0254-6124/2021/41(5)-800–08 *Chin. J. Space Sci.* 空间科学学报

GUO Xingliang, CUI Ruifei, ZHU Yaguang, TIAN Chao, JIANG Jianmin, YUE Fulu. A missing values imputation method for radiation data from Beidou MEO satellite sensors (in Chinese). *Chin. J. Space Sci.*, 2021, **41**(5): 800-807. DOI:10.11728/cjss2021.05.800

# 北斗 MEO 卫星辐射剂量探测数据缺失值处理方法*

郭兴亮 崔瑞飞 朱亚光 田 超
姜健民 岳甫璐

(宇航动力学国家重点实验室 西安 710600)

**摘 要** 针对北斗 MEO 卫星辐射剂量探测数据出现连续性缺失的问题, 开展缺失值处理方法研究. 提出一种叠加正弦波的线性样条回归方法, 即引入样条函数, 对各数据连续缺失的时间段进行分段处理, 每段样条采用叠加正弦波的线性方程填充缺失值. 结果表明: 利用该方法处理缺失值, 每段样条中填充曲线与探测曲线在增长趋势、周期性波动等方面具有较高的一致性; 相比前向插值法和线性插值法, 其填充值与真实值误差更小, 关联性更高. 该方法较好地解决了数据连续缺失的问题, 形成了完整性好、准确性高的北斗 MEO 卫星辐射剂量数据集, 为后续数据的发布、建模和可视化展示等奠定了基础.

**关键词** MEO 卫星, 辐射剂量, 缺失值, 线性样条, 回归分析

**中图分类号** V 520

# A Missing Values Imputation Method for Radiation Data from Beidou MEO Satellite Sensors

GUO Xingliang CUI Ruifei ZHU Yaguang TIAN Chao
JIANG Jianmin YUE Fulu

(*State Key Laboratory of Astronautic Dynamics, Xi'an* 710600)

**Abstract** Radiation dose refers to the sum of ionization energy deposited by various incident high-energy particles in unit mass material. The radiation dose data detected by Beidou MEO satellites suffers from the problem of missing values that tend to occur continuously and there is little difference in the radiation dose detection values in the three directions in the satellite cabin. Firstly, the detector 3 data with strong regularity is selected as the research object. In order to fill the missing values, a sine-adjusted linear spline regression method is proposed, in which each continuous data-missing time period is individually processed in a spline and the missing values are filled with the sine-adjusted

* 国防科技创新特区项目资助 (1916321TS00101206)
2020-06-11 收到原稿, 2021-04-09 收到修定稿
E-mail: crf_dqq@126.com

linear equation. The experiment shows that with our approach the filling curve is highly consistent with the true curve in terms of growth trend and periodic fluctuation. And our method performs significantly better with respect to errors and correlation coefficient between true and filled values, than alternative methods such as forward interpolation and linear interpolation. This paper selects the data of different time periods to analyze the filling effect, and draws the above conclusions. The proposed method properly solves the problem of continuous data-missing, leading to a data set with good completeness and high accuracy, which lays a foundation for the subsequent tasks like data release, modeling and visualization.

**Key words** MEO satellite, Radiation dose, Missing values, Linear spline, Regression analysis

## 0 引言

中地球轨道 (Medium Earth Orbit, MEO) 是指高度在 2000 km 以上, 地球静止轨道 (约 3600 km) 以下的轨道区域. 该轨道多应用于卫星导航业务, 例如美国 GPS、俄罗斯 GLONASS、欧洲 Galileo、中国北斗导航系统等 [1]. 由于 MEO 轨道的卫星每天约 4 次穿越地球外辐射带中心, 受辐射带内空间带电粒子影响的风险较大. 因此, 空间带电粒子与 MEO 卫星的相互作用是影响其在轨寿命及安全性的重要因素 [2−5].

空间带电粒子与构成卫星的元器件和材料发生撞击时, 可通过电离作用将部分甚至全部能量传递给元器件和材料, 使其出现性能衰退、参数漂移等现象, 导致电离总剂量效应. 电离总剂量效应具有长期累积的特点, 是造成卫星元器件性能衰退和失效的主要原因之一 [6]. 实际应用中, 通常采用辐射剂量来定量描述电离总剂量效应, 即各种入射高能粒子在单位质量物质材料中沉积的电离能量的总和. 开展辐射剂量探测, 对研究星内剂量分布、分析卫星元器件异常原因、合理选择元器件与材料、进行抗辐照设计等, 具有重要的理论意义和应用价值.

基于上述原因, 中国发射的北斗 MEO 卫星搭载了可用于探测辐射剂量的空间环境载荷. 然而, 在实际应用过程中, 辐射剂量探测数据存在部分时间段内数据缺失的问题, 其原因可能是载荷工作异常、通信信道拥挤、数据下传不及时或原始探测数据解遥测失败等. 探测数据的缺失不仅影响卫星舱内辐射剂量的监测, 更是会对后续航天器设计和器件筛选、在轨卫星状况评估等任务造成严重影响. 因此, 研究数据质量控制方法, 实现缺失值的填充, 是北斗 MEO 卫星辐射剂量数据应用过程中必须要解决的问题.

数值分析中常用的缺失值填充方法有前向插值、后向插值、线性插值等, 这些方法通常用于处理个别数据随机性丢失的问题 [7]. 然而, 北斗 MEO 卫星辐射剂量数据缺失具有明显的连续性特点, 即如果在某一时刻数据开始缺失, 那么数据大概率会在接下来一段时间内连续缺失. 这一特点给传统的数据插值方法造成极大影响, 尤其是当数据连续缺失时间较长时, 传统方法几乎失效. 为了解决这一问题, 本文根据北斗 MEO 卫星辐射剂量数据变化规律及缺失特点, 提出了一种叠加正弦波的线性样条回归方法, 研究使用该方法实现缺失值的填充, 并与前向插值及线性插值方法进行对比分析, 形成完整性好、准确性高的北斗 MEO 卫星辐射剂量数据集.

本文根据北斗 MEO 卫星辐射剂量数据变化规律及缺失特点, 提出了一种叠加正弦波的线性样条回归方法, 利用该方法填充数据缺失值, 并与前向插值、线性插值进行对比分析.

## 1 探测数据

分析采用北斗某 MEO 卫星辐射剂量探测数据. 该卫星一共搭载三台辐射剂量探测仪 (以下简称探测仪 1、探测仪 2、探测仪 3), 可探测获取卫星舱内三个方向的辐射剂量, 数据分辨率为 20 s, 单位为拉德 (rad).

图 1 给出的是 2019 年全年卫星舱内三个方向辐射剂量变化曲线. 比较三台探测仪的探测结果可以发现, 三个方向辐射剂量探测值差别不大, 尤其是 9 月之前差别很小, 9 月之后差别稍有增大但是数据增长趋势仍然较为一致. 分析数据缺失特点可以发现, 三个方向辐射剂量数据缺失的时刻基本相同, 且数据缺失呈现出明显的连续性特征, 即如果在某一时刻数

据开始缺失, 那么数据大概率会在接下来一段时间连续缺失. 例如, 2019 年数据大致存在 1 月 28 日至 2 月 10 日 (粗略估计, 故只精确到日) 等 13 个连续缺失的时间段, 具体情况列于表 1 (由于篇幅限制, 缺失时间段没有完全列出).

为了更清晰地认识数据变化规律, 图 2 给出了 2019 年前 2 个月 (1 月 1 日至 2 月 28 日) 卫星舱内三个方向辐射剂量变化曲线. 图 2 对原始数据进行了平滑, 平滑方法为对探测数据按取中位数的方法进行重新采样, 采样频率为 1 h. 从图 2 可以看出, 1 月 28 日至 2 月 10 日数据出现连续性缺失, 在此期间只有部分较短时间段内存在探测数据. 分析数据变化规律可以发现, 数据呈现高频率的波动状态, 探测仪 3 的数据波动规律性较强, 波动周期、振幅分别稳定在 12 h, 100 rad 左右.

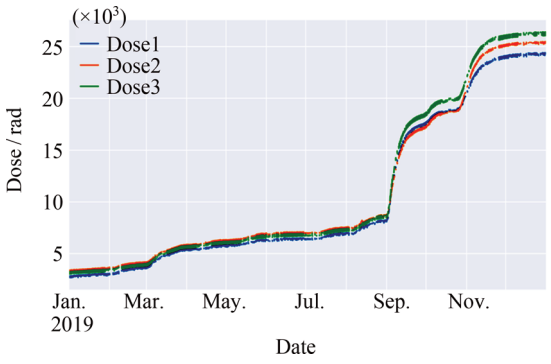



图 1 2019 年卫星舱内三个方向辐射剂量变化曲线

Fig. 1 Radiation dose curve in three directions in satellite cabin in 2019

需要说明的是, 探测仪探测的随时间累积的辐射剂量理论上应是单调递增的, 但是实际探测结果却呈现出上下涨落的规律性波动. 这是因为探测仪选用的传感器属于温度敏感器件, 当卫星环境温度变化时, 探测仪输出也发生变化. 探测仪采用 RADFET 场效应管传感器, 其基本工作原理是场效应管受辐照后, 产生电子空穴对, 空穴被场效应管绝缘层 (二氧化硅) 俘获, 并且在栅极氧化层产生新的电荷界面, 引起栅极电压发生变化. 栅极电压的变化与传感器接受的辐射剂量之间存在对应关系, 利用这个对应关系, 通过监测栅极电压变化从而获得器件受照剂量大小. 当温度升高时, 被俘获的空穴数目减少, 从而引起栅

**表 1 2019 年数据连续缺失时间段情况统计**

**Table 1 Summary of continuous data-missing time period in 2019**

| 序号 | 1 | 2 | 3 | ··· | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|
| 缺失时间段 | 1 月 28 日至 2 月 10 日 | 3 月 4 日至 3 月 17 日 | 4 月 7 日至 4 月 21 日 | ··· | 10 月 30 日至 11 月 3 日 | 11 月 11 日至 11 月 26 日 | 12 月 10 日至 12 月 28 日 |

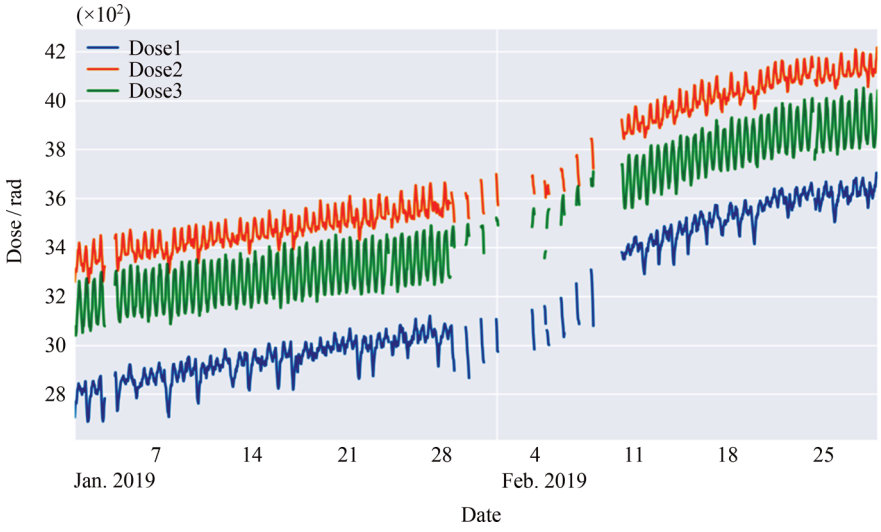



图 2 2019 年 1 月 1 日至 2 月 28 日卫星舱内三个方向辐射剂量变化曲线

Fig. 2 Radiation dose curve in three directions in satellite cabin from 1 January to 28 February in 2019

极电压变化减小, 这就是 RADFET 场效应管的温度效应 [8,9]. RADFET 场效应管的温度效应与其制造工艺、灵敏层厚度、面积等有关. 图 3 所示为 2019 年 1 月 1−11 日卫星舱内三个方向辐射剂量探测仪传感器输出电压与仪器温度的变化情况. 从图 3 可以看出, 总剂量输出与探测仪温度存在比较明显的负相关性: 温度每升高 1°C, 探测仪输出电压大约降低数毫伏. 为了降低温度的影响, 在实际应用过程中, 可以根据地面温度试验数据对星上探测数据进行相应的温度修正, 最大限度降低温度对探测仪输出数据的影响.

本文研究对象是辐射剂量探测仪的原始探测数据, 探索如何高效地实现缺失值的填充, 形成完整性好、准确性高的原始探测数据集. 在原始探测数据集上实现缺失值的填充, 可以最大限度地保留探测数据最原始的特征, 便于分析探测数据周期性波动的原因及规律等. 同时该项工作也是后续进行地面温度效应修正的基础: 现有地面温度效应修正模型依托原始探测数据作为输入. 本文首先选择规律性较强的探测仪 3 数据作为研究对象, 探测仪 1 和 2 的数据将作为后续的研究内容, 本文暂不讨论.

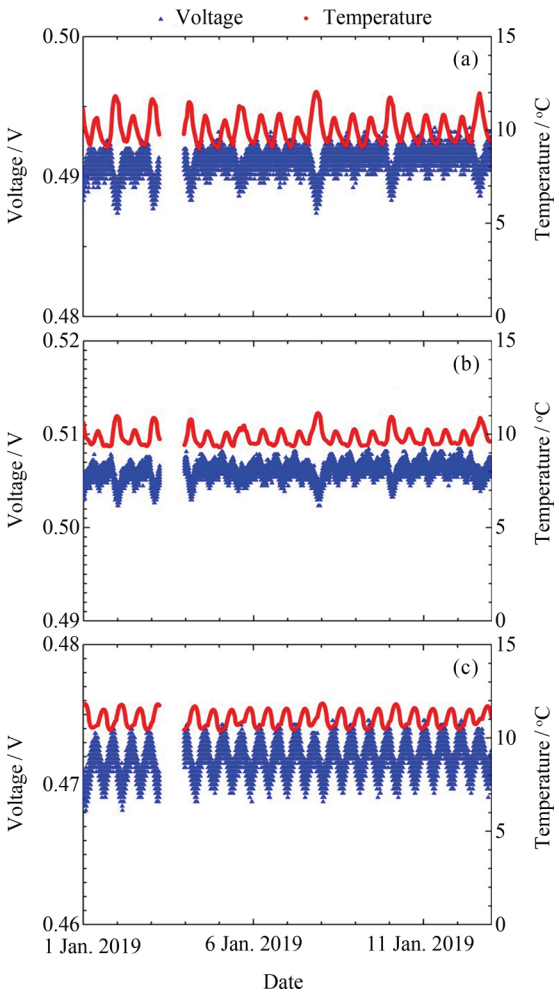



图 3 卫星舱内三个方向辐射剂量探测仪传感器输出电压与仪器温度的变化

Fig. 3 Temperature and output voltage of three detector sensors in three directions in satellite cabin

## 2 缺失值处理方法

上述分析认为, 辐射剂量数据呈现高频率波动状态. 规律性较好的探测仪 3 数据, 其波动周期、振幅较为稳定. 如果将辐射剂量 $S$ 看作随时间 $t$ 变化的函数, 其规律可近似为一个基本函数上叠加一个正弦波, 表达式可以写为 $S(t)=f(t)+c\sin(\omega t+d)$. 因此, 辐射剂量数据缺失值填充可用回归分析的方法实现, 步骤如下.

**步骤 1** 确定基本函数表达式 $f(t)$.

**步骤 2** 在 $f(t)$ 上叠加正弦波 $c\sin(\omega t+d)$.

**步骤 3** 估计 $S(t)$ 中的未知参数.

**步骤 4** 用 $S(t)$ 填充缺失值.

从图 1 反映的 2019 年数据全年变化趋势可以看出, 基本函数 $f(t)$ 呈现高度非线性特征, 想要寻找一个非线性函数 $f(t)$ 反映全年数据整体变化趋势是非常困难的. 但是如果只关注某一小段时间数据, 例如图 2 中的 1 月 7−27 日, 其整体变化趋势基本可视为线性的, 即 $f(t)$ 为线性函数. 基于以上分析, 本文引入数值分析中常用的样条函数 (spline), 提出一种叠加正弦波的线性样条回归方法 (sine-adjusted linear spline regression), 对各数据连续缺失的时间段进行分段处理, 每段样条采用叠加正弦波的线性方程填充缺失值.

### 2.1 线性样条

在数值分析中样条是一种特殊函数, 由多项式分段定义 [10,11]. 例如, 给定 $n+1$ 个点 $t_0,t_1,t_2,\cdots,t_n$ 并且满足 $t_0<t_1<t_2<\cdots<t_n$, 这些点称为节点; 同时假定一个整数 $k\geqslant 0$, 具有节点 $t_0,t_1,t_2,\cdots,t_n$ 的一个 $k$ 次样条函数是指满足下列条件的函数 $f(t)$:

(1) 在每一个区间 $[t_i,t_{i+1}]$ 上, $f(t)$ 是一个次数 $\leqslant k$ 的多项式;

(2) 在 $[t_0,t_n]$ 上 $f(t)$ 有 $k-1$ 阶导数.

常见的样条有一次样条、$B$-样条、三次样条等.

最简单的样条是一次样条 ($k=1$), 即线性样条, 其表达式可以记为

$$f(t)=a_it+b_i,\ t\in[t_i,t_{i+1}],\ \forall I\in\{0,1,2,\cdots,n-1\}.$$

其中 $a_i$, $b_i$ 为常数, $t_i$, $t_{i+1}$ 为节点. 例如, 具有节点 0, 10, 20, 30, 40 的一个线性样条表达式可以记为

$$f(t)=\begin{cases}t, & t\in[0,10];\\ 0.5t+5, & t\in[10,20];\\ 2t-25, & t\in[20,30];\\ -t+95, & t\in[30,40].\end{cases}$$

该线性样条如图 4 所示.

通过上述对数据的分析可以发现, 如果只关注某一小段时间的辐射剂量数据, 其整体变化趋势可视为线性的. 因此本文使用线性样条函数作为基本函数 $f(t)$.

### 2.2 正弦波叠加

正弦型函数是实践中广泛应用的一类重要函数, 表达式为 $y=c\sin(\omega t+d)$, $\omega=2\pi/T$, 其中 $c$, $\omega$, $T$, $d$ 为常数, 物理意义分别代表振幅、频率、周期、初相. 探测仪 3 数据呈现高频率波动状态, 波动振幅、周期较为稳定. 因此, 可在基本函数 $f(t)$ 上叠加一个正弦型函数作为辐射剂量函数 $S(t)$ 的表达式, 即

$$S(t)=a_it+b_i+c_i\sin\left(\frac{2\pi}{T_i}t+d_i\right),$$

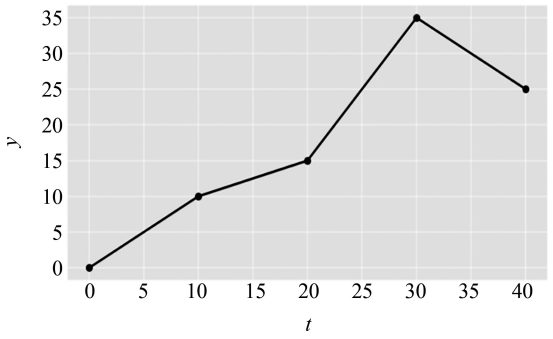



图 4 线性样条

Fig. 4 Linear spline function

$$t\in[t_i,t_{i+1}],\quad \forall i\in\{0,1,2,\cdots,n-1\}.$$

其中, $a_i$, $b_i$, $c_i$, $T_i$, $d_i$ 为待确定的未知参数, $t_i$, $t_{i+1}$ 为节点.

### 2.3 参数估计

#### 2.3.1 节点选取

关于样条节点的选取, 常用的方法有在数据变化剧烈的位置设置节点、在数据变化复杂的地方多设置节点、平均分配节点等 [12−14]. 本文重点要处理 13 个数据连续缺失的时间段 (见表 1), 因此样条节点取在数据连续缺失时间段前后. 同时为了保证具有充足的样本作为训练集, 样条节点取为数据连续缺失起始 (终止) 时间点前 (后) 1∼2 天. 以处理 1 月 28 日至 2 月 10 日这一数据连续缺失的时间段为例, 样条节点取为 1 月 27 日 00:00:00 LT 和 2 月 13 日 00:00:00 LT. 本文一共选取了 28 个样条节点, 在存在数据连续性缺失的 13 段样条中, 采用叠加正弦波的线性方程填充缺失值. 具体的样条节点选取情况列于表 2, 受篇幅限制, 部分样条节点没有列出.

#### 2.3.2 其他参数

样条节点选取后, 在存在数据连续性缺失的 13 段样条中, 采用叠加正弦波的线性方程填充缺失值. 在每段样条 $[t_i,t_{i+1}]$ 内, 需要估计的未知参数有 $a_i$, $b_i$, $c_i$, $T_i$, $d_i$. 本文采用最小二乘法估计未知参数, 通过最小化误差平方和寻找数据的最佳函数匹配.

假定某段样条 $[t_i,t_{i+1}]$ 内共有 $m$ 个样本点, 点 $t=t_j$ 处探测值为 $S_j$, 填充值为 $\hat{S}_j=a_it_j+b_i+c_i\sin\left(\frac{2\pi}{T_i}t_j+d_i\right)$, 那么目标就是找到一组 $\hat{a}_i$, $\hat{b}_i$, $\hat{c}_i$, $\hat{T}_i$, $\hat{d}_i$, 使得损失函数的值最小, 即解决如下优化问题:

$$\hat{a}_i,\cdots,\hat{d}_i=\underset{\hat{a}_i,\cdots,\hat{d}_i}{\arg\min}\frac{1}{m}\sum_{j=1}^{m}(\hat{S}_j-S_j)^2.$$

Python Scipy 数值计算库的 optimize 模块提供了众多的拟合与优化算法, 本文使用 optimize 模块提供的 curve_fit() 函数进行最小二乘法参数估计.

**表 2 样条节点选取情况统计**

**Table 2 Summary of selected spline knots**

| 节点 | $t_0$ | $t_2$ | $t_3$ | $t_4$ | $\cdots$ | $t_{24}$ | $t_{25}$ | $t_{26}$ | $t_{27}$ |
|---|---|---|---|---|---|---|---|---|---|
| 时刻 | 1 月 | 1 月 | 2 月 | 3 月 | $\cdots$ | 11 月 | 12 月 | 12 月 | 12 月 |
| (00:00:00 LT) | 1 日 | 27 日 | 13 日 | 2 日 | $\cdots$ | 25 日 | 10 日 | 28 日 | 31 日 |

# 3 缺失值处理结果

## 3.1 缺失值填充结果

以 2019 年 1 月 28 日至 2 月 10 日、4 月 8 日至 4 月 20 日这 2 个数据连续缺失的时间段为例, 按照取缺失时间段前后 1~2 天数据作为训练集的原则, 样条节点分别取 1 月 27 日 00:00:00 LT 和 2 月 13 日 00:00:00 LT, 4 月 6 日 00:00:00 LT 和 4 月 23 日 00:00:00 LT. 在两段样条内, 分别使用最小二乘法估计未知参数 $a_i$, $b_i$, $c_i$, $T_i$, $d_i$, 然后使用叠加正弦波的线性方程填充缺失值. 计算结果分别为

$$\begin{cases} a_1 = 5.16 \times 10^{-3}, \\ b_1 = 3.33 \times 10^{3}, \\ c_1 = -88.86, \\ T_1 = 2.32 \times 10^{3}, \\ d_1 = 2.51, \end{cases} \quad \begin{cases} a_2 = 4.39 \times 10^{-3}, \\ b_2 = 5.69 \times 10^{3}, \\ c_2 = -1.13 \times 10^{2}, \\ T_2 = 2.32 \times 10^{3}, \\ d_2 = -1.93. \end{cases}$$

图 5 给出了 2 段样条内缺失值填充结果, 其中绿色曲线表示探测数据变化曲线 (简称探测曲线), 红色曲线表示使用叠加正弦波的线性样条回归方法学习得到的曲线 (简称填充曲线). 从图中可以看出, 填充曲线能够非常好地拟合真实数据的变化规律; 填充曲线与探测曲线在增长趋势、周期性波动等方面具有较高的一致性. 因此, 使用填充曲线可以完成探测数据缺失值的填充.

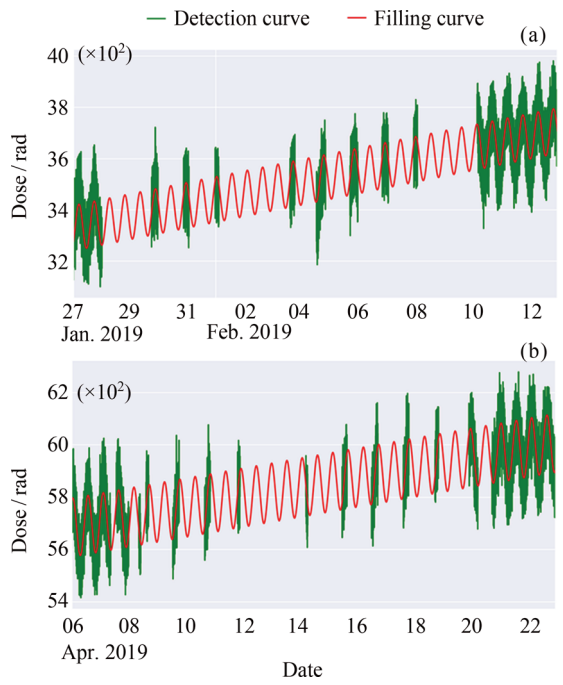



图 5 2019 年 1 月 27 日至 2 月 13 日和 4 月 6 日至 4 月 23 日辐射剂量变化曲线

Fig. 5 Radiation dose curve from 27 January to 13 February and from 6 April to 23 April in 2019

## 3.2 填充效果分析

为了分析缺失值填充效果, 验证上述方法的有效性和精度, 本节将使用前向插值和线性插值填充结果与本文所提方法进行对比. 前向插值方法是指用数据缺失前最后一个点的值代表缺失时间段的值, 线性插值方法是指用连接缺失前最后一个点和缺失后第一个点的直线段上的值代表缺失时间段的值. 对比实验的具体步骤如下: (1) 选择一段不含缺失值的完整数据片段, 在其中剔除部分时间段数据, 用来模拟数据连续缺失的现象; (2) 被剔除的数据作为测试集, 用于评估学习方法的性能; (3) 未被剔除的数据作为训练集, 基于训练集分别使用上述三种方法学习得到填充曲线; (4) 将填充值与真实值进行比较, 评估缺失值填充效果.

选取 1 月 5 日 00:00:00 LT 至 1 月 20 日 00:00:00 LT 这一不含缺失值的完整数据段作为实验数据, 剔除 1 月 11 日 00:00:00 LT 至 1 月 15 日 00:00:00 LT 4 天数据. 被剔除的 4 天数据作为测试集, 未被剔除的数据作为训练集, 基于训练集分别使用上述三种方法学习得到填充曲线.

图 6 给出了三种方法的填充曲线、探测曲线及测试集数据曲线 (简称测试集曲线) 对比情况. 从图 6 可以看出, 叠加正弦波线性样条回归方法的填充曲线为一条周期性波动曲线, 前向插值、线性插值方法的填充曲线则为一条直线. 对比填充曲线与测试集曲线可以发现, 叠加正弦波的线性样条回归方法的填充曲线体现了测试集曲线的周期性波动特征, 且波动周期、振幅与测试集曲线具有较高的一致性; 前向插值、线性插值方法的填充曲线为一条直线, 不能体现出测试集曲线的周期性波动.

为了定量描述三种方法填充值与真实值的偏差, 本文计算了填充值与真实值的误差 (Error, $E$)、相对误差 (Relative Error, $E_{\mathrm{R}}$)、均方根误差 (Root Mean Squared Error, $E_{\mathrm{RMS}}$) 及相关系数 (Correlation Coefficient, $R$), 表达式分别如下:

$$E = y_{\mathrm{fill}} - y_{\mathrm{true}}, \quad E_{\mathrm{R}} = \frac{|y_{\mathrm{fill}} - y_{\mathrm{true}}|}{y_{\mathrm{true}}},$$

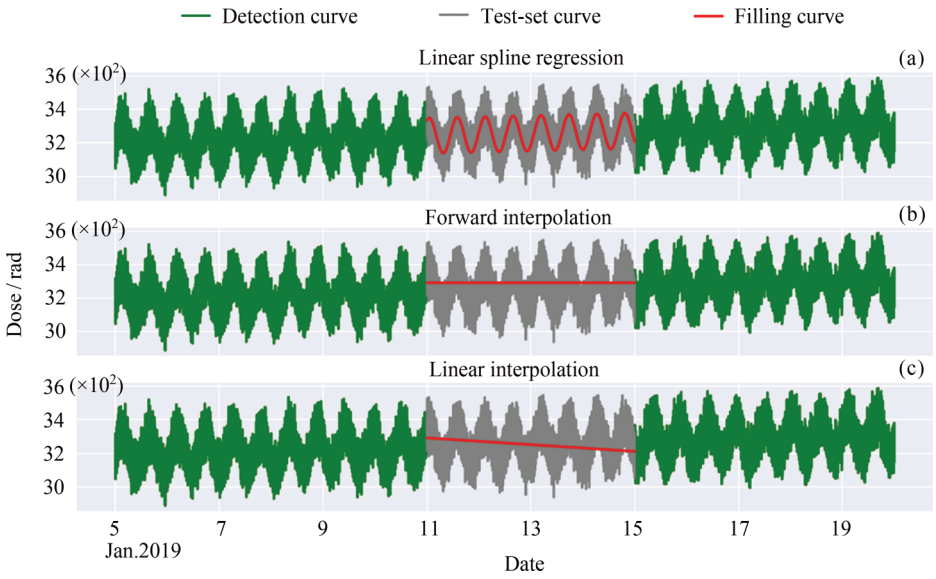



图 6 叠加正弦波的线性样条回归、前向插值、线性插值三种方法缺失值填充结果

Fig. 6 Result of sine-adjusted linear spline regression method, forward interpolation and linear interpolation

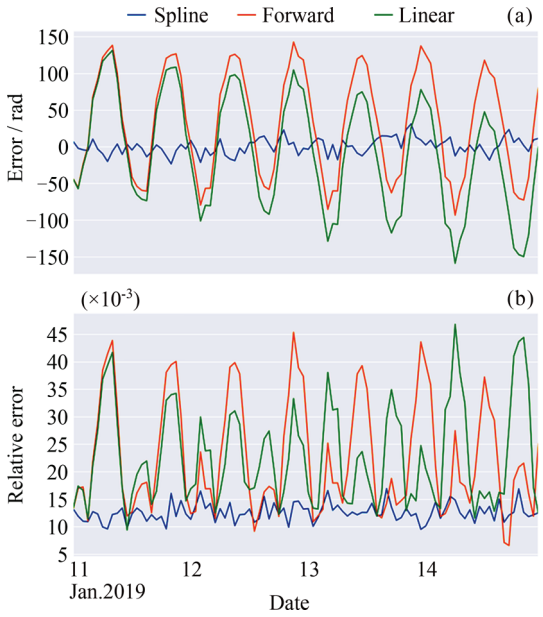



图 7 三种方法填充值与真实值误差 (a) 及相对误差 (b) 的对比情况

Fig. 7 Comparison of the error (a) and relative error (b) between filled values and real values of three methods

$$E_{\mathrm{RMS}}=\sqrt{\frac{\sum_{i=1}^{n}(y_{i\text{-fill}}-y_{i\text{-true}})^2}{n}},$$

$$R=\frac{\mathrm{Cov}(y_{\mathrm{fill}},y_{\mathrm{true}})}{\sqrt{D(y_{\mathrm{fill}})D(y_{\mathrm{true}})}}.$$

**表 3 三种方法填充值与真实值均方根误差 ($E_{\mathrm{RMS}}$)、相对误差平均值 ($E_{\mathrm{RM}}$) 和相关系数 ($R$)**

**Table 3 Root mean square error ($E_{\mathrm{RMS}}$), mean relative error ($E_{\mathrm{RM}}$) and correlation coefficient ($R$) between filled values and real values of three methods**

| 填充方法 | 叠加正弦波的线性样条回归 | 前向插值 | 线性插值 |
|---|---|---|---|
| $E_{\mathrm{RMS}}$ | 59.34 | 99.65 | 99.72 |
| $E_{\mathrm{RM}}$ | 0.015 | 0.025 | 0.025 |
| $R$ | 0.78 | 0 | −0.07 |

其中, $y_{\mathrm{true}}$, $y_{\mathrm{fill}}$ 分别为真实值和填充值, $\mathrm{Cov}(\cdot)$, $D(\cdot)$ 分别代表协方差和方差.

图 7 为三种方法填充值与真实值误差及相对误差对比情况, 从图 7 可以看出, 叠加正弦波的线性样条回归方法填充值与真实值误差在 (−25, 25) 区间内微小波动, 相对误差也基本集中在 (0.010, 0.015) 区间内; 前向插值、线性插值方法的填充值与真实值误差波动振幅较大, 最大偏差可达 150 rad 左右, 相对误差也明显大于叠加正弦波的线性样条回归方法. 从均方根误差、相对误差平均值和相关系数计算结果 (见表 3) 也可以看出, 叠加正弦波的线性样条回归方法均方根误差、相对误差平均值明显小于前向插值和线性插值方法, 相关系数则明显大于其他二种填充方案.

综上所述, 利用叠加正弦波的线性样条回归方法处理探测数据缺失值, 填充曲线与探测曲线在增长趋势、周期性波动等方面具有较高的一致性, 能够反映探测数据的变化规律. 同时相较于前向插值、线性插值, 其填充值与真实值误差更小, 关联性更高. 因此, 利用叠加正弦波的线性样条回归方法填充探测数据缺失值, 总体上优于前向插值和线性插值, 能够形成完整性好、准确性高的北斗 MEO 卫星辐射剂量数据集. 需要说明的是, 本文选取不同时间段的数据进行填充效果分析, 均得出了上述结论, 由于篇幅限制不再一一描述.

## 4 结论与讨论

针对北斗 MEO 卫星辐射剂量探测数据出现连续性缺失的问题, 提出一种叠加正弦波的线性样条回归方法, 利用该方法处理缺失值, 并与前向插值法和线性插值法填充结果进行了比较, 主要结论如下.

(1) 卫星舱内三个方向辐射剂量探测值差别不大, 均存在数据缺失问题, 且三个方向数据缺失的时间基本一致, 数据缺失呈现出明显的连续性特征.

(2) 针对数据变化规律与缺失特点, 可用叠加正弦波的线性样条回归方法处理缺失值, 即引入样条函数, 对各数据连续缺失的时间段进行分段处理, 每段样条采用叠加正弦波的线性方程填充缺失值.

(3) 利用叠加正弦波的线性样条回归方法处理缺失值, 每段样条内填充曲线与探测曲线在增长趋势、周期性波动等方面具有较高的一致性; 同时相较于前向插值和线性插值, 其填充值与真实值误差更小, 关联性更高. 该方法的运用形成了完整性好、准确性高的北斗 MEO 卫星辐射剂量数据集.

需要说明的是, 本文只选取了规律性较强的探测仪 3 数据进行缺失值处理试验. 理论上探测仪 1 和探测仪 2 数据可采用相似的方法处理, 但是仅仅叠加正弦波曲线或许不能刻画其中的周期性波动规律, 一种可行的方案是利用傅里叶特征取代正弦函数以获得更一般的周期性特征. 如何高效处理探测仪 1 和探测仪 2 数据的缺失值, 是下一步研究中要解决的问题.

## 参考文献


[1] YANG Yuanxi. Progress, contribution and challenges of Compass/Beidou satellite navigation system [J]. *Acta Geodaet. Cartograph. Sin.*, 2010, **39**(1): 1-6

[2] YANG Xiaochao, ZHU Guangwu, FU Yang, *et al.* Analysis of high energy electron environment properties on the middle earth orbit [J]. *Chin. J. Space Sci.*, 2015, **35**(5): 581-587 (杨晓超, 朱光武, 符养, 等. 中地球轨道高能电子辐射环境特性分析 [J]. 空间科学学报, 2015, **35**(5): 581-587)

[3] WANG Chunqin, ZHANG Xianguo, WANG Shijin, *et al.* Comparison of high energy particles observation results between FY-3A satellite and NOAA series satellites [J]. *Chin. J. Space Sci.*, 2010, **30**(1): 49-54 (王春琴, 张贤国, 王世金, 等. FY-3A 卫星与 NOAA 系列卫星高能带电粒子实测结果的比较 [J]. 空间科学学报, 2010, **30**(1): 51-56)

[4] BAKER D N, BLAKE J B, CALLIS L B, *et al.* Relativistic electron acceleration and decay time scales in the inner and outer radiation belts: SAMPEX [J]. *Geophys. Res. Lett.*, 1994, **21**(3): 409-412

[5] WANG Zhenhe, SUN Baosheng, LIU Shengpeng. Influence of high energy electrons on geosynchronous orbit satellites [J]. *J. Spacecraft TT&C Technol.*, 2017, **36**(3): 201-206

[6] HU Jianhang, FENG Ying, HAN Jianwei, *et al.* Simulation and validation of composite shielding for total ionizing dose [J]. *Chin. J. Space Sci.*, 2014, **34**(2): 180-185 (胡鉴航, 冯颖, 韩建伟, 等. 电离总剂量复合屏蔽模拟仿真及验证试验 [J]. 空间科学学报, 2014, **34**(2): 180-185)

[7] QIAO Zhufeng, TIAN Fengzhan, HUANG Houkuan, *et al.* A comparison study of missing value datasets processing methods [J]. *J. Comput. Res. Develop.*, 2006, **43**: 171-175

[8] BHAT B R, UPADHYAYA N, KULKARNI R. Total radiation dose at geostationary orbit [J]. *IEEE Trans. Nucl. Sci.*, 2005, **52**(2): 530-534

[9] SHIN Y H, MIN K W, RHEE J G, *et al.* Analysis of anomalous TDE data on-board the KITSAT-1 [J]. *IEEE Trans. Nucl. Sci.*, 2002, **46**(6): 1586-1594

[10] SCHUMAKER, LARRY L. Spline models for observational data (Grace Wahba) [J]. *SIAM Rev.*, 1991, **33**(3): 502

[11] JOHNSON S A, STEDINGER J R, SHOEMAKER C A, *et al.* Numerical solution of continuous-state dynamic programs using linear and spline interpolation [J]. *Operat. Res.*, 1993, **41**(3): 484-500

[12] YI Taihe, LIU Bing, YI Dongyun. Method for optimization of spline nodes and its application in trajectory representation [J]. *J. Spacecraft TT&C Technol.*, 2012, **031**(005): 54-58

[13] QIN Kaihuai, GUAN Youjiang. Two algorithms for inserting knots into B-spline curves [J]. *Chin. J. Comput.*, 1997, **6**: 556-561

[14] ZHU Jübo, HE Mingke, RAN Chengqi, *et al.* Data smoothing and filtering technique by free node spline function [J]. *Acta Electron. Sin.*, 2004, **032**(001): 135-138