0254-6124/2024/44(6)-1012-09 *Chin. J. Space Sci.* 空间科学学报

JIAO Qirong, ZHANG Dianjun, LIU Wenlong. *Kp* Index Forecast Model Based on GBR Method (in Chinese). *Chinese Journal of Space Science*, 2024, **44**(6): 1012–1020. CSTR:32142.14.cjss.2024–0011. DOI:10.11728/cjss2024.06.2024–0011

基于 GBR 方法的 *Kp* 指数预报模型*

焦琦融　张典钧　刘文龙

(北京航空航天大学空间与环境学院　北京　102206)

(北京航空航天大学 空间环境监测与信息处理工信部重点实验室　北京　102206)

摘　要　地磁 *Kp* 指数是空间天气预警的重要指标, 也是研究太阳风–磁层耦合的关键参数. 采用梯度提升回归 (GBR) 算法和随机森林 (RF) 两种机器学习方法, 构建了以太阳风、行星际磁场参数、历史 *Kp* 值和太阳黑子数据为输入的 3 h 地磁 *Kp* 指数预报模型. 预报结果表明, 两种方法均可提前 1 h 预报地磁 *Kp* 指数, 预测结果与观测值之间的相关系数为 0.90, 其中 GBR 方法在均方根误差上表现出更好的效果, 均方根误差为 0.56. *Kp* 指数预报模型在太阳活动周不同相位的预测结果存在差异, 在活动周下降阶段模型预测结果与观测数据的相关系数更高. 比较了不同地磁扰动下模型的预测情况, 相比中等磁暴和超强磁暴, 模型对强磁暴 ($6 \leqslant Kp < 7$) 的预报准确度最高.

关键词　*Kp* 指数, 机器学习, 空间天气预报

中图分类号　P353

Kp Index Forecast Model Based on GBR Method

JIAO Qirong　ZHANG Dianjun　LIU Wenlong

(*School of Space and Environment, Beihang University, Beijing* 102206)

(*Key Laboratory of Space Environment Monitoring and Information Processing of MIIT, Beihang University, Beijing* 102206)

Abstract　The solar wind transports solar activity energy to interplanetary space, causing changes in the spatial structure of the Earth's magnetosphere and causing disastrous space weather. The *Kp* index is an important indicator for space weather alerts and a key parameter for the coupling between solar wind and the magnetosphere. With the development of machine learning methods, more and more space weather forecasting works adopt this method. In this paper, two machine learning methods, Gradient Boosting Regression (GBR) algorithm and Random Forest (RF), are used to construct a 3-hour *Kp* index prediction model with solar wind, interplanetary magnetic field parameters, historical *Kp* values and sunspot data as inputs. The forecast results show that our methods can predict the *Kp* index one hour in advance and the correlation coefficient is 0.90 between the *Kp* index of the optimal case recommended by the model and the actual value. The GBR model performs better, the root Mean Square Error (E_{rms}) is 0.56, and the Prediction Efficiency (P) is 0.81. The *Kp* index prediction model shows varying performances in different solar cycle phases, with better result during the cycle descending phase. The high-

* 国家自然科学基金项目资助 (42304175, 41974194)

2024-01-18 收到原稿, 2024-05-23 收到修定稿

©The Author(s) 2024. This is an open access article under the CC-BY 4.0 License (https://creativecommons.org/licenses/by/4.0/)

speed solar wind drive dominates the magnetospheric dynamics, and the model with solar wind as the main input parameter in the cycle descending phase has a better prediction effect. The model prediction situations under different geomagnetic disturbances have been compared. Compared with moderate and super severe magnetic storms, the model has the highest prediction accuracy for severe magnetic storms ($6 \leqslant Kp < 7$). In this study, the results of different prediction models are compared and analyzed. The prediction model can not only provide early warning of severe space weather, but also better understand the relationship between geomagnetic index and solar wind input energy, which provides more methods and theoretical basis for the research work of solar wind-magnetosphere coupling.

Key words *Kp* index, Machine learning, Space weather forecast

0 引言

Kp 指数是衡量地磁活动整体水平的重要指标之一[1,2], 由德国科学家 Bartels[3] 利用早期地磁观测资料首次提出 *Kp* 指数概念. *Kp* 指数是由全球 13 个地磁站的 *K* 指数的加权平均值计算得到的, 时间分辨率为 3 h, 每日 8 个值. *Kp* 指数按从 0 到 9 共分为 28 个等级: 0, 0^+, 1^-, 1, 1^+, ..., 9^-, 9, 用以表示每个时间段内磁场偏离磁静日的程度. 美国国家海洋和大气管理局 (National Oceanic and Atmospheric Administration, NOAA) 将 $Kp>5$ 认证为发生了地磁暴事件, 并按照 *Kp* 值的大小将磁暴划分为 5 个等级. 因此准确预测 *Kp* 指数对空间天气预警具有十分重要的意义.

Kp 指数发布时间通常有一定的延迟, 为适用实时应用需求, 研究开发了现报和短期预报模型. Takahashi 等[4] 利用 *Kp* 计算方法, 通过处理 9 个地磁台站的数据得到 *Kp* 估计值, 与 *Kp* 最终测量值相关系数为 0.94, 标准误差为 0.5, 具有很好的实时性. Luo 等[5] 基于太阳风–磁层耦合函数, 建立了 *Kp* 短期预报模型并给出相应的经验公式. 统计研究工作表明, *Kp* 指数与太阳风参数之间存在很好的相关性, 如太阳风速度和行星际磁场南向分量等. 随着人工智能的发展, 越来越多基于机器学习的 *Kp* 预报模型被建立, 这些模型的代表包括 Costello 模型[6]、Bala 模型[7,8] 以及 NARX 模型[9] 等, 主要利用上游太阳风数据和现报 *Kp* 值作为输入参数. Wing 等[10] 采用神经网络方法, 以 L1 点太阳风参数为输入建立了 APL 模型, 重点改进了对地磁活跃时段的预报. Tan 等[11] 使用长短时神经网络 (Long Short Term Memory, LSTM) 的方法, 将分类模型和回归模型相结合来预测 *Kp* 指数, 模型的均方根误差为 0.64, 相关系数 *C* 为 0.82. Wang 等[12] 使用统计方法针对 *Kp* 指数大于等于 4 的磁暴事件概率进行预测. Liu 等[13] 使用神经网络方法, 提前 1～3.5 h 预报 *Kp* 指数, 最高相关系数为 0.90, 均方根误差为 0.62. 目前短期 *Kp* 指数预报结果由 NOAA 的空间天气预报中心 (Space Weather Prediction Center, SWPC) 和中国科学院国家空间科学中心的空间环境预报中心 (Space Environment Prediction Center, SEPC) 发布, 每半小时更新一次.

Johnson 等[14] 研究发现地磁活动水平在太阳活动周的不同阶段差异明显. 在太阳活动高年, 受到行星际日冕物质抛射 (Interplanetary Coronal Mass Ejection, ICME)、太阳风高速流等行星际事件控制, 地磁活动较强; 在太阳活动低年, 地磁活动相对较弱, 主要受地球磁层动力学的影响. Le 等[15] 研究了地磁扰动事件在太阳活动周的分布特性, 发现在活动周下降阶段, 地磁扰动尤为显著, 这与低纬冕洞产生的高速太阳风有关. 太阳风到达地球后和磁层发生复杂的相互作用, 增加了 *Kp* 指数预报的难度. 太阳风作为磁层扰动的主要来源, 其能量注入的多少决定了磁层系统的总体行为. 当太阳风高速流冲击磁层顶时, 通过磁场重联和黏滞作用进入磁层, 引起磁层各结构的变化[16,17]. 大量研究表明, *Kp* 指数与磁层中一些现象相关联, 例如等离子体片边界和等离子体层顶位置等[18]. 磁层系统对太阳风能量注入的响应特征表现为时间上的滞后性和非线性, 时间滞后为 1 h 的太阳风能量和地磁指数的相关性最大, 较长的滞后时间意味着输入能量存在存储释放的过程[19,20]. 太阳风能量输入的不确定性和磁层响应的复杂性使得 *Kp* 预测的难度增加.

Kp 指数短期预报面临的问题主要来自以下几个

方面. 首先是数据发布的不确定性, 预测模型需要大量实时数据, 包括太阳风、行星际磁场参数、历史 Kp 指数等, 由于 Kp 指数在数据发布上存在一定时间延迟, 会影响预报的准确性. 其次是地磁场的复杂度, 地磁指数受到多种因素的影响, 包括太阳活动、地球磁层中电流系统等, 这些复杂系统使得难以确定 Kp 指数预报的直接影响因子. 然后是磁暴的非线性性, 即磁暴的发生是一个非线性过程, 因此需要发展更适应非线性关系的预报模型. 最后是太阳活动的随机性, 太阳活动是地球磁场的直接扰动源. 然而, CME 等剧烈太阳活动本身具有较强的随机性, 这也为 Kp 指数的预报带来不确定性.

本研究采用非线性的机器学习方法, 基于梯度提升回归和随机森林两种算法构建 Kp 指数预测模型. 通过参数选择的方法, 对比分析不同输入参数对预测模型的影响, 确定最佳输入参数组合. 在以往的研究中可以发现不同纬度区域的太阳黑子对近地太阳风速度的影响[21], 为探究太阳活动对地磁 Kp 指数的影响, 在预测模型中还将增加不同纬度区域的黑子数据作为输入参数.

1 数据处理与过程

研究中使用的太阳风和行星际参数来自 OMNI 数据库, 这些数据是由日地拉格朗日 1 点轨道上的航天器 (ISEE 3, Wind, ACE) 收集得到的, 并根据太阳风传播速度将数据时间由 L1 点调整至到太阳风首次接触到弓激波的时间[22]. OMNI 低分辨率 (Low Resolution OMNI, LRO) 数据集记录了从 1963 年至今每小时平均的太阳风磁场和等离子参数数据. LRO 数据集提供了太阳风速度 ($v_{\rm sw}$)、密度 ($N_{\rm sw}$)、压强 ($P_{\rm sw}$)、GSM 坐标下的磁场分量 (B_x, B_y, B_z)、行星际磁场强度 (B)、电场强度 (E) 等太阳风行星际参数. 研究中使用的数据时间范围为 1963 年 1 月至 2017 年 12 月的数据. 为保证数据的连续性和完整性, 将缺失和异常的数据点用去除后剩余数据的平均值代替.

地磁 Kp 指数数据来自 OMNI 数据库整合的德国地球科学研究中心的数据, 时间分辨率为 3 h. 标准的 Kp 指数值为 0, 0^+, 1^-, 1, 1^+, …, 9^-, 9, 但在 OMNI 数据集中, 相应的 Kp 被存为 0, 3, 7, 10, 13, …, 87, 90. 为满足预测模型的需求, 将数据除以 10, 得到范围在 0 到 9 的 Kp 数据.

太阳黑子数据来自格林尼治皇家天文台 (Royal Greenwich Observatory, RGO) 和美国国家海洋和大气管理局的在线数据. 该数据记录项目自 1874 年 5 月开始至 2017 年 12 月结束. 该太阳黑子数据集将每个黑子群编号, 记录每日各黑子群活动区的面积、日面位置等参数. 通过计算每日记录的黑子活动区编号, 得到每日群黑子数, 其数量大约是国际太阳黑子数的 1/10 [23]. 按照纬度将太阳黑子数据划分为五组, 包括全日面黑子数 $R_{\rm all}$、纬度小于 20°的太阳黑子数 R_{20}、纬度小于 15°的太阳黑子数 R_{15} 和纬度小于 10°的太阳黑子数 R_{10}. 考虑到地球公转轨道与太阳自旋赤道之间 7.2°的夹角, 增加了纬度小于 7.2°的太阳黑子数 $R_{7.2}$. 黑子面积数据也按照相同的方法划分, 即 $A_{\rm all}$, A_{20}, A_{15}, A_{10}, $A_{7.2}$. 其中黑子面积的单位为太阳半球的百万分之一 (μ Hem). 研究模型在太阳活动周不同阶段的表现时, 划分太阳活动周相位使用的黑子数据为月平滑黑子数, 数据来源于太阳影响数据分析中心 (Sunspot Number , SILSO).

2 模型数据与性能指标

研究使用的梯度提升算法 (Gradient Boosting regression, GBR) 是一种机器学习集成算法[24]. Gradient 代表一种优化方法, 用于最小化损失函数以找到函数的局部最小值, 通过迭代的方式, 从当前点开始, 沿着目标函数下降最快的方向 (梯度的反方向), 逐步接近最优解. 集成方法是指将几个弱学习器结合成一个强学习器的方法. 根据个体学习器的生成方式, 目前集成学习的方法大致分为两类, 一类为串联集成方法, 即个体学习器之间存在强依赖关系, 必须串行生成的序列化方法; 另一类为并联集成方法, 就是个体学习器之间不存在强依赖关系、可同时生成的并行化方法. 前者的代表是 Boosting, 后者的代表是 Bagging 和随机森林. GBR 算法属于 Boosting 集成方法, 通过循环训练预测器, 对前序学习器做出改正. 在训练过程中根据当前学习器损失函数的负梯度信息来训练新加入的学习器, 然后将训练好的学习器以累加的形式结合到现有模型中来降低预测误差. GBR 集成算法属于串联集成算法, 该方法的优势在于能够处理复杂的非线性关系, 用于回归预测模型时表现出色[25,26].

为了对比分析不同机器学习模型的性能, 引入了随机森林算法 (Random Forest, RF)[27]. 该算法是一种常用的机器学习方法, 属于集成学习中的并行 Bagging 的一种扩展方法, 随机森林方法是以决策树为基础学习器的集成算法, 与 GBR 算法不同点在于, 随机森林的集成方法属于并行集成方法, 在训练过程中引入随机属性, 通过组合多个决策树来进行分类和回归任务. 其结合了决策树的高灵活性和集成学习的优势, 具有较高的预测准确性.

GBR 算法主要适用于回归问题, 通过串联迭代过程中样本分布为训练样本赋予权重来降低偏差, 不断优化模型性能, 而随机森林方法在分类和回归问题中均有所应用. 两种方法在处理复杂数据关系、非线性关系时表现出色, 具有较高的预测准确性.

2.1 特征参数选择

在参数时间选取中, 以 OMNI 数据库提供的 3 h 段 Kp 值中的最后时刻 Kp 值作为预测值, 记作 Kp_t. 预测值 Kp_t 对应的时刻分别为 03:00 UT, 06:00 UT, …, 24:00 UT. 考虑到磁层系统对太阳风能量注入的响应时间, 选取 Kp_t 时刻前 3 h 内太阳风及行星际参数的极大值作为输入, 以 t 时刻为 06:00 UT 为例, 太阳风和行星际磁场参数对应时刻为 t–1 (05:00 UT), t–2 (04:00 UT), t–3 (03:00 UT), 并选取这三个时段中各参数的最大值. 之前的研究发现黑子活动区可以作为太阳风的起源[28–31], 因此黑子参数也被加入到输入参数以探究其对地磁指数的影响. 计算地磁 Kp 指数日平均值与黑子面积参数 $A_{\rm all}$ 在不同时间延迟下的相关系数, 结果表明延迟为 3 天时二者的相关系数最大, 因此选择 3 天前的太阳黑子参数作为输入变量.

使用了多个输入参数来构建模型, 包括太阳风速度 ($v_{\rm sw}$)、密度 ($N_{\rm sw}$)、压强 ($P_{\rm sw}$), GSM 坐标下的磁场分量 (B_x, B_y, B_z)、磁场强度 (B)、电场强度 (E)、不同纬度区域的黑子参数. 由于太阳风行星际磁场参数较多, 特别是不同纬度区域的黑子参数, 将所有 5 组纬度范围的黑子参数都输入模型不仅会增加模型的运算时长, 还会导致输入参数冗余. 因此, 引入参数评估指标来选择合适的输入参数. 使用 F 检验作为评估函数对输入数据进行排序筛选. 首先, 对目标函数 Kp 指数做多元线性回归, 然后对回归方程做显著性检验, 并计算 F 统计量. F 统计量的具体计算过程为先得到各输入参数与 Kp 指数的皮尔逊相关系数 r, 然后将 r 转换为 F 统计量, F 统计量的计算公式如下:

$$F = \frac{r^2}{(1-r^2)}(n-2).$$

其中, n 表示样本数量. F 统计量越大, 说明该输入参数对目标变量 Kp 指数的影响越大.

各参数评分结果用 F 统计量表示, 图 1 给出了以 Kp_t 指数为预测值的结果. 横坐标代表各输入变量与目标变量 Kp_t 的 F 统计量计算结果, 采用以 10 为底的对数表示. 参数评分越高表示该参数与 Kp_t 的相关性越强. 纵坐标按照得分从大到小排列, 对应各输入参数的名称. 参数选择结果表明行星际电场强度 E 与 Kp 值相关性最强. 其次的主要扰动源分别为太阳风速度 $v_{\rm sw}$, 行星际磁场强度 B, 太阳风压强 $P_{\rm sw}$. 在行星际磁场的三个方向分量中, B_z 分量的得分明显高于其他两个方向的磁场分量. 在之前的 Kp 预测模型中, 行星际磁场的 B_z 分量也是人工神经网络的重要输入参数[32]. 此外, 许多研究结果也证明南向行星际磁场强度对地磁扰动的影响较大[33,34]. 不同纬度区域的太阳黑子参数中, 全日面黑子面积 ($A_{\rm all}$) 与地磁参数评分最高. 因此, 在研究中选择黑子数据 $A_{\rm all}$ 作为模型输入参数.

2.2 数据集构建

为研究不同参数输入组合对 Kp 指数预测模型的影响, 基于参数选择的结果, 将建立三种输入数据集:

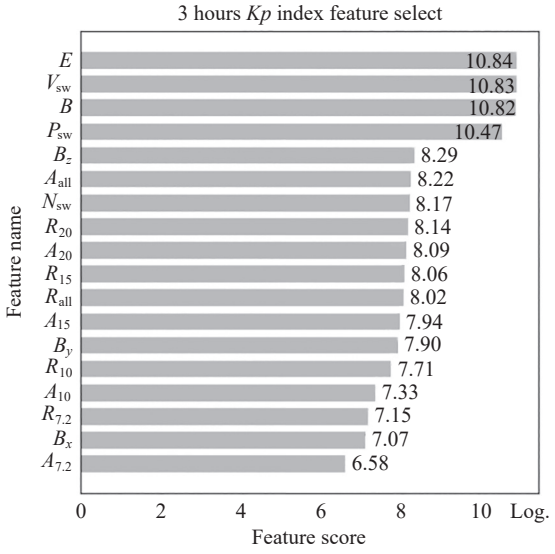


图 1 参数评分

Fig. 1 Parameter scoring results

数据集 1 输入参数为 t–1 到 t–3 时刻行星际电场强度 E, 太阳风速度 $v_{\rm sw}$, 行星际磁场 B, 太阳风压强 $P_{\rm sw}$ 和太阳风密度 $N_{\rm sw}$ 的极大值; 数据集 2 的输入是在数据集 1 的基础上加入 t–3 时刻 Kp 值, 记作 $Kp_{t\text{-}3}$; 数据集 3 在数据集 2 的基础上增加了 Kp_t 对应时间的三天前的黑子参数 $A_{\rm all}$.

太阳风和行星际参数记录从 1963 年到现在, 黑子数据集记录从 1874 年 5 月至 2017 年 12 月, 按照数据集重叠期, 最终确定数据集的时间范围从 1963 年 1 月 1 日至 2017 年 12 月 28 日. 研究中使用各数据集中前 80% 的数据作为训练集, 用于模型的训练; 后 20% 的数据作为测试集, 用以评估模型的性能. 然后对数据进行标准化处理, 将其转化为均值为 0、标准差为 1 的分布. 通过标准化, 可以使不同特征之间的数值范围一致, 避免某些特征对模型的影响过大. 此外, 标准化还可以加快模型的训练速度, 提高模型的收敛效果. 使用以下公式可以完成标准化过程:

$$X_{\rm norm}=\frac{x-\mu}{\sigma}.$$

其中, x 为原始数据, μ 为数据均值, σ 为数据标准差, $X_{\rm norm}$ 为标准化后的数据.

为防止过拟合情况的发生, 研究中 GBR 算法的参数设置如下: 学习率设置为 0.1, 决策树的个数为 100, 决策树的深度设置为 5, 损失函数为最小平方差. 随机森林方法中参数设置如下: 树的个数为 100, 节点分裂最小样本数为 2, 最小叶子节点样本数为 1.

2.3 性能指标

采用三种参数来评估模型的预测性能, 分别为线性相关系数

$$C=\frac{\sum_{i=1}^{n}\left(y-\overline{y}\right)\left(y_{\rm pre}-\overline{y}_{\rm pre}\right)}{\sqrt{\sum_{i=1}^{n}\left(y-\overline{y}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_{\rm pre}-\overline{y}_{\rm pre}\right)^2}},$$

预测效率

$$P=1-\frac{\sum_{i=1}^{n}\left(y-y_{\rm pre}\right)^2}{\sum_{i=1}^{n}\left(y-\overline{y}\right)^2},$$

以及均方根误差

$$E_{\rm rms}=\sqrt{\frac{\sum_{i=1}^{n}\left(y-y_{\rm pre}\right)^2}{n}}.$$

其中, y 表示 Kp 观测值, $y_{\rm pre}$ 表示 Kp 预测值, $\overline{y}$ 和 $\overline{y}_{\rm pre}$ 分别表示观测值和预测值的平均值.

3 预报结果

基于 GBR 算法和 RF 方法的 3 h 地磁 Kp 指数预报结果展示在表 1 中. 模型 1 到模型 3 分别对应不同的输入数据集. GBR1 模型的观测值与预测值的相关系数 C 为 0.89, 预测效率 P 为 0.79, 均方根误差 $E_{\rm rms}$ 为 0.58; RF1 模型的预测结果 C 为 0.87, P 为 0.75, $E_{\rm rms}$ 为 0.65. GBR2 模型的预测结果 C 为 0.90, P 为 0.81, $E_{\rm rms}$ 为 0.56; RF2 模型的 C 为 0.89, P 为 0.79, $E_{\rm rms}$ 为 0.59. 输入参数增加 $Kp_{t\text{-}3}$ 后两种模型的预测效果都有所提升. 增加了黑子参数后 GBR3 模型相对于 GBR2 模型预测结果保持不变, 但 RF3 模型预测结果 C 从 0.89 提升至 0.90, 有微小提升.

图 2 给出了 GBR3 和 RF3 模型的测试集预测结果对比. 图 2 中蓝色表示 Kp 实测值与预测值之间对比的散点图, 红色虚线表示 1:1 的对角线, 蓝色虚线表示散点的线性拟合结果. 虽然两种模型的预测 C 值

表 1 输入参数和 3 h 地磁 *Kp* 指数预测模型结果 (提前 1 h)

Table 1 Data set variables and comparison of the prediction model results of 3-hour *Kp* index (predict 1 h)

模型名称	输入参数	C	P	$E_{\rm rms}$
GBR1	$v_{\rm sw}$, $P_{\rm sw}$, $N_{\rm sw}$, B_z, B, E	0.89	0.79	0.58
RF1		0.87	0.75	0.65
GBR2	$v_{\rm sw}$, $P_{\rm sw}$, $N_{\rm sw}$, B_z, B, E, $Kp_{t\text{-}3}$	0.90	0.81	0.56
RF2		0.89	0.79	0.59
GBR3	$v_{\rm w}$, $P_{\rm sw}$, $N_{\rm sw}$, B_z, B, E, $Kp_{t\text{-}3}$, $A_{\rm all}$	0.90	0.81	0.56
RF3		0.90	0.80	0.58

注 $Kp_{t\text{-}3}$ 表示3 h 前Kp 指数.

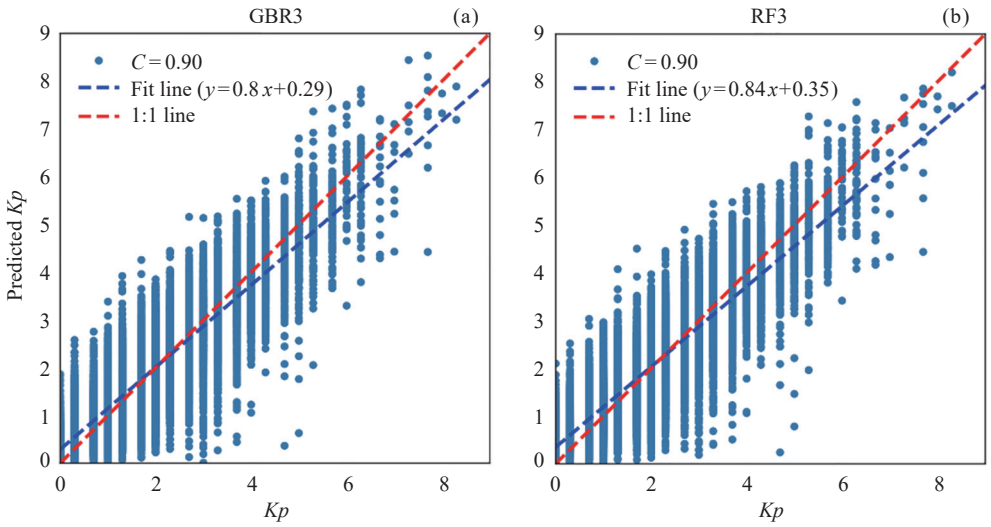


图 2　观测 *Kp* 值与 GBR3 和 RF3 模型的预测 *Kp* 值对比 (红色虚线表示对角线, 蓝色虚线表示线性拟合直线相关系数为 0.90)

Fig. 2　*Kp* observations and predictions, with observed *Kp* values on the horizontal axis and predicted values on the vertical axis (The red dashed line represents the diagonal line, and the blue dashed line represents the linear fit result of the scatter, with a correlation coefficient of 0.90)

均为 0.90, 但 GBR3 的预测值与观测值的线性拟合结果斜率大于 RF1 的拟合结果, 且 GBR3 模型的 $E_{\rm rms}$ 小于 RF1 模型. 测试结果表明, 太阳黑子参数对 RF 方法模型有一定提升, 但 GBR 算法在三种数据集的表现均优于 RF 方法. 在以往的研究中还有多种机器学习方法被用来预测 *Kp* 指数, 如 Tan 等[11] 使用长短时记忆神经网络算法预测结果 C 为 0.81, $E_{\rm rms}$ 为 0.64; Sexton 等[35] 使用递归神经网络算法, 预测结果 C 为 0.92, $E_{\rm rms}$ 最低为 1.32; Wang 等[36] 使用机器学习相似度算法, 预测结果 C 为 0.93, $E_{\rm rms}$ 为 0.79; Liu 等[13] 使用神经网络方法, 提前 1 h 到 3.5 h 预报 *Kp* 指数, 最高相关系数为 0.90, 均方根误差为 0.62. GBR 算法预测结果在均方根误差的表现要优于传统的 RF 方法, 是因为算法自身的优越性, 通过串联学习器在加权迭代过程中最小化损失函数, 因此, GBR3 模型的预测结果 C 值最高并且有较低的均方根误差.

4　讨论

与现有的 *Kp* 预测模型进行比较, 结果列于表 2. 表 2 中传统预报模型使用的算法为人工神经网络算法, 本文介绍的基于 GBR 算法的 *Kp* 指数预测模型在预测相关系数和均方根误差上表现均优于传统的预报模型. 相较于文献 [4] 的现报模型, 在时间上可以提前一小时进行 *Kp* 指数预报. 但模型输入参数数量上比传统预报模型多, 未来可以在参数输入量上进行优化.

研究中测试集的数据范围自 2006 年 12 月到 2017 年 12 月, 包含第 24 太阳周大部分活动阶段. 为了研究各模型在太阳活动周不同阶段的预测表现, 将活动周划分为三个阶段, 分别为上升期、极大期和下降期. 图 3 给出了 24 活动周的阶段划分, 红色部分表示活动周极大期, 绿色部分表示活动周的上升阶段, 蓝色部分表示下降阶段. 划分依据为: 月平滑黑子数超过极大年黑子数的 70% 为极大期, 极大期之前为上升阶段, 之后为下降阶段. 按照图 3 所示的时间分割点, 将数据划分为对应的三个测试集, 并计算了不同阶段下各种模型的预测结果. 选取线性相关系数 C 作为评价指标, 表 3 给出了 3 h 地磁 *Kp* 指数预测结果. 从表 3 可以看出, 在太阳活动周的不同阶段, 模型的预测性能也会有所变化. 三种输入数据集的预测性能在太阳活动周下降阶段均优于其在极大期和上升阶段. 研究表明, 地磁活动在同一太阳活动周内通常呈现双峰结构[33,37], 在活动周下降阶段的地磁扰动峰值多与低纬冕洞引起的太阳风高速流相关. 下降相冕洞高速流相对稳定且周期性较强, 因此以太阳风参

表 2 3 h 地磁 *Kp* 指数 GBR 预测模型与现有模型结果对比

Table 2 Comparison of the GBR prediction model of 3-hour *Kp* index with the existing models

文献	输入参数	提前时间/h	C	$E_{\rm rms}$
[7]	$v_{\rm sw}$, B_z, B	1	0.75	—
[32]	$v_{\rm sw}$, $N_{\rm sw}$, B_z	3	0.77	0.99
[9]	Boyle index, Kp	1	0.86	0.71
[4]	9个地磁台站Kp估计值	0	0.94	0.50
GBR3	$v_{\rm sw}$, $P_{\rm sw}$, $N_{\rm sw}$, B_z, B, E, Kp_{t-3}, $A_{\rm all}$	1	0.90	0.56

表 3 3 h 模型在太阳活动周不同相位的预测结果相关系数

Table 3 Correlation coefficient during different solar cycle phase of the 3-hour prediction model

	GBR1	RF1	GBR2	RF2	GBR3	RF3
上升阶段	0.89	0.88	0.89	0.88	0.89	0.88
极大期	0.88	0.87	0.89	0.88	0.89	0.88
下降阶段	0.89	0.89	0.90	0.89	0.90	0.90

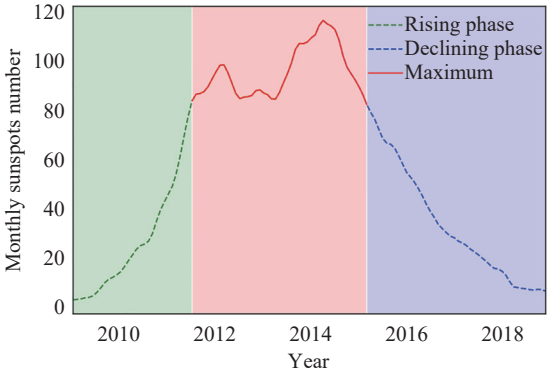


图 3 第 24 太阳活动周相位划分

Fig. 3 Phase division of the 24th solar activity cycle

数为主要输入的模型在活动周下降阶段预测 C 值更高.

地磁 Kp 指数预报主要针对强地磁扰动时预警作用, 将不同地磁扰动强度下的 Kp 指数预测结果进行对比. 表 4 列出了三种强度地磁扰动情况下, GBR3 模型对 Kp 指数的预测. 表 4 分别给出了不同强度扰动的事件总数, 预测模型预警成功的次数, 准确比例以及模型预测结果与真实值的线性相关系数. 表中展示了测试集数据在第 24 太阳活动内, 共发生磁暴事件 541 起, 模型准确预测 148 起, 预测准确率为 27%. 中等磁暴事件 ($5\leqslant Kp<6$)413 起, 预测准确率 26%; 强磁暴事件 ($6\leqslant Kp<7$)104 起, 预测准确率 35%; 超强磁暴事件 ($Kp\geqslant 7$)24 起, 预测准确 6 次, 准确率为 25%. 相比中等磁暴和超强磁暴, 模型对强磁暴 ($6\leqslant Kp<7$) 的预报准确度最高.

表 4 不同地磁扰动情况下 3 h 地磁 *Kp* 指数预测结果

Table 4 Comparison of the prediction results of 3-hour *Kp* index under different geomagnetic disturbance

	数据总数	预测次数	准确比例/(%)
$5\leqslant Kp<6$	413	106	26
$6\leqslant Kp<7$	104	36	35
$Kp\geqslant 7$	24	6	24

5 结论

为满足地磁指数的预报需求, 基于 GBR 算法构建了地磁 Kp 指数预报模型, 确立了以太阳风、行星际磁场参数 ($v_{\rm sw}$, $P_{\rm sw}$, $N_{\rm sw}$, B_z, B, E)、历史 Kp 值和太阳黑子数据 ($A_{\rm all}$) 为输入的参数组合. 本文介绍的 GBR 模型能够提前 1 h 对 Kp 指数进行预测, 线性相关系数 C 为 0.90, 均方根误差 $E_{\rm rms}$ 为 0.56, 预测效率 P 为 0.81.

地磁 Kp 指数 3 h 预测结果表明短期地磁活动主要受到太阳风、行星际磁场参数的影响[38–40]. 在之前的研究中发现不同纬度区域的黑子参数与近地太阳风速度存在相关性[21], 但这种影响主要表现在中低速

太阳风上, 其他情况下影响微小, 可忽略不计. 太阳黑子参数对短期 *Kp* 指数预报模型没有明显改善. 该结果表明太阳黑子活动区的面积与数量对短期地球磁场变化的影响较小, 地球磁场的主要扰动源还是来自剧烈太阳活动产生的高速流.

对比太阳活动周不同相位的 3 h 地磁 *Kp* 指数预报模型, 结果表明在太阳活动周下降阶段, 地磁 *Kp* 指数的预测结果均优于活动周上升阶段和活动周极大期. 有统计研究表明地磁活动峰值通常也出现在太阳活动周下降阶段[33], 此时的地磁扰动多与低纬冕洞产生的太阳风高速流相关. 高速太阳风驱动主导了磁层动力学, 因此在活动周下降阶段以太阳风为主要输入参数的模型预测效果更好. 对比不同地磁扰动强度下的 *Kp* 指数预测结果, 相较于中等磁暴和超强磁暴, 模型对强磁暴 ($6 \leqslant Kp < 7$) 的预报准确度最高.

通过这项研究, 推荐一种基于 GBR 算法的 *Kp* 指数预测模型, 可以实现提前 1 h 的 *Kp* 指数预报. 基于参数间的相关关系提高了 *Kp* 指数的预测准确性并降低了预测误差, 为 *Kp* 预测提供了新的方法. *Kp* 指数预测结果证明短期磁层扰动主要来自太阳风的能量输入[41–43]. 本文通过对比分析不同预测模型的结果, 不仅可以对恶劣空间天气提前预警, 还可以更好地理解地磁指数与太阳风输入能量的关系, 为太阳风–磁层耦合的研究工作提供更多方法和理论基础.

参考文献

[1] CAO J B, DUAN J T, DU A M, *et al*. Characteristics of middle- to low-latitude Pi2 excited by bursty bulk flows[J]. *Journal of Geophysical Research: Space Physics*, 2008, **113**(A7): A07S15

[2] CAO J B, YAN C X, DUNLOP M, *et al*. Geomagnetic signatures of current wedge produced by fast flows in a plasma sheet[J]. *Journal of Geophysical Research: Space Physics*, 2010, **115**(A8): A08205

[3] . BARTELS J. The standardized index, Ks, and the planetary index Kp[J]. IATME Bull 12b, IUGG Publ Office, Paris 1949 : 97–112; reprinted in: IAGA Bull 12i, IUGG Publ Office, Paris 1955: 88–101

[4] TAKAHASHI K, TOTH B A, OLSON J V. An automated procedure for near-real-time *Kp* estimates[J]. *Journal of Geophysical Research: Space Physics*, 2001, **106**(A10): 21017-21032

[5] LUO B, LIU S, GONG J. Two empirical models for short-term forecast of *Kp*[J]. *Space Weather*, 2017, **15**(3): 503-516

[6] COSTELLO K A. Moving the Rice MSFM into A Real-Time Forecast Mode Using Solar Wind Driven Forecast Modules[D]. Houston: Rice University, 1998

[7] BALA R, REIFF P. Improvements in short-term forecasting of geomagnetic activity[J]. *Space Weather*, 2012, **10**(6): S06001

[8] BALA R, REIFF P H, LANDIVAR J E. Real-time prediction of magnetospheric activity using the Boyle Index[J]. *Space Weather*, 2009, **7**(4): S04003

[9] AYALA SOLARES J R, WEI H L, BOYNTON R J, *et al*. Modeling and prediction of global magnetic disturbance in near-earth space: a case study for *Kp* index using NARX models[J]. *Space Weather*, 2016, **14**(10): 899-916

[10] WING S, JOHNSON J R, JEN J, *et al*. *Kp* forecast models[J]. *Journal of Geophysical Research: Space Physics*, 2005, **110**(A4): A04203

[11] TAN Y, HU Q H, WANG Z, *et al*. Geomagnetic index *Kp* forecasting with LSTM[J]. *Space Weather*, 2018, **16**(4): 406-416

[12] WANG J J, ZHONG Q Z, LIU S Q, *et al*. Statistical analysis and verification of 3-hourly geomagnetic activity probability predictions[J]. *Space Weather*, 2015, **13**(12): 831-852

[13] LIU Yang, LUO Bingxian, LIU Siqing, *et al*. *Kp* forecast models based on neural networks[J]. *Manned Spaceflight*, 2013, **19**(2): 70-80 (刘杨, 罗冰显, 刘四清, 等. 基于神经网络方法的 Kp 预报模型 [J]. 载人航天, 2013, **19**(2): 70-80)

[14] JOHNSON J R, WING S. A solar cycle dependence of nonlinearity in magnetospheric activity[J]. *Journal of Geophysical Research: Space Physics*, 2005, **110**(A4): A04211

[15] LE G M, CAI Z Y, WANG H N, *et al*. Solar cycle distribution of great geomagnetic storms[J]. *Astrophysics and Space Science*, 2012, **339**(1): 151-156

[16] PERREAULT P, AKASOFU S I. A study of geomagnetic storms[J]. *Geophysical Journal International*, 1978, **54**(2): 547-573

[17] KAN J R, LEE L C. Energy coupling function and solar wind-magnetosphere dynamo[J]. *Geophysical Research Letters*, 1979, **6**(7): 577-580

[18] THOMSEN M F. Why *Kp* is such a good measure of magnetospheric convection[J]. *Space Weather*, 2004, **2**(11): S11004

[19] BARGATZE L F, BAKER D N, MCPHERRON R L, *et al*. Magnetospheric impulse response for many levels of geomagnetic activity[J]. *Journal of Geophysical Research: Space Physics*, 1985, **90**(A7): 6387-6394

[20] XU Wenyao. Energy budget in the coupling processes of the solar wind, magnetosphere and ionosphere[J]. *Chinese Journal of Space Science*, 2011, **31**(1): 1-14 (徐文耀. 太阳风–磁层–电离层耦合过程中的能量收支 [J]. 空间科学学报, 2011, **31**(1): 1-14)

[21] JIAO Q R, LIU W L, ZHANG D J, *et al*. Relation between latitude-dependent sunspot data and near-earth solar wind speed[J]. *The Astrophysical Journal*, 2023, **958**(1): 70

[22] KING J H, PAPITASHVILI N E. Solar wind spatial scales in and comparisons of hourly wind and ACE plasma and magnetic field data[J]. *Journal of Geophysical Research: Space Physics*, 2005, **110**(A2): A02104

[23] WOLF W. Abstract of his latest Results[J]. *Monthly Notices of the Royal Astronomical Society*, 1861, **21**(3): 77-78

[24] BIAU G, CADRE B, ROUVIÈRE L. Accelerated gradient boosting[J]. *Machine Learning*, 2019, **108**(6): 971-992

[25] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. *The Annals of Statistics*, 2001, **29**(5): 1189-1232

[26] NATEKIN A, KNOLL A. Gradient boosting machines, a tutorial[J]. *Frontiers in Neurorobotics*, 2013, **7**: 21

[27] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, **45**(1): 5-32

[28] HE J S, MARSCH E, TU C Y, *et al*. Intermittent outflows at the edge of an active region: a possible source of the solar wind[J]. *Astronomy & Astrophysics*, 2010, **516**: A14

[29] KOJIMA M, FUJIKI K, OHMI T, *et al*. Low-speed solar wind from the vicinity of solar active regions[J]. *Journal of Geophysical Research: Space Physics*, 1999, **104**(A8): 16993-17003

[30] ZANGRILLI L, POLETTO G. Evolution of active region outflows throughout an active region lifetime[J]. *Astronomy & Astrophysics*, 2016, **594**: A40

[31] FU H, LI B, LI X, *et al*. Coronal sources and *in situ* properties of the solar winds sampled by ACE during 1999-2008[J]. *Solar Physics*, 2015, **290**(5): 1399-1415

[32] . BOBERG F, WINTOFT P, LUNDSTEDT H. Real time Kp predictions from solar wind data using neural networks [J]. *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial & Planetary Science*, 2000, **25**(4): 275-280

[33] BERGIN A, CHAPMAN S C, MOLONEY N R, *et al*. Variation of geomagnetic index empirical distribution and burst statistics across successive solar cycles[J]. *Journal of Geophysical Research: Space Physics*, 2022, **127**(1): e2021JA029986

[34] ZHANG Dianjun, LIU Wenlong, DU Junfeng, *et al*. Simulation study on the Impulsive Electric Field induced by an interplanetary shock[J]. *Chinese Journal of Geophysics*, 2023, **66**(11): 4429-4436 (张典钧, 刘文龙, 杜俊峰, 等. 行星际激波导致内磁层脉冲电场的模拟研究 [J]. 地球物理学报, 2023, **66**(11): 4429-4436)

[35] SEXTON E S, NYKYRI K, MA X Y. *Kp* forecasting with a recurrent neural network[J]. *Journal of Space Weather and Space Climate*, 2019, **9**: A19

[36] WANG Zisiyu, SHI Liqin, LIU Siqing, *et al*. *Kp* index prediction based on similarity algorithm of machine learning [J]. *Chinese Journal of Space Science*, 2022, **42**(2): 199-205. DOI: 10.11728/cjss2022.02.210316030 (王子思禹, 师立勤, 刘四清, 等. 基于机器学习相似度算法的 *Kp* 指数预报 [J]. 空间科学学报, 2022, **42**(2): 199-205. DOI: 10.11728/cjss2022.02.210316030)

[37] GONZALEZ W D, GONZALEZ A L C, TSURUTANI B T. Dual-peak solar cycle distribution of intense geomagnetic storms[J]. *Planetary and Space Science*, 1990, **38**(2): 181-187

[38] CROOKER N U, GRINGAUZ K I. On the low correlation between long-term averages of solar wind speed and geomagnetic activity after 1976[J]. *Journal of Geophysical Research: Space Physics*, 1993, **98**(A1): 59-62

[39] PAPITASHVILI V O, PAPITASHVILI N E, KING J H. Solar cycle effects in planetary geomagnetic activity: analysis of 36-year long OMNI dataset[J]. *Geophysical Research Letters*, 2000, **27**(17): 2797-2800

[40] GARRETT H B, DESSLER A J, HILL T W. Influence of solar wind variability on geomagnetic activity[J]. *Journal of Geophysical Research*, 1974, **79**(31): 4603-4610

[41] LIU W, SARRIS T E, LI X, *et al*. Solar wind influence on Pc4 and Pc5 ULF wave activity in the inner magnetosphere[J]. *Journal of Geophysical Research: Space Physics*, 2010, **115**(A12): A12201

[42] MAYAUD P N. Derivation, Meaning, and Use of Geomagnetic Indices[M]. Washington: American Geophysical Union, 1980

[43] ZHANG D J, LIU W L, DU J F, *et al*. Response of electric field in terrestrial magnetosphere to interplanetary shock[J]. *The Astrophysical Journal*, 2022, **938**(1): 70

作者简介



焦琦融 女, 1993 年 10 月出生于内蒙古呼和浩特, 现为北京航空航天大学博士研究生, 主要研究方向为太阳物理过程、空间天气预报等.
E-mail: JiaoQirong@buaa.edu.cn



张典钧（通信作者） 男, 1994 年 10 月出生于北京市西城区, 现为北京航空航天大学讲师, 主要研究方向为空间物理、内磁层动力学等.
E-mail: diandian@buaa.edu.cn