

基于时间序列的宇宙电磁波接收模型

曹泊豪

(湖北经济学院, 湖北省武汉市, 430205; 3443192224@qq.com)

摘 要: 其一, 首先对数据进行有效性检验, 在确定所有数据均有效且无缺失值后利用题中所给的功率谱密度函数建立功率谱分析模型。其中进行了基于 (Root Mean Square, RMS) 的红噪声设定以及频谱分析等过程, 最后为了使数据拟合度高于 95%, 利用最小二乘法调整参数 P_0 , f_c 和 q , 进行未知参数的确定。通过模型可视化展现, 可以表明模型的拟合性优良。其二, 通过附件相邻 MJD 相减, 可知此数据并非时间步长相等的时序。为了使后续模型精确化以及统一时间单位, 本问进行了拉格朗日插值, 生成时间步长为 1 天的均匀时间序列。对于 PT-TT 的预测, 首先想到用 ARIMA 模型进行预测。最后经过检验, 发现短短一年后 MJD 的 95% 置信区间长达 1s, 难以进行长期预测, 于是本问考虑了用更适合长期预测的 LSTM 预测模型。通过交叉验证及 6000 余次迭代, 最后在长达三年的预测集和真实值的 RMSE 低至 0.002, 模型表现良好。其三, 首先基于 Saatanainen 模型与 Herring 模型建立一个初步的大气折射估计模型, 之后对模型中的 P (近地面大气压)、 e (水汽压) 等大气参数在现实允许范围下进行鲁棒性调整, 使模型在极端天气下具有足够的稳健性。在时延 ($0 < \Delta\tau < 7.69$) 的限制下进行蒙特卡洛模拟的两步模拟: 一、对 $[0, 4]$ km 海拔高度的电磁波接受点进行随机取值, 评估海拔高度与频率 f 的关系。二、为了尽可能真实地模拟宇宙噪声, 对频率 f 为 $[20, 250]$ GHz (包括宇宙中最普遍的 160 GHz 微波背景辐射) 随机模拟取值, 评估海拔不同的电磁波接收点接收效果的优劣性。其四, 本题在上一个模型基础上引进了大气曲率的影响, 并在此基础上进行大气分层建模。本题首先构建了有关地球曲率的大气折射指数 n 的海拔模型, 通过路径积分模型进行大气分层建模, 生成 10 组均匀的在仰角在 10 度以下的精确延时数据。为了将第三问模型进行矫正, 之后在理论值的基础上进行映射参数 a , b , c 的最小二乘估计, 计算分别为 $e+5 \times (0.06866 \quad -3.3163)$ 。最后将矫正前延时与矫正后延时进行比较与分析, 准确回答原始模型小仰角误差出现的原因, 证明了模型的正确性。

关键词: 时间序列; 宇宙电磁波接收模型

引言

问题背景

脉冲星是快速旋转的中子星, 具有稳定的自转, 被誉为“宇宙灯塔”。它们在深空探测器的导航和时间标准的维持中起着至关重要的作用。将脉冲星时间应用于原子钟计时, 有望提高本地原子钟的稳定性和可靠性, 是计时技术未来发展的长期方向。然而, 脉冲星计时噪声是影响脉冲星计时精度和稳定性的关键挑战之一。这里参考了建模所需的数据(一段时间接收信号时间和接收信号值)。引进了模型有关定义: 红噪声 (Red Noise)[1]: 一种在低频段具有更大功率的噪声, 常出现在脉冲星计时噪声中。均方根 (RMS, Root Mean Square): 一种衡量数据集变化幅度的统计量。到达时间 (TOA, Time of Arrival): 探测到脉冲星脉冲的时间。修正儒略日 (MJD, Modified Julian Date)[2]: 天文学中使用的日期计数系统, 是表示特定日期的一种方法。PT-TT (Pulsar Timing - Terrestrial Time)[3]是脉冲星计时和地球时间 (Terrestrial Time, TT) 之间的转换过程与关系。在脉冲星研究中, PT-TT 主要用于将脉冲星发射的信号时间与地球收到的时间进行精确对比。

问题重述

脉冲星计时噪声模拟: 使用脉冲星计时噪声数据, 建立一个功能模型来模拟该噪声, 目标是模型拟合度达到 95% 以上。通常假设红噪声的强度与 RMS 值成正比, 但并不相等。

脉冲星计时噪声预测：对脉冲星计时噪声的短期（几天到一个月）和长期（几个月到几年）趋势进行预测。预测结果的验证数据在附件 1 中提供。

电波折射时间延迟建模：建立一个模型来模拟射电观测频率高于 20 GHz 时的电波折射时间延迟，确保天顶延迟小于等于 7.69 ns。

大气时间延迟建模：建立一个模型来模拟低仰角观测（10 度或更小）时的大气时间延迟，以提高到达时间(TOA)的精度。需要提供模型以及对模型考虑因素和可实现目标的描述。

1 模型假设

本文基于以下合理假设：

假设平流层折射指数的变化与对流层的统一于一个模型

假设气候在认知范围内变化，不会产生灾难性的剧变

假设同一层大气密度变化均匀

2 符号说明

符号	意义
P_0	红噪声强度
f	电磁波频率
P	近地面大气压
e	水气压
$\Delta\tau_i$	第 i 层大气下的折射时延
z	天顶角
ϵ	仰角
h_i	第 i 层大气的高度
T	气温
n_i	第 i 层大气的折射指数

2.1 具体建模步骤

Step1: 功率谱密度模型的确定

根据题目公式，功率谱密度模型为式（1）：

$$P(f) = \frac{P_0}{(1+(f/f_c)^q)} \quad (1)$$

N 为信号长度， $|X(f)|$ 为 FFT 结果的复数模值。

Step2: 基于 RMS 的红噪声设定

根据题目描述，RMS 与红噪声强度 P_0 呈正比，因此我们引入一个常量 k ，有（式 2）：

$$P_0 = kRMS \quad (2)$$

通过不同的 RMS 值，可以对 P_0 进行估算。

Step3: 频谱分析过程。

首先设定频率范围 $[a, b] \ni f$ 。

通过功率谱密度（PSD）可以生成不同频率下的噪声幅度。

使用傅里叶逆变换，将频域噪声转换到时域，生成一个模拟的噪声时间序列

通过连续傅里叶变换将此时间序列转换成频域，见式（3）：

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{-j2\pi fT} dt \quad (3)$$

其中：

$X(f)$ 为频域中的信号， $x(t)$ 为时域中的信号， T 为时间序列的总长度。

Step4: 通过最小二乘拟合优化参数

为了使数据拟合度高于 95%，这里使用最小二乘法调整参数 P_0 ， f_c 和 q 。

使残差平方和最小，见式（4）：

$$SSR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

其中：

y_i 是实际的观测值。

$\hat{y}_i$ 是模型预测的值。

Step5: 建模可视化展示：

红噪声时间序列见图 1：

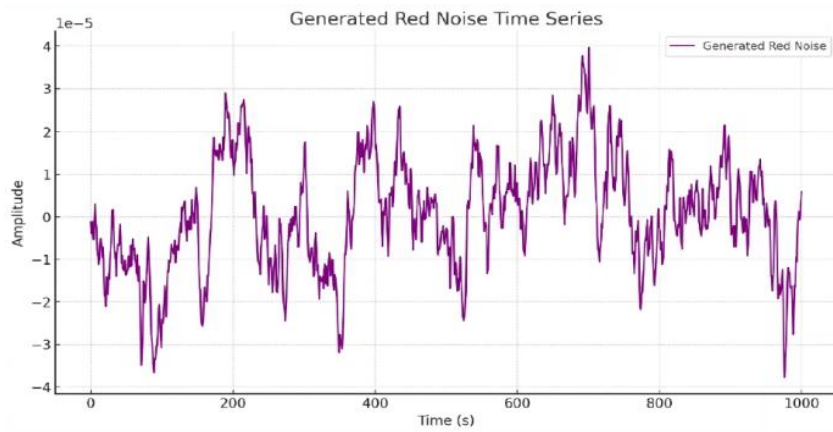


图 1 红噪声时间序列

所生成红噪声和拟合模型的功率谱密度函数与时频的关系，见图 2：

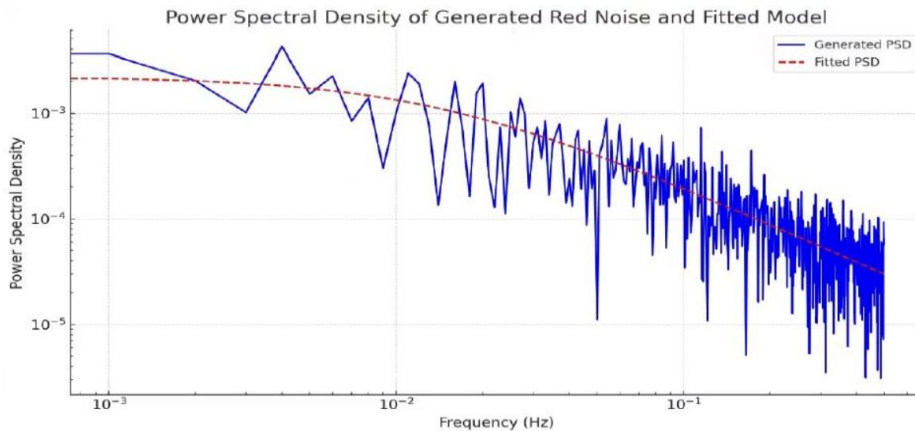


图 2 功率谱密度函数与时频的关系

2.2 模型特点分析：

1) **频率域信息：**功率谱分析可以将信号从时间域转换到频率域，使得分析周期性信号的频谱特征变得更加直观。

2) **噪声分析**: 在噪声背景下, 功率谱能够有效提取有用信号的频率成分, 帮助识别信号中的噪声特性。

3) **信号特征提取**: 通过功率谱, 能够提取出信号的主要频率成分和特征, 并对其进行分析与比较。

4) **最小二乘法的使用**: 可以随意控制模型拟合优度, 在高要求环境下比PLANA模型更加便利且精确, 提供了更高的灵活度。

5.总的来说, 功率谱分析模型较傅里叶曲线拟合更为复杂, 但提供了更高的精确度以及灵活性, 在实际应用的效果中, 功率谱分析模型往往比傅里叶曲线拟合模型更为优秀。

3 问题二模型的建立与求解

3.1 整体建模思路分析

首先将引用数据改成 MJD 并不是步长一致的时间序列, 为了更好地建立时间序列模型, 这里运用拉格朗日插值法, 将 MJD 转换成以天为步长的时间序列。通过观察 MJD 与 PT-TT 的关系图, 猜测若建立 ARIMA 时间序列预测模型, 应能取得良好的预测效果。现建立以下流程, 见图 3:

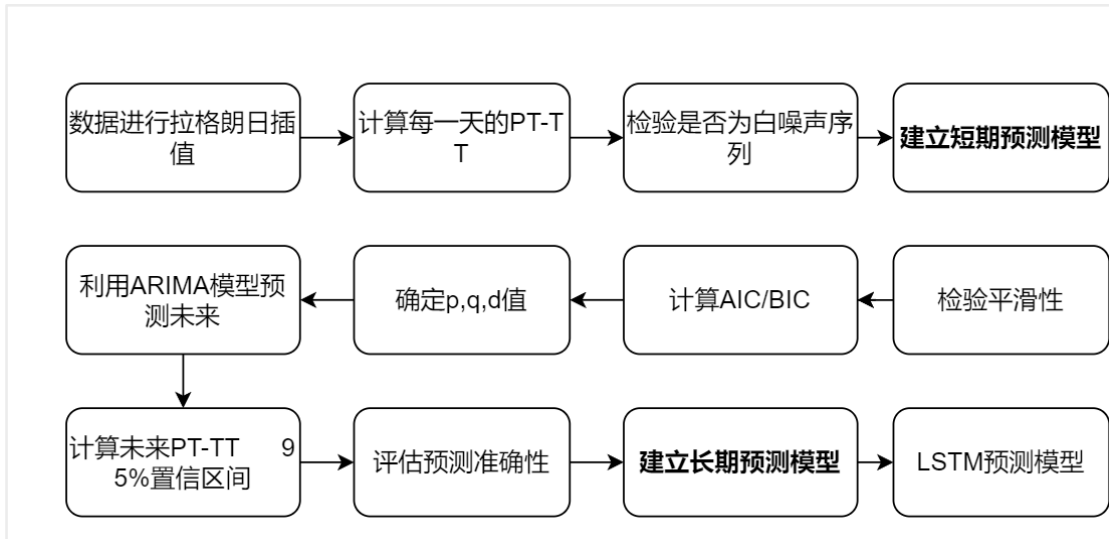


图3 问题二模型流程图

3.2 对数据进行拉格朗日插值

Step1: 确定数据点:

选择需要插值的数据点, 通常表示为 $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$, 其中 x_i 是自变量, y_i 是因变量。在本题中, x_i 为 MJD, y_i 为 PT-TT。

Step2:构造拉格朗日基多项式:

拉格朗日插值多项式 $P(x)$ 可以表示为式(5):

$$P(x) = \sum_{i=0}^n y_i L_i(x) \quad (5)$$

其中, $L_i(x)$ 是第 i 个拉格朗日基多项式, 定义为式(6):

$$L_i(x) = \prod_{\substack{0 \leq j \leq n \\ j \neq i}} \frac{x - x_j}{x_i - x_j} \quad (6)$$

这意味着需要为每个数据点 i 计算一个基多项式 $L_i(x)$ 。

Step3: 计算每个基多项式

计算每个 $L_i(x)$ 的值, 确定 x 的值, 为每个 i 和所有其他 j 的数据点计算乘积。

Step4: 组装插值多项式

将所有的 $y_i L_i(x)$ 加在一起，得到最终的插值多项式

$$P(x) = y_0 L_0(x) + y_1 L_1(x) + \cdots + y_n L_n(x) \quad (7)$$

Step5: 进行插值

一旦函数 $P(x)$ 完成，就可以用这个多项式来插值任何 x 值，特别是那些在原始数据点之间的值。插值结果如图 4:

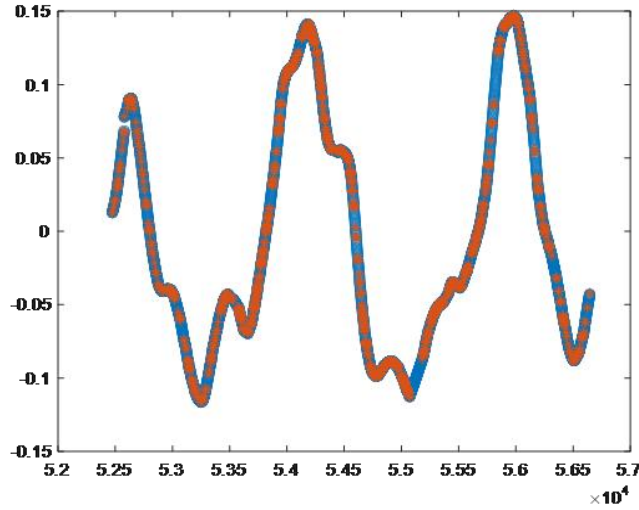


图 4 插值前后数据可视化

3.3 短预测建模（ARIMA）

3.3.1 白噪声检验

Step1: 数据准备

收集数据：获取待检验的时间序列数据，可以是模型的残差或其他时间序列数据。

可视化数据：绘制时间序列图和自相关图（ACF）以观察其特征。

Step2: 自相关函数 (ACF) 计算

计算 ACF：自相关函数用来衡量一个时间序列自身在不同滞后期的相关性。公式：自相关函数的计算公式为式（8）：

$$\rho_k = \frac{\sum_{t=k+1}^n (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2} \quad (8)$$

其中：

ρ_k 是滞后 k 的自相关系数， X_t 是时间序列的值， $\bar{X}$ 是时间序列的均值。 n 是序列的长度。

Step3: Ljung-Box 检验

原假设：时间序列为白噪声。

检验统计量，见式（9）：

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k} \quad (9)$$

其中：

Q 是检验统计量， n 是样本大小， m 是滞后期的数量， $\hat{\rho}_k$ 是滞后 k 的样本自相关系数， p 值：根据 Q 值，进行卡方分布检验，得到 p 值。

Step4: 结果解释

判断原假设:

如果 p 值 > 显著性水平 (如 0.05), 不能拒绝原假设, 认为序列是白噪声。

如果 p 值 ≤ 显著性水平, 拒绝原假设, 认为序列不是白噪声。

Step5: 结果呈现

用 MATLAB 可以算出 Ljung-Box 测试值:

Ljung-Box Test Results, 见表 1:

表 1 Ljung-Box 测试值

h	1
p-value	0
Q-statistic	41749.8181
Critical Value	18.307

即拒绝原假设, 数据不是白噪声,

其 ACF 以及时间序列图如图 5 所示:

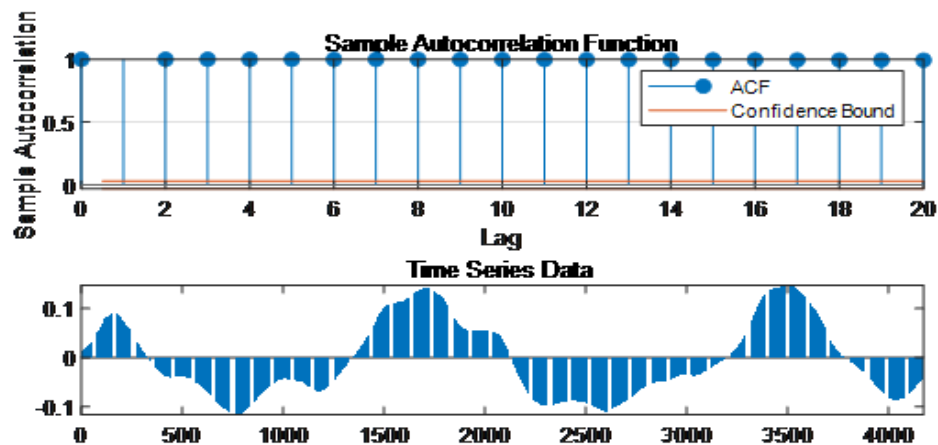


图 5 ACF 以及时间序列图

3.4 建立 ARIMA 模型

Step1: Augmented Dickey-Fuller (ADF) 平稳性检验

1) 模型形式: ADF 检验的基本回归形式为式 (10):

$$\Delta y_t = \alpha + \beta_t + \phi y_{t-1} + \sum_{i=1}^p \gamma_i \Delta y_{t-i} + \epsilon_t \quad (10)$$

其中: y_t 是时间序列观测值, y_t 在本题中为 PT-TT, $\Delta y_t = y_t - y_{t-1}$ 是当前值与前一个值的差, α 是常数项, β 是时间趋势的系数, ϕ 是单位根的参数, ϵ_t 是白噪声误差项。

2) 假设检验:

原假设 H_0 : 存在单位根 (时间序列是非平稳的)。

备择假设 H_1 : 不存在单位根 (时间序列是平稳的)。

3) 经过计算:

是否拒绝原假设 (序列非平稳): 0; p 值: 0.4650。表明序列为平稳序列。

Step2: 通过 **BIC/AIC 准则** 可以准确评估 ARIMA p,d,q 值的优劣处:

1) BIC/AIC 准则的基本原理:

Akaike Information Criterion (AIC), 见式 (11):

$$AIC = -2\log(L) + 2k \quad (11)$$

其中: L 是模型的最大似然估计 (MLE), k 是模型中的参数个数, Bayesian Information Criterion (BIC), 见式 (12):

$$BIC = -2\log(L) + k\log(n) \quad (12)$$

其中: n 是样本量, BIC/AIC 越低, 其对应的 ARIMA 模型越优。

2) 下面代入足够多的 p, q 值, 进行 AIC 的计算, 见表 2:

表 2 不同 p, q 下的 ARIMA 模型组合

ARIMA (3,0,1)	ARIMA (2,0,1)	ARIMA (0,0,1)	ARIMA (1,0,2)
ARIMA (3,0,2)	ARIMA (2,0,3)	ARIMA (0,0,3)	ARIMA (1,0,1)
ARIMA (3,0,3)	ARIMA (2,0,2)	ARIMA (0,0,2)	ARIMA (1,0,3)
ARIMA (3,0,4)	ARIMA (2,0,4)	ARIMA (0,0,4)	ARIMA (1,0,4)

其对应 AIC 的热力图如图 6:

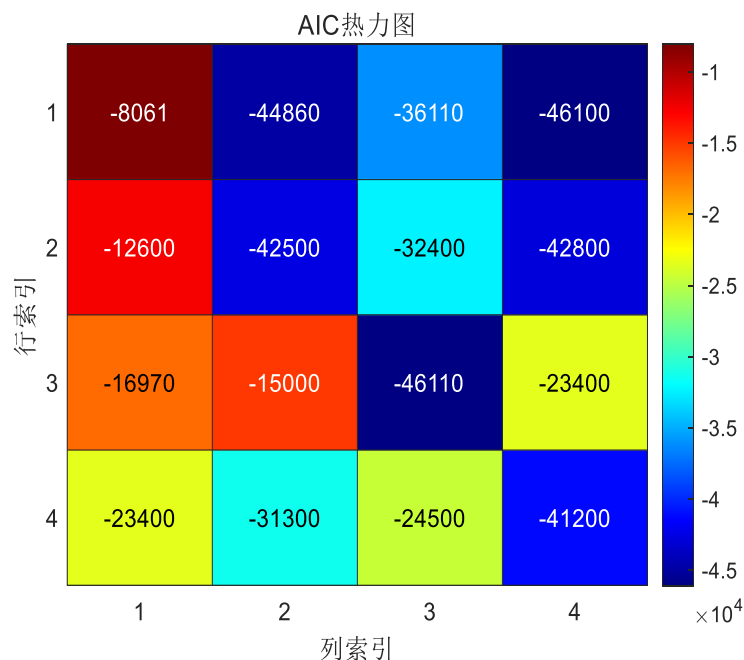


图 6 AIC 的热力图

由图可得, 最低 AIC 为 -46110, 此时对应的模型为 ARIMA (0,0,2), 即 MA (2)。

Step3: ARIMA 模型建立的过程

1) ARIMA 模型由以下部分组成:

AR (AutoRegressive): 自回归部分, 表示当前值与其过去值的关系。

(Integrated): 积分部分, 表示序列的差分次数。ARIMA(0, d, q) 模型中的 'd' 是为了使序列平稳而进行的差分次数。

MA (Moving Average): 滑动平均部分, 表示当前值与过去误差的关系。

ARIMA(0, 0, 2) 表示: 自回归部分 ($p = 0$), 差分次数 ($d = 0$), 移动平均部分 ($q = 2$)

2) 方程描述

因为对象为平稳序列, 可以省略差分过程, 故 ARIMA(0, 2, 2) 模型**的方程可以写作式 (13):

$$Y_t^{(2)} = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t \quad (13)$$

其中, ϵ_t 是白噪声误差项, θ_1, θ_2 是需要估计的参数。

Step4: 对模型的检验与分析

1) 预测值 95%置信区间的确定

对于给定的预测值 $\hat{y}_t + h$ (h 步预测), 95%置信区间的公式为式 (14):

$$\text{下限} = \hat{y}_t + h - z \cdot SE(\hat{y}_t + h)$$

$$\text{上限} = \hat{y}_t + h + z \cdot SE(\hat{y}_t + h) \quad (14)$$

2) 评估结果呈现

本题预测时间步长为 $h = (700)$ 天, 预测的 PT-TT 以及 95%置信区间如表 3 所示:

表 3 预测的 PT-TT 以及 95%置信区间

预测时间	$h=30$	$h=300$	$h=650$
95%置信上界	0.1152	1.5935	3.0259
95%置信下界	-0.1480	-1.3740	-2.6568
预测值 (PT-TT)	-0.0380	0.1104	0.1311

生成的预测值以及置信边界如图 7 所示:

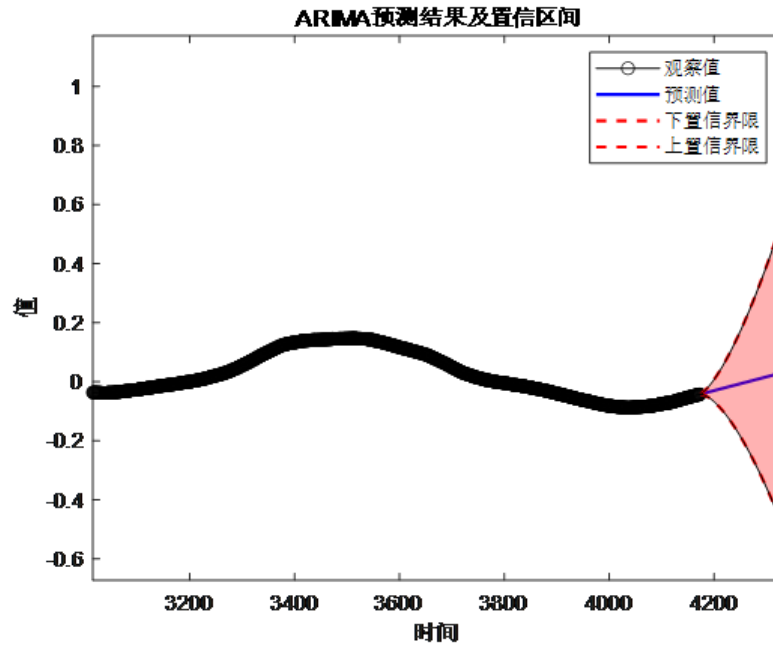


图 7 预测值以及置信边界

3) ARIMA 模型评价与总结

PT-TT 整体变化趋势：在未来 700 天内，95%置信上界的绝对值恒大于置信下界的绝对值，说明未来 PT-TT 总体呈上升趋势。随着天数的增加，脉冲实际到达时间大概率增加。

模型精度分析：通过表格与图可知，随着天数的增加，PT-TT 的置信区间呈几何倍数的增大，在未来几天到一个月内（短期）置信区间远小于几个月到几年的置信区间（长期），这进一步说明随着时间的增加，模型预测的精准度几乎是呈几何倍数下降。

模型的评估：本题构建的 ARIMA 模型适合短期预测，在一个月内 95%置信区间小于 0.28，其预测效果十分可观，但不建议进行以年为单位的长期预测。在短期预测中，虽然因为数据的插值化处理，可能会有极小的误差，但几乎可以忽略不计。

3.5 长期预测的建模（LSTM）：

在长期预测中，未来未知因素造成的影响偶然性强，随着时间的推移预测函数呈现出极强的不确定性。为了更精确地挖掘历史信息，构建更为精确的数学模型，下面运用深度学习中的 LSTM（长短期记忆网络）来进行建模并检验模型预测的精确度。

3.5.1 LSTM 的原理介绍

LSTM（长短期记忆网络）是一种特殊类型的循环神经网络（RNN），专门设计用来处理和预测时间序列数据。它能够有效地捕捉数据中的长期依赖关系。以下是 LSTM 预测原理的详细介绍：

1) 循环神经网络（RNN）基础

RNN 是一种用于序列数据的神经网络。与传统的前馈神经网络不同，RNN 可以处理动态长度的输入序列，并通过隐状态（hidden state）将先前的信息传递给当前的状态。

问题：标准 RNN 在长序列上训练时容易遇到梯度消失或梯度爆炸的问题，导致它难以捕获长期依赖关系。LSTM 的设计是为了解决这个问题。

2) LSTM 的基本单元

LSTM 通过引入门控机制来调节信息的流动，从而就能够记忆重要信息并遗忘不重要的信息。每个 LSTM 单元由以下几个组成部分构成：

细胞状态（Cell State）：存储长期记忆。它在整个链条中传递，可以向前或向后调整。

遗忘门（Forget Gate）：决定哪些信息将从细胞状态中删除。其输入是当前输入和前一个隐藏状态，经过一个 sigmoid 激活函数，输出一个 0 到 1 之间的数值。

输入门（Input Gate）：决定哪些信息将存储到细胞状态中。它包含两个部分：一个 sigmoid 层决定哪些值更新，一个 tanh 层创建新的候选值向量。

输出门（Output Gate）：决定细胞状态的哪个部分将输出到下一层或作为下一步的隐藏状态。

3) LSTM 单元的计算流程

在时间步 t ，LSTM 单元的工作流程如下：

遗忘门 f_t : $f_t = \sigma(W_f \cdot [h_{t-1}, x_t])$ 其中， W_f 是遗忘门的权重， h_{t-1} 是前一个隐状态， x_t 是当前输入。

输入门 i_t : $i_t = \sigma(W_i \cdot [h_{t-1}, x_t])$ $\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t])$ 其中， W_i 是输入门的权重， $\tilde{C}_t$ 是候选值向量。

更新细胞状态 (C_t): $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$

输出门 (o_t): $o_t = \sigma(W_o \cdot [h_{t-1}, x_t])$ 计算输出状态: $h_t = o_t * \tanh(C_t)$

4) 预测过程

输入序列：从时间序列数据中提取输入特征，并形成适合模型的格式。

通过 LSTM 进行前向传播：输入序列通过 LSTM 网络，每个时间步的计算都会更新隐状态和细胞状态。

输出预测：最后时间步的隐状态 h_t 可以通过另一层（如全连接层）转换为最终的预测结果。

5) 训练 LSTM

LSTM 模型使用反向传播算法（包括时间传播）进行训练，损失函数通常使用均方误差（MSE）等方法来优化模型。通过调整模型的权重，LSTM 可以有效地学习序列数据中的模式。

3.5.2 的展示及分析

我们将 70%数据集作为训练集，将剩下 30%作为测试集，在迭代 6000 余次后得到了一个精确的模型，其中测试集的 **RMSE 为 0.0020402**，见图 8。

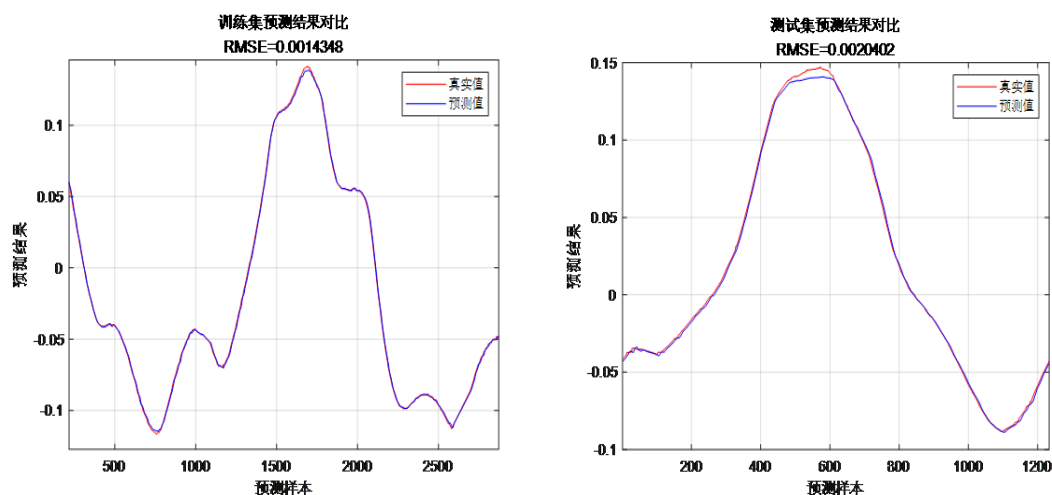


图 8 训练集测试集预测结果对比

其中，在 6000 余次迭代中 RMSE 与损失变化如图 9：

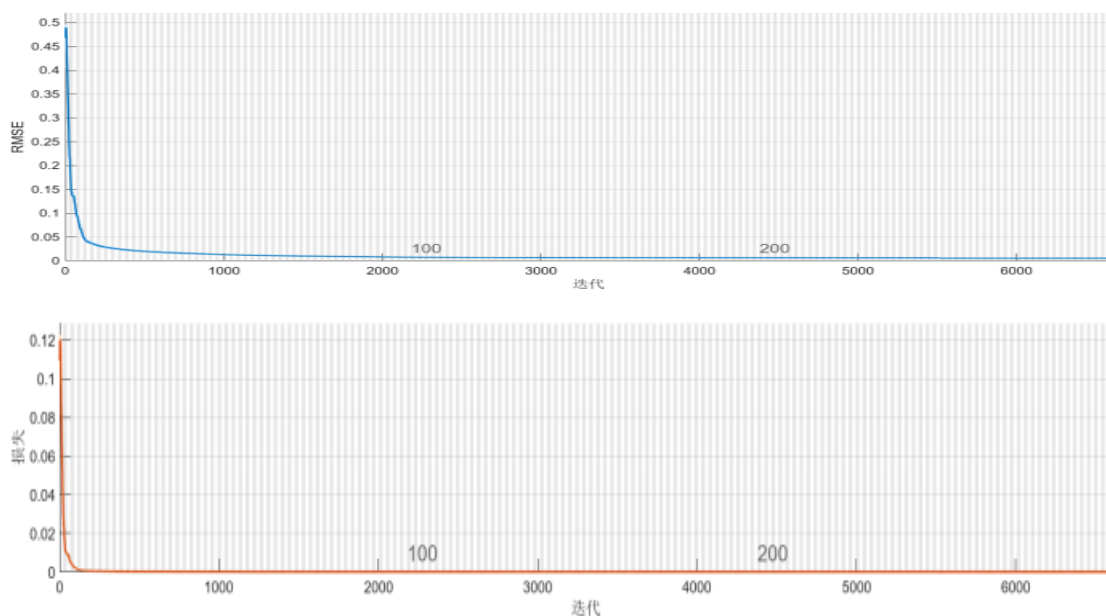


图 9 6000 余次迭代中 RMSE 与损失变化

在迭代次数为 200 时，所对应的 RMSE 点为曲线的肘点，其后 RMSE 无限趋近于 0
在迭代次数为 150 时，所对应的 RMSE 点为曲线的肘点，其后 RMSE 无限趋近于 0
为了更直观的观察预测值的精确度，这里进行了 Q-Q 图的绘制，见图 10：

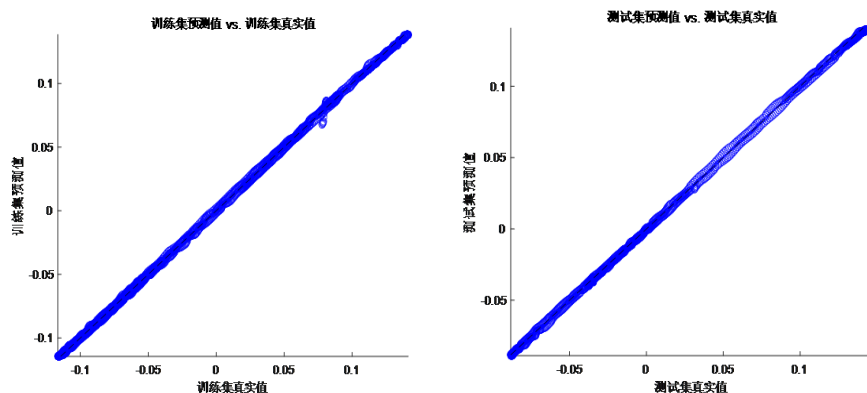


图 10 训练集与测试集的 Q-Q 图

两图的点集在图上基本沿 $y = x$ 分布，可见预测值拟合状况良好。

3.6 模型的评价与总结

本题用了 1200 天（约等于 3 年）作为测试集，在三年尺度下，进行 6000 余次迭代，RMSE 仅有 0.002，可见此模型对于长期预测是有足够准确性的。在短期预测下，此模型的计算成本太大，且 LSTM 作为神经网络，计算迭代过程为黑盒子，难以描述并将过程可视化。总的来说，在以年为单位的时间序列预测情况下，建议用 LSTM 作为预测模型，但以日或月为基本单位的预测情况下，建议使用 ARIMA 模型。

4 问题三模型的建立与求解

4.1 整体建模思路与步骤

本题主干思路：（模型流程见图 11）通过具体分析，我们基于 *Saatamoinen* 模型与 *Herring* 模型建立一个输入-输出模型，变量为环境参数 P , e , ϕ , h 其余大气指标通过计算，为一个基础常量。考虑到现实中观测台为一个定点，在非极端天气下全球温热。

带表面 P （气压）与 e （水气压）的变化微乎其微，结合现实考虑， h 值（接受高度）基本不变，本题将 P , e , h 取一个常量，通过查阅资料以全球温热带平均值作为它们的值，通过调整 ϕ 的值来保证天顶折射时延不超过 7.69 纳秒。

模型的鲁棒性调整：为了将模型在现实应用化，我们将大气参数看成客观参数，在现实中观测时不能主观调整，所以我们将模型中唯一能进行主观改变的 h 值（接受高度）进行确定，在输入-输出模型下将所有大气参数在现实可能发生的范围下调整为令时延最长的参数，让模型在现实环境变化中具有足够的稳健性。

对宇宙电磁噪音进行模拟：模型构建完成后，用蒙特卡洛随机模拟法在一定区间内变化 f 值，尽可能模拟宇宙微波环境，测试在不同海拔高度 h 下宇宙射线天顶时延落在 $(0ns < 7.69ns)$ 的概率。

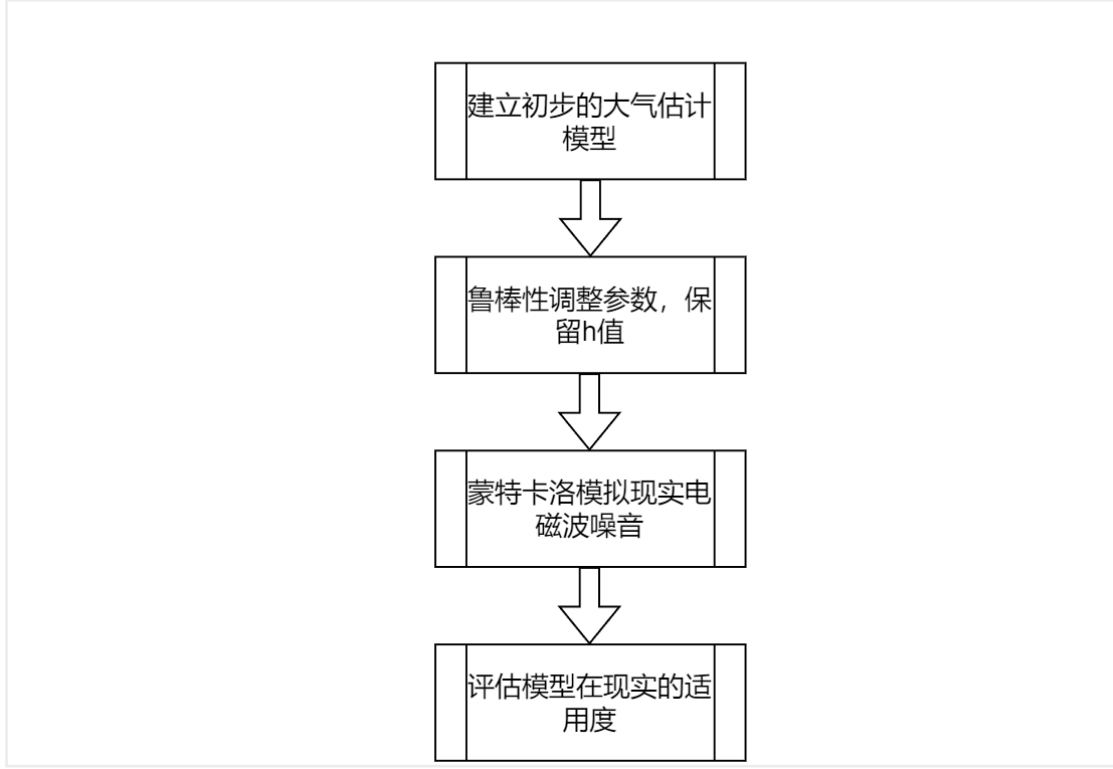


图 11 问题三模型流程图

4.2 创建大气折射估计模型:

Step1: 折射的传播路径的时延表达式为式 (15):

$$\Delta\tau = \int_{h_0}^{h_2} (n(h) - 1) \frac{dl}{c} \quad (15)$$

它一般与大气折射指数 $n(h)$ 有关。

其中: $n(h)$ 是高度 h 处的大气折射指数; dl 是沿传播路径的微分距离; h_0 和 h_2 分别是地面和传播路径的最高点高度; c 是光速。

Step2: Saastamoinen 模型的引入:

由题目可知, 其模型通常用于区分干湿延迟, 通过查阅可以发现, 总延迟可以看成干湿延迟之和, 见式 (16):

$$\Delta\tau = \Delta\tau_{dry} + \Delta\tau_{wet} \quad (16)$$

并且, 通过已有文献可知, 干延迟公式见式 (17):

$$\Delta\tau_{dry} = \frac{0.0022768P}{\cos z} \quad (17)$$

湿延迟公式, 见式 (18):

$$\Delta\tau_{wet} = \frac{0.002277e}{(1 - 0.00266 \cos 2z - 0.00028h)} \quad (18)$$

参考附件, 见式 (19):

$$e = H_w \left(\frac{273.15 + T}{247.1} \right)^{18.36} \quad (19)$$

其中： P 为地面气压(hPa)， z 为观测点天顶角（因为本题要求天顶延迟，所以电磁波垂直于地平线射下，忽略地球曲率影响，相当于天顶角为 0° ）， h 为观测点高度(km)， e 为地面水气压(hPa)， z 为天顶角（当电磁波从天顶方向射入时值为 0° ）， H_w 为大气湿度。

Step3: 频率的依赖性:

通过查阅资料可以得知,实际折射时延随着射电频率 f 的增加而减小，具体公式可以描述为式（20）：

$$\Delta\tau(f) = \Delta\tau \cdot \left(\frac{1}{f}\right) \quad (20)$$

$\Delta\tau$ 是频率 f 下的折射时延； f 为当前信号的频率。

从公式我们可以得知随着频率 f 的增加，折射时延迅速减小，所以在 $f \geq 20GHz$ 的条件下，湿延迟的影响微乎其微。这里，为了将影响具象化，选取 $30GHz$ 为测试频率，结合上述模型可以得出以下结果，见图 12：

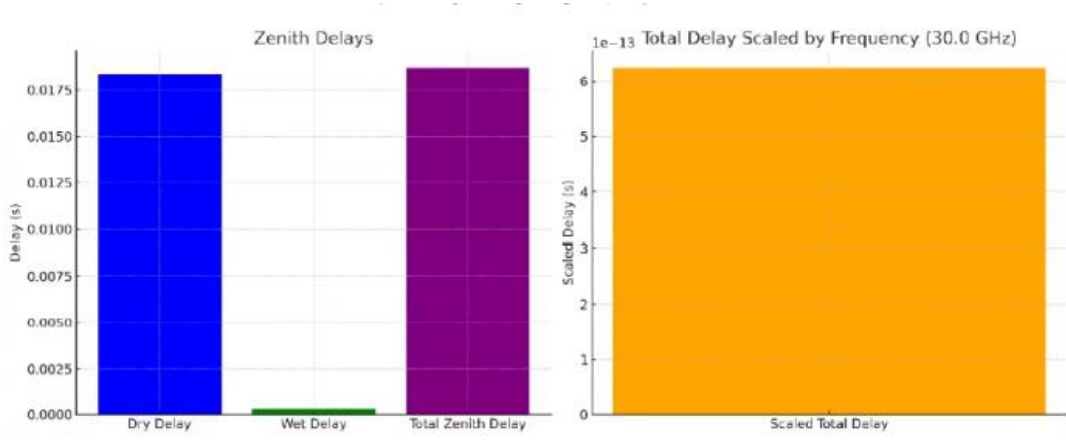


图 12 湿时延的影响可视化

在干延迟约为 0.018s 的情况下，湿延迟仅为 0.00033s，从可视化图像推断湿延迟的影响可以忽略。

Step4: 映射函数的计算

天顶方向的时延需要通过映射函数修正到任意仰角 ϵ 或天顶角 z 下的路径延迟，根据题中所给条件， H_{erring} 映射函数可以表示为式（20）：

$$m(\epsilon) = \frac{1}{\sin \epsilon + a \left(\frac{\cos \epsilon}{1 + b/\sin \epsilon + c} \right)} \quad (21)$$

其中： a , b , c 为经验参数， $\epsilon = 90^\circ - z$ 为仰角。

注：因为本题要求天顶延迟，所以当电磁波从天顶方向射入时， ϵ 应为 90°

经过映射后可以得到折射时延见式（22）：

$$\Delta\tau(\epsilon) = \Delta\tau(f) \cdot m(\epsilon) \quad (22)$$

Step5: 确定大气压与海拔关系:

从经验考虑，这里只考虑观测海拔高度<4000 米的情况，有以下线性关系，见式（23）

$$P = 1013 - 100h \quad (23)$$

Step4: 确定约束条件，见式（24）：

$$\Delta\tau_{weailh} \leq 7.69ns \quad (24)$$

4.3 基于蒙特卡洛模拟法对宇宙微波进行随机模拟

4.3.1 模拟过程

Step1: 参数的鲁棒性估计

为了将模型在现实应用化, 我们将大气参数看成客观参数, 在现实中观测时不能主观调整, 所以我们将模型中唯一能进行主观改变的 h 值(接受高度)进行确定, 在输入-输出模型下将所有大气参数在现实可能发生的范围下调整为令时延最长的参数, 让模型在现实环境变化中具有足够的稳健性。

通过寻阅资料, 将以下参数进行鲁棒性估计, 见表 4:

表 4 不同大气参数的鲁棒性估计

符号	估计值	单位	说明
P	1020	hPa	大气压
T	39	$^{\circ}C$	温度
e	0.8		大气湿度

Step2: 海拔 h 值的, 模拟: 以现实取样, 当今射电望远镜海拔普遍在 $[0, 2]$ (km) 之间, 先进行 10000 次的随机模拟, 当海拔逐渐增加时, 计算频率落在限制区间的概率, 见表 5:

表 5 当海拔逐渐增加时, 频率落在限制区间的概率

f (GHz)	p
20	0.5966
30	0.9350
40	1
.....	
100	1

由模型分析可知, 频率在 40GHz 下变化明显, 当超过此数值, 在现实允许发生的情况下海拔与大气参数无论怎么变化, 其射线时延恒小于 7.65ns。

Step3: 随机 f 值的模拟:

通过查阅资料, 宇宙射线频段大多在 160GHz 前后, 现在宇宙射线频率假设 $[20, 250]$ 在此区间随机均匀随机分布, 先进行 10000 次的随机模拟, 当海拔逐渐增加时, 计算频率落限制区间的概率, 见式 25:

$$p = 10000ps \quad (25)$$

其中: p 为概率, ps 为满足时延小于 7.69ns 的频数。

表 6 当海拔逐渐增加时，频率落限制区间的概率

h (km)	p
0.5	0.95476
0.9	0.95932
1.5	0.96697
1.9	0.97254
.....	
3.9	0.99849

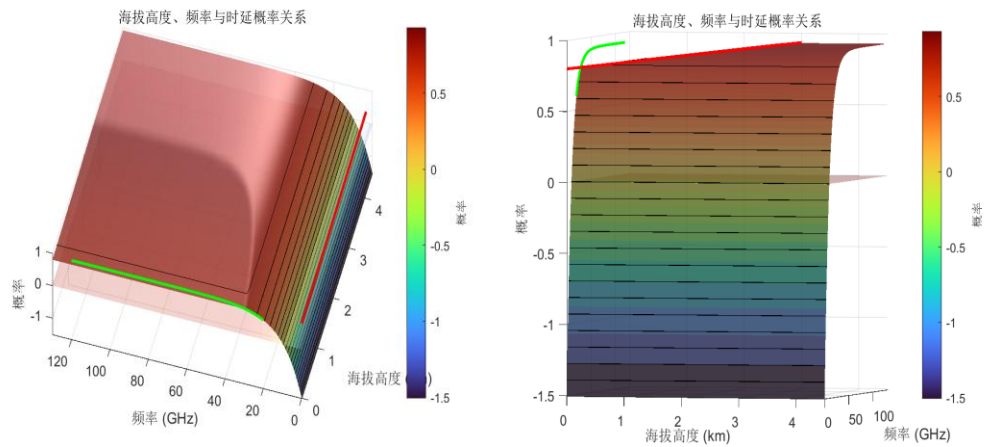


图 13 海报高度，时延与频率落在限制区间概率

4.4 模型的现实意义

通过此模型的建立，若想保证天顶时延恒低于 7.69ns，海拔高度应该在 4km 以上。仅从射线接受效果考虑，海拔 4km 以上大气能见度高，大气折射小，能最小化湿延迟的影响，是进行天文观测的绝佳高度。但通过现实情况可以得知，若要考虑成本以及人员健康问题，实际科研观测应该在 1.5~2km 之间。虽然有小概率出现 40GHz 以下频段电磁波，天顶时延小于 7.69ns，但这种情况难以避免。

5 问题四模型的建立与求解

5.1 对于延迟精度降低因素的深入挖掘

模型构建流程见图 14:

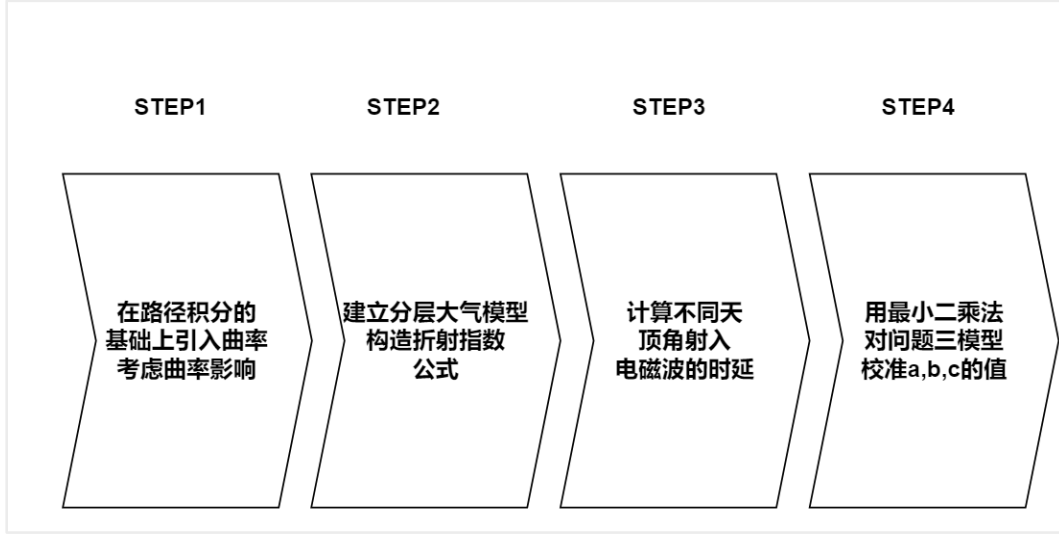


图 14 问题四的模型构建流程

1) 不可控因素: 低仰角下折射路径显著增长, 路径的增长带来大气因素的不确定性增强, 尤其是对流层, 在这种因素下产生的 TOA 误差难以通过模型的形式消除, 所以我们主要考虑通过基于可控因素调整模型, 尽可能的减小误差。

2) 可控因素: 根据附件资料, 引起低仰角误差的根本原因在于大气层的低仰角加剧了大气层曲率对于折射路径的改变程度。在高仰角下, 大气层曲率对于光线路径的影响不大, 在仰角趋近为 90° 的情况下, 大气层的曲率甚至可以忽略。所以, 本题的目的在于引入有关大气曲率的公式, 对于问题二建立的延时模型进行校准。

5.2 对路径积分进行曲率修正:

5.2.1 折射时延的分层积分模型

参考附件一, 在路径积分基础上进行地球曲率引入 (式 23):

$$\Delta\tau_i = \frac{1}{c} \int_{h_i}^{h_{i+1}} (n_i - 1) \cdot \sqrt{1 + \left(\frac{R}{R+h}\right)^2 \sin^2 z} dh \quad (23)$$

其中, 对大气折射指数 $n(h)$ 进行了分层处理, 将大气分成若干层, 分别计算延迟累加, 这里假设第 i 层的厚度为 h 。

5.2.2 对模型进行修正过程

在题目三建立的输入-输出模型的基础上:

Step1: 对大气层分层, 这里将大气层分为 10 层, 即 $i=10$, 通过累加高度 (h) , 确定总延迟。

注: 其他大气参数取现实中的平均值, 频率 f 设定为 30。

Step2: 估计折射指数:

$$n_i = 1.003 - \sum_{i=1}^{10} 0.003 \times 0.1^{0.2} \quad (24)$$

在式 (24) 中, 假设光线能发生路径偏折的高度接近 50km。在高度等于 50km 时, n_i 趋近于 1。

Step3: 引入修正前的路径积分（式 25）：

$$\Delta\tau_i1=\int_{b_i}^{h_i+1}(n_i(h)-1)\frac{dh}{c} \tag{25}$$

5.2.3 对映射函数中 a,b,c 值进行最小二乘估计

为了使新模型更加精确，这里使用最小二乘法调整映射函数函数参数 a，b 和 c。
先输入初始参数：a=1，b=1，c=1.通过优化以下函数：
使残差平方和最小：

$$SSR=\sum_{i=1}^n(y_i-\hat{y}_i)^2 \tag{26}$$

式（26）中：
 y_i 引入 Step2 计算的模型精确值， $\hat{y}_i$ 为在初始模型下的观察值
结果展示与模型评估：

Result: 建立最小二乘估计拟合函数可视化（见图 15）：

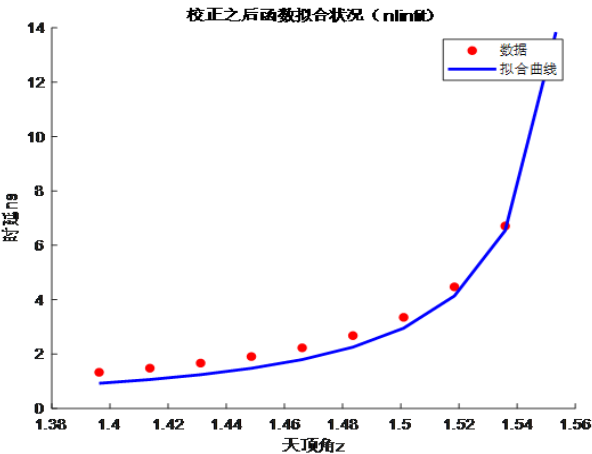


图 15 校正之后的模型拟合情况

校正后得到的映射函数的 a，b，c 值：e+5×（0 0.6866 -3.3163）
其中，MSE 为 0.1722，模型拟合程度良好。
差异性比较（见图 16）：

表 7 校正前后 $\Delta\tau$ 对比

z（度）	校正前 $\Delta\tau$	校正后 $\Delta\tau$
80	0.8989	0.9226
81	0.9923	1.0625
82	1.1886	1.2414
83	1.2021	1.4763
84	1.1512	1.7962
85	1.1932	2.2531
86	2.5513	2.9518
87	3.8898	4.1366
88	6.0989	6.5422
89	11.2485	13.8435

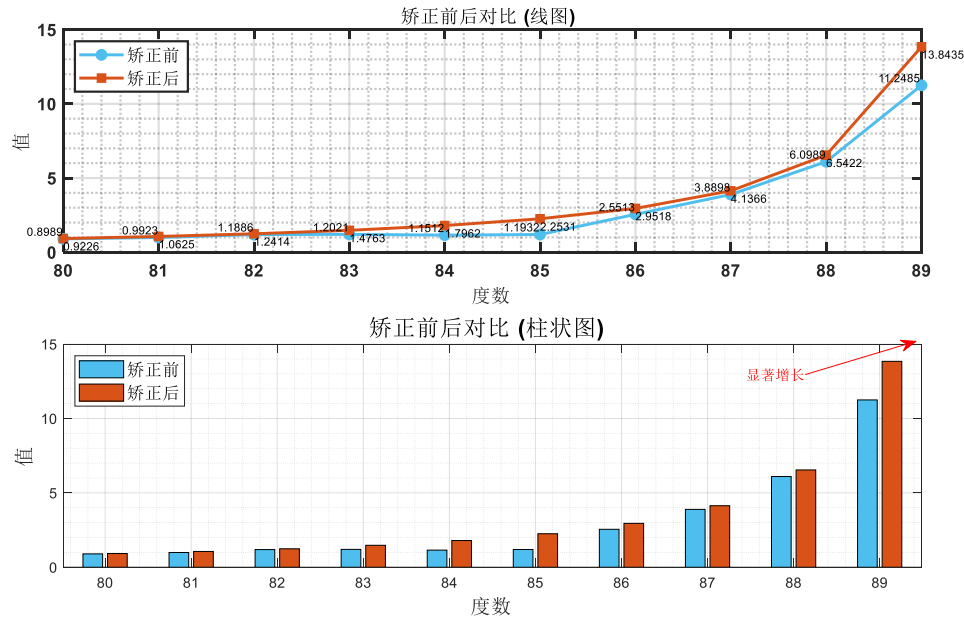


图 16 矫正前后数据可视化对比

5.3 模型分析

据表 7 可知, 在电磁波频率为 160GHz (宇宙最普遍的微波辐射) 时, 当小仰角的情况下, 矫正前时延较矫正后时延短, 根据分析照成这种情况可能是在地球曲率的影响下每一层大气的电磁波入射角增大, 仰角的减小, 加长了电磁波的传播路径, 这种现象再一次证明了模型的正确性。其中, 我们可以观察到当随着仰角的减小, 矫正的误差越大。这个模型完美适用于小仰角情形, 并可以把它扩展到大仰角情形。

参考文献

- [1] KAY S M, MARPLE S L. Spectrum analysis—a modern perspective [J]. Proceedings of the IEEE, 1981, 69(11): 1380-1419.
- [2] SAASTAMOINEN J. Atmospheric correction for the GPS observations [C]// Proceedings of the Third International Workshop on Satellite Altimetry. 1972: 278-290.
- [3] HERRING T A. Modeling the GPS atmosphere: a comparison of two approaches [J]. Journal of Geodesy, 1992, 66(2): 103-113.
- [4] HELLINGS R W, DOWNS G S. Upper limits on the isotropic gravitational radiation background from pulsar timing [J]. The Astrophysical Journal, 1983, 265: L39-L42.
- [5] NEWHALL X X, WILLIAMS J G, DICK A. The new International Astronomical Union (IAU) standard for time [J]. Astronomy and Astrophysics, 1983.
- [6] HOBBS G, et al. Pulsar timing arrays: a review [J]. Classical and Quantum Gravity, 2010, 27(8): 084013.