本文的预测方法具有更高的预测精度以及更小的预测误差.

致谢 感谢评审人提出的建设性意见.

参考文献

1 Zhao M Y, Chen J Q, Liu Y, et al. Statistical analysis of sunspot groups and flares for solar maximum and minimum (in Chinese). Sci Sin-Phys Mech Astron, 2014, 44: 109–120 [赵明宇, 陈军权, 刘煜, 等. 太阳活动峰年和谷年期间黑子群与耀斑的统计分析. 中国科学: 物理学 力学 天文学, 2014, 44: 109–120]

2 Wang J X, Zhang Y Z, He H, et al. Cluster of solar active regions and onset of coronal mass ejections. Sci China-Phys Mech Astron, 2015, 58: 599601

3 Tang J, Zhang W. Prediction of smoothed monthly mean sunspot number based on chaos theory (in Chinese). Acta Phys Sin, 2012, 61: 169601 [唐洁, 张维. 基于混沌理论的太阳黑子数平滑月均值预报. 物理学报, 2012, 61: 169601]

4 Gkana A, Zachilas L. Sunspot numbers: Data analysis, predictions and economic impacts. J Eng Sci Technol Rev, 2015, 8: 79–85

5 Kane R P. An estimate for the size of sunspot cycle 24. Sol Phys, 2013, 282: 87–90

6 Silbergleit V M. Probable value for the next sunspot minimum. Adv Astron, 2014, 2014: 1–4

7 Wang J L, Gong J C, Liu S Q, et al. The prediction of maximum amplitudes of solar cycles and the maximum amplitude of solar cycle 24. Chin J Astron Astrophys, 2002, 2: 557–562

8 Du Z L, Wang H N. The prediction method of similar cycles. Res Astron Astrophys, 2011, 11: 1482–1492. arXiv: 1110.6436

9 Du Z L, Wang H N. Predicting the solar maximum with the rising rate. Sci China-Phys Mech Astron, 2012, 55: 365–370. arXiv: 1112.5721

10 Jiang J, Chatterjee P, Choudhuri A R. Solar activity forecast with a dynamo model. Mon Not R Astronl Soc, 2007, 381: 1527–1542. arXiv: 0707.2258

11 Ding L, Jiang Y, Lan R. Prediction of the smoothed monthly mean sunspot area using artificial neural network. In: Proceedings of International Conference on Information and Computing Science. Washington DC: IEEE, 2012. 33–36

12 Zhang D Q, Ning X X, Liu X N. On-line prediction of nonlinear time series using RBF networks (in Chinese). Control Theory Appl, 2009, 26: 151–155 [张冬青, 宁宣熙, 刘雪妮. 基于神经网络的非线性时间序列在线预测. 控制理论与应用, 2009, 26: 151–155]

13 Zhang D, Han Y. Time series prediction with rbf neural networks. Inf Tech J, 2013, 12: 2815–2819

14 Yusoff M H, Jin Y C. Modeling neural plasticity in echo state networks for time series prediction. In: Proceedings of 14th UK Workshop on Computational Intelligence. Piscataway: IEEE, 2014. 6930163

15 Tian Z D, Gao X W, Shi T. Combination kernel function least squares support vector machine for chaotic time series prediction (in Chinese). Acta Phys Sin, 2014, 63: 160508 [田中大, 高宪文, 石彤. 用于混沌时间序列预测的组合核函数最小二乘支持向量机. 物理学报, 2014, 63: 160508]

16 Ismail S, Shabri A, Samsudin R. A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting. Expert Syst Appl, 2011, 38: 10574–10578

17 Wang X, Zhang Q B. Evidence of solar signals in tree rings of Smith fir from Sygera Mountain in southeast Tibet. J Atmospheric Solar-Terrestrial Phys, 2011, 73: 1959–1966

18 Arlt R, Weiss N. Solar activity in the past and the chaotic behaviour of the dynamo. Space Sci Rev, 2014, 186: 525–533. arXiv: 1406.7628

19 Khatibi R, Sivakumar B, Ghorbani M A, et al. Investigating chaos in river stage and discharge time series. J Hydrology, 2012, 414-415: 108–117

20 Wolf A, Swift J B, Swinney H L, et al. Determining Lyapunov exponents from a time series. Phys D-Nonlinear Phenomena, 1985, 16: 285–317

21 Han M, Xu M L. A hybrid prediction model of multivariate chaotic time series based on error correction (in Chinese). Acta Phys Sin, 2013, 62: 120510 [韩敏, 徐美玲. 一种基于误差补偿的多元混沌时间序列混合预测模型. 物理学报, 2013, 62: 120510]

22 Sun G Q, Wei Z N, Zhai W X. Short term wind speed forecasting based on RVM and ARMA error correcting (in Chinese). Trans China Electrotech Soc, 2012, 27: 187–193 [孙国强, 卫志农, 翟玮星. 基于RVM与ARMA误差校正的短期风速预测. 电工技术学报, 2012, 27: 187–193]

23 Ongenae F, Van Looy S, Verstraeten D, et al. Time series classification for the prediction of dialysis in critically ill patients using echo statenetworks. Eng Appl Artif Intell, 2013, 26: 984–996
24 Sheng C, Zhao J, Liu Y, et al. Prediction for noisy nonlinear time series by echo state network based on dual estimation. Neurocomputing, 2012, 82: 186–195
25 Li D C, Han M, Wang J. Chaotic time series prediction based on a novel robust echo state network. IEEE Trans Neural Netw Learning Syst, 2012, 23: 787–799
26 Ozturk M C, Xu D, Príncipe J C. Analysis and design of echo state networks. Neural Comput, 2007, 19: 111–138
27 Jaeger H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science, 2004, 304: 78–80
28 Karaboga D, Basturk B. A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. J Glob Optim, 2007, 39: 459–471
29 Zhou X Y, Wu Z J, Wang M W. Artificial bee colony algorithm based on orthogonal experimental design (in Chinese). J Software, 2015, 26: 2167–2190 [周新宇, 吴志健, 王明文. 基于正交实验设计的人工蜂群算法. 软件学报, 2015, 26: 2167–2190]
30 Imanian N, Shiri M E, Moradi P. Velocity based artificial bee colony algorithm for high dimensional continuous optimization problems. Eng Appl Artif Intell, 2014, 36: 148–163
31 Kang F, Li J, Li H. Artificial bee colony algorithm and pattern search hybridized for global optimization. Appl Soft Computing, 2013, 13: 1781–1791
32 Draa A, Bouaziz A. An artificial bee colony algorithm for image contrast enhancement. Swarm Evolut Comput, 2014, 16: 69–84
33 Miao J, Gong J C, Li Z T, et al. The prediction of maximum amplitude of solar cycle 25 (in Chinese). Sci Sin-Phys Mech Astron, 2015, 45: 099601 [苗娟, 龚建村, 李志涛, 等. 第25太阳周太阳黑子数峰值预测. 中国科学: 物理学 力学 天文学, 2015, 45: 099601]

A hybrid prediction model of smoothed monthly mean sunspot number

TIAN ZhongDa[1*], LI ShuJiang[1], WANG YanHong[1], WANG XiangDong[1] & SHA Yi[2]

[1] *College of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China;*
[2] *College of Computer Science and Engineering, Northeastern University, Shenyang 110819, China*

A hybrid prediction model of smoothed monthly mean sunspot number is proposed in this paper. The maximum predictable time span of smoothed monthly mean sunspot number is calculated by the maximum Lyapunov exponent method. The results show that the maximum predictable time span of smoothed monthly mean sunspot number is 42 months. The smoothed monthly mean sunspot number time series has linear and nonlinear components. Auto regressive moving average model is utilized to predict the linear component. The nonlinear error component can be obtained by the actual value of smoothed monthly mean sunspot number minus the prediction value by auto regressive moving average model. Then, echo state network with good nonlinear prediction ability is utilized to predict nonlinear error component. At the same time, artificial bee colony algorithm is used for parameters optimization of echo state network prediction model. The final prediction value of smoothed monthly mean sunspot number will be obtained by the prediction value of auto regressive moving average model added to the prediction value of echo state network. The simulation of the twenty-third solar cycle smoothed monthly mean sunspot number shows that the proposed model has higher prediction accuracy. At the same time, smoothed monthly mean sunspot number of twenty-fourth solar cycle is predicted, the results show that the twenty-fourth solar cycle will end in February 2020.

smoothed monthly mean sunspot number, hybrid prediction, auto regressive moving average, echo state network

PACS: 96.50.Wx, 96.60.-j, 96.60.Q-, 96.60.qd

doi: 10.1360/SSPMA2016-00191