

DOI: 10.19650/j.cnki.cjsi.J2413355

基于随机森林的 GNSS 观测粗差拟准检定方法*

高 旺¹, 刘裕荐¹, 郭雅娟², 陶贤露¹, 潘树国¹

(1. 东南大学仪器科学与工程学院 南京 210096; 2. 国网江苏省电力有限公司电力科学研究院 南京 210008)

摘 要:在复杂观测环境中,全球导航卫星系统(GNSS)可用卫星数量显著减少,且观测信号易受多维粗差影响,导致粗差识别难、定位性能显著劣化。针对此问题,提出一种基于随机森林自适应分类的拟准检定抗差滤波方法。首先,基于城市环境下的实测动态观测数据,构建了一个基于 GNSS 观测多维数据特征的随机森林分类器,实现将观测数据自适应划分为拟准观测值和非拟准观测值两大类;其次,基于该分类器发展了一种新的 GNSS 拟准检定抗差定位方法,通过分类过程的拟准得分确定拟准观测解算的随机模型,并采用标准化后的真误差代替常规最小二乘残差进行后续的中国科学院地质与地球物理研究所的第 3 种抗差方案(IGG3)抗差检验。以伪距差分定位为例,在两种具有不同量级观测误差的定位场景中对所提方法的有效性进行了评估,实验结果表明,与常规的 IGG3 和拟准检方法相比,该方法定位性能显著提升,不同测试场景下的定位精度提升 16%~51%。所提出的随机森林分类模型有效提高了拟准观测值的识别准确性,因此能够有效提升复杂观测环境下的定位性能。

关键词:GNSS;随机森林;拟准检定;粗差;定位

中图分类号:TH-3 **文献标识码:**A **国家标准学科分类代码:**460.40

Quasi-accurate detection method for GNSS observation gross errors based on random forest

Gao Wang¹, Liu Yujian¹, Guo Yajuan², Tao Xianlu¹, Pan Shuguo¹

(1. School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China;

2. Research Institute, State Grid Jiangsu Electric Power Co., Ltd, Nanjing 210008, China)

Abstract: In complex observation environments, the number of available GNSS satellites is significantly reduced, and the observed signals are susceptible to multidimensional outliers, resulting in challenges in outlier detection and a marked deterioration in positioning performance. To address this issue, this paper proposes a robust filtering method for quasi-accurate detection based on adaptive classification using Random Forests. Initially, we construct a Random Forest classifier utilizing multidimensional data characteristics derived from dynamic observational data collected in urban settings. This classifier facilitates the adaptive categorization of observational data into two principal classes: quasi-accurate observations and non-quasi-accurate observations. Following this classification, a novel robust positioning method for GNSS quasi-accurate detection is developed. This method leverages the quasi-accurate scores obtained during the classification process to determine the random model for quasi-accurate solutions, substituting the standardized true errors for conventional least-squares residuals in the subsequent IGG3 robustness test. Taking pseudorange differential positioning as an example, we evaluate the effectiveness of the proposed method across two positioning scenarios characterized by different magnitudes of observational errors. Experimental results demonstrate that, in comparison to traditional IGG3 and quasi-accurate detection methods, the proposed approach significantly enhances positioning performance, with accuracy improvements ranging from 16% to 51% across various testing scenarios. The random forest classification model proposed in this paper effectively improves the identification accuracy of quasi-accurate observations, and therefore can effectively improve the positioning performance in complex observation environments.

Keywords:GNSS; random forest; quasi-accurate detection; gross errors; positioning

0 引言

随着各全球导航卫星系统(global navigation satellite system, GNSS)的不断建设和完善,卫星导航进入多系统多频融合应用时代。依托充足的卫星数量和丰富的信号频率资源,卫星定位在室外开阔环境下已取得了较好的应用效果。但在城市峡谷、浓密林荫等复杂观测条件下,卫星信号易受遮挡,可用卫星少、观测冗余低;同时易受多路径效应和非直射信号等影响,观测值误差分布非正态性显著,且蕴含较多的粗差^[1]。传统的抗差模型,通常假设正常观测值数量远大于粗差观测值的数量,从而能够依靠残差检验方法检测并剔除存在粗差的观测值。但在复杂观测环境下,卫星数量少、观测冗余低,且粗差出现频繁,常规抗差模型的应用具有明显的局限性,容易造成粗差的误检和漏检。

近年来,众多研究在传统抗差模型的基础上进行扩展,以构建适应复杂环境的抗差模型,例如随机模型的优化、中国科学院地质与地球物理研究所的第3种抗差方案(the 3rd robust scheme of Institute of Geodesy & Geophysics, IGG3)等价权函数优化、粗差观测值的识别与处理以及运动学模型约束。针对随机模型的优化,Zhang等^[2]通过对测站周边环境遮挡信息的建模,建立了由方位角映射截止高度角、信噪比映射等效高度角的复合随机模型;李岚等^[3]深入分析了不同城市环境下的观测信号特性,构建了适应不同场景的随机模型,从而有效减轻部分粗差观测值的影响;蔡亲青等^[4]发展了一种结合高度角和信噪比的加权双因子混合随机模型,进一步提升模型性能;Bahadur等^[5]采用改进的方差模型,提出了一种随机模型细化方法来处理安卓智能手机的码观测数据。在IGG3等价权函数的优化方面,Zhang等^[6]利用稳健的Helmert方差分量估计方法,为不同的观测值分配合理的权重;Yuan等^[7]则结合后验残差和信噪比,提出了一种基于双因子的五段重加权扩展抗差方法;Zhao等^[8]基于新息向量引入抗差因子,通过重新构造等价权重矩阵的方式来调整增益矩阵,从而减少粗差对系统的影响;Zhu等^[9]使用等价权重矩阵和自适应因子,放大了异常观测值的方差和协方差,并利用IGG3等价权函数调节先验信息,有效提升了精度。

复杂环境中粗差观测值的识别和处理是一个具有挑战性的课题,Wang等^[10]开发了一种增强卡尔曼滤波(Kalman filter, KF)过程,通过预测的先验残差和量测的后验残差替代基于高度角的随机模型,自适应估计量测噪声;Yu等^[11]引入单位权方差的先验约束,在迭代中自适应调整阈值以分离未建模误差,实验结果表明该方法可以有效提升浮点解和固定解的精度;陈熙源等^[12]利用

相邻帧间的平均新息构造量测协方差预测值,作为粗差判据来实现稳健估计;Hsu等^[13]则利用基于卡方测试的一致性检验对信噪比和残差进行分析,以剔除非视距(none-line-of-sight, NLOS)观测值,最终选择一致性最好的观测值子集。通过运动学模型约束辅助抗差同样是提升抗差性能的有效手段,张建等^[14-15]通过计算异常值的检验统计量,构造自适应动态调节因子,调整状态预报和量测更新过程,减弱粗差观测值对滤波的不良影响;蒋晨^[16]提出利用惯性导航系统(inertial navigation system, INS)予以辅助,构成泛用性更高的GNSS/INS组合导航系统的抗差区间KF算法,确保滤波器在受到异常观测污染时仍能提供良好的解算效果;王东宇等^[17]根据基于GNSS/INS紧组合KF的新息的协方差,对量测噪声方差进行调节,从而解决多遮挡环境下紧组合定位精度大幅下降的问题。

在上述研究中,高度角、信噪比、残差等特征值被认为与误差高度相关,并在抗差过程中得到广泛应用,为了探究这些特征与误差之间更深层次的关系,同时应对更加复杂的观测环境和更加庞大的数据规模,一些学者尝试引入机器学习技术来辅助抗差模型。Wang等^[18-19]均采用了K均值++(K-means++)聚类方法来识别正常观测信号、多路径和NLOS,通过对NLOS进行分离,在提高模型鲁棒性上取得一定优势,尽管非监督学习方法具有实现简便的优点,但模型可解释性相对较差。Yozevitch等^[20-22]分别使用了决策树、朴素贝叶斯和支持向量机作为分类模型对NLOS和视距(line-of-sight, LOS)卫星进行监督式分类,并且它们在高度角、信噪比、伪距残差的基础上都引入了多普勒一致性等额外特征。Zheng等^[23]提出基于集成堆叠学习模型的NLOS检测方法,通过融合不同基础模型减少偏差并降低泛化误差。Zhu等^[24-25]进一步探讨了GNSS观测值的更多特征值,系统比较了支持向量机、多层感知机、卷积神经网络、递归神经网络等分类模型在场景识别中的表现,但并未过多讨论模型在观测值误差方面的识别能力。

多粗差场景中,许多抗差模型效果的发挥受到限制,拟准检定方法(quasi-accurate detection, QUAD)作为一种较新的抗差模型提供了一种解决思路。该方法具有较强的粗差探测能力,在系统中存在多个粗差观测值的情况下,依然能够实现准确的探测^[26]。该方法的核心在于确定拟准观测值(quasi-accurate observations, QUO),黄立人^[27]首次提出了一种适用性较强的QUO确定方法,并将其应用到地壳运动的观测中;Luo等^[28]在QUAD框架下提出了一种自适应的KF方法,通过运动学推导的先验残差来识别QUO,并将QUAD与抗差KF结合,展示了该方法在处理粗差观测值和运动学突变方面的优越性能;赵俊等^[29]采用观测值的L1范数和中位数确定

QUO,并运用均值漂移模型对观测值进行修复;Qu等^[30]通过计算每个冗余观测值的方差膨胀系数,使用K-means++聚类方法对QUO和非QUO进行区分,并通过蒙特卡洛模拟验证了该方法在异常值占比超过50%时的有效性。

传统的QUAD研究中对于QUO的确定通常依赖于充分冗余的观测值,在观测值不足的复杂环境下鲁棒性较差。由此,本文提出引入随机森林机器学习模型,探索和挖掘更多的GNSS特征值,并建立用于识别QUO的分类器。这一方法在现有研究的基础上进行扩展,通过将随机森林(random forest, RF)与QUAD结合,提出了一种基于随机森林识别的QUAD抗差滤波方法,旨在为复杂环境下的不同场景提出更具有自适应性的抗差手段。

1 拟准检定方法

1.1 拟准检定方法原理

Gauss-Markov模型的构造为:

$$\mathbf{L} = \mathbf{A}\mathbf{\beta} + \mathbf{\Delta} \quad (1)$$

式中: $\mathbf{L}$ 为 $n \times 1$ 的观测值向量; $\mathbf{A}$ 为 $n \times t$ 的设计矩阵; $\mathbf{\beta}$ 为 $n \times 1$ 的待估参数向量; $\mathbf{\Delta}$ 为 $n \times 1$ 的真误差向量。观测值的随机特性可以表示为:

$$D(\mathbf{L}) = \sigma_0^2 \mathbf{P}^{-1} \quad (2)$$

式中: $D(\mathbf{L})$ 是 $\mathbf{L}$ 的方差; σ_0^2 是单位权方差; $\mathbf{P}$ 是 $\mathbf{L}$ 的权矩阵。

以伪距差分单历元定位模型为实例,上述模型中的向量和矩阵可分别表示为:

$$\mathbf{L}_i = [\Delta \nabla p_{ij} - \Delta \nabla \rho_{ij} + \Delta \nabla(v_{ion})_{ij} + \Delta \nabla(v_{trop})_{ij}] \quad (3)$$

$$\mathbf{A}_i = [-\Delta e_i^x \quad -\Delta e_i^y \quad -\Delta e_i^z] \quad (4)$$

$$\mathbf{P}_{ii} = \left(a + \frac{b}{\sin(ele_i)} \right)^{-1} \quad (5)$$

式中: p 是伪距观测值; ρ 是站星距; v_{ion} 和 v_{trop} 分别是电离层和对流层误差改正项; e^x 、 e^y 、 e^z 分别为测站到卫星方向向量在 x 、 y 、 z 3个方向上的分量; ele 为卫星高度角; a 、 b 为高度角随机模型的两个预设系数; $\Delta \nabla$ 为星间、站间双差算子; Δ 为星间单差算子; i 、 j 分为非参考星和参考星的编号。

通过最小二乘(least squares, LS)的方法对该模型进行求解,待估参数的估值可以表示为:

$$\hat{\mathbf{\beta}} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P} \mathbf{L} \quad (6)$$

对应的残差为:

$$\mathbf{V} = \mathbf{A}\hat{\mathbf{\beta}} - \mathbf{L} \quad (7)$$

将式(6)代入式(7),可以得到由观测值空间到残差空间的投影公式为:

$$\mathbf{V} = -\mathbf{R}\mathbf{L} \quad (8)$$

式中: $\mathbf{R} = \mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}$,称为平差因子阵; $\mathbf{I}$ 为 $n \times n$ 的单位矩阵。在式(1)两边同时左乘 $\mathbf{R}$,得到:

$$\mathbf{R}\mathbf{\Delta} = \mathbf{R}\mathbf{L} \quad (9)$$

式(9)表明真误差与观测值在经过 $\mathbf{R}$ 变换后的投影是相等的,因此可以通过 $\mathbf{R}$ 计算出真误差的大小,又由于 $\mathbf{R}$ 秩亏,求解真误差需要构造额外的约束条件。

假设可以在观测值集合中确定大小为 r 的子集,其是未被粗差污染的干净观测值,且残差服从正态分布,构造最小化残差的目标函数为:

$$\min(\mathbf{V}_r^T \mathbf{P}_r \mathbf{V}_r) \quad (10)$$

式中: $\mathbf{V}_r$ 、 $\mathbf{P}_r$ 分别为观测值子集的残差向量和权矩阵,大小为 $r \times 1$ 和 $r \times r$ 。对式(10)进行求导,当导数等于0时目标函数取得最小值,即:

$$\mathbf{A}_r^T \mathbf{Q}_r^{-1} \mathbf{V}_r = 0 \quad (11)$$

式中: $\mathbf{A}_r$ 、 $\mathbf{Q}_r^{-1}$ 分别为观测值子集的设计矩阵和噪声矩阵, $\mathbf{Q}_r^{-1} = \mathbf{P}_r$ 。因为所选的观测值子集是无粗差的,所以可以将残差用真误差近似代替,同时考虑到观测值全集与子集之间的转移,有:

$$\mathbf{A}_r^T \mathbf{Q}_r^{-1} \mathbf{W} \mathbf{\Delta} = 0 \quad (12)$$

式中: $\mathbf{\Delta}_r = \mathbf{W} \mathbf{\Delta}$, $\mathbf{W}$ 是大小为 $r \times n$ 的转移矩阵。联合式(9)得到:

$$\begin{cases} \mathbf{R}\mathbf{\Delta} = \mathbf{R}\mathbf{L} \\ \mathbf{G}\mathbf{\Delta} = 0 \end{cases} \quad (13)$$

式中: $\mathbf{G} = \mathbf{A}_r^T \mathbf{Q}_r^{-1} \mathbf{W}$ 。式(13)即为完整的真误差约束方程组,对其使用最小二乘方法可以得到真误差的估计公式,即:

$$\hat{\mathbf{\Delta}} = (\mathbf{P}\mathbf{R} + \mathbf{G}^T \mathbf{G})^{-1} \mathbf{P}\mathbf{R}\mathbf{L} \quad (14)$$

在首次计算出各观测值的真误差后,可根据真误差的大小重新确定观测值子集,并再次计算真误差,重复此步骤直到连续两次确定的子集相同或达到最大重复次数为止。计算第 i 个观测值真误差的检验指标如下:

$$\mathbf{W}_i = \frac{|\hat{\mathbf{\Delta}}_i|}{C} \quad (15)$$

其中:

$$C = \left(\frac{\mathbf{\Delta}_r^T \mathbf{P}_r \mathbf{\Delta}_r}{r-1} \right)^{\frac{1}{2}} \quad (16)$$

式中: $\mathbf{\Delta}_r$ 是所选观测值子集的真误差向量。若该检验指标小于一定门限值,即将该观测值纳入下一轮计算的子集。参考文献[31]可知,该门限值可取值在2.2~2.3。

上述拟准检定方法,所选择的观测值子集称为拟准观测值。拟准检定的本质是通过一组相对稳定的拟准观测值,去估计整个观测值集合的真误差,因此估计结果的好坏很大程度依赖于拟准观测值的选取。现有的拟准检定方法一般基于先验或后验残差的大小来确定拟准观测值,但在一个平差系统内,一旦误差发生转移,会使得根

据残差客观评价观测值变得困难,从而无法准确判定拟准观测值。为了减少这种偏差并准确评估观测值的好坏,本文提出一种基于随机森林分类方法的评分机制,通过引入系统外因素以优化拟准观测值的判定。

1.2 基于随机森林的GNSS拟准检定方法

如图1所示,随机森林是一种由多个弱学习者构成的集成学习器,每个弱学习者为一棵决策树。从样本特征矩阵中随机选择一部分样本和特征,形成子矩阵,每个子矩阵负责一棵决策树的训练。在对测试样本进行分类时,每棵决策树都会给出自己的分类结果。通过统计所有结果的众数,来确定这个样本的最终类别。如果某一类别的众数出现次数较多,那么该类别得分较高,意味着这个样本的类别特征更加明显。随机森林具有如下适用于GNSS观测值的优良特性。

1) 每棵决策树都只根据样本的一部分特征进行类别判断,所以模型允许样本有缺失特征,加大了对训练样本的利用率和分类的覆盖率。

2) 观测值的误差具有随机性,存在不可忽略的噪声,而随机森林对噪声样本的容忍度使它能更好地学习观测值特征。

3) 模型中的每个弱学习者都是非线性的决策树,因此对样本特征的缩放没有要求,基于同样原因,它也对离群样本不敏感。

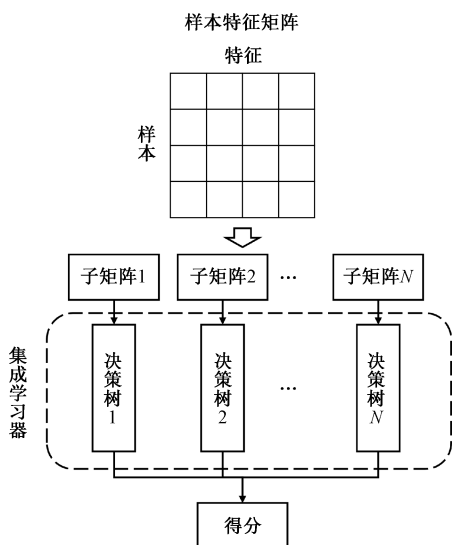


图1 随机森林模型结构

Fig. 1 Schematic diagram of the random forest model structure

在处理训练样本时,根据观测值的误差大小来决定其是否为拟准观测值,并以是否为拟准观测值作为样本标签。通常误差是区间为 $(0, +\infty)$ 的连续变化量,而样本标签是0或1的离散取值,当误差接近拟准观测值的阈值时,样本特征的显著性会变低,样本特征与样本标签之间的相关性会减弱。为了削弱这类观测值对样本训

练的影响,本文引入特征与标签的相关性因子,构造了一个分段函数来表示这种相关性,如下所示:

$$w = \begin{cases} ae + b, & e \leq \text{thresh} \\ \ln e + c, & e > \text{thresh} \end{cases} \quad (17)$$

式中: w 为特征与标签之间的相关性因子; e 为误差; thresh 为选择拟准观测值的阈值; a, b, c 都是分段函数的系数,可根据实际情况调整。该函数称为观测值误差与样本特征-标签相关性函数,函数曲线如图2所示。

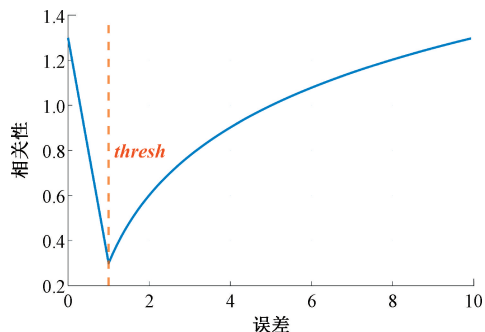


图2 观测值误差与样本特征-标签相关性

Fig. 2 Observation error and sample feature-label correlation

为了尽可能全面地描述一个观测值的信息,本文选择了如下数据作为样本特征:高度角、信噪比、信噪比衰减、几何贡献、信号完整率、差分无电离层组合(difference of melbourne-wübbena combination, dMW)观测值、差分无几何组合(difference of geometry-free combination, dGF)观测值、载波一致性、多普勒一致性、单点定位残差、失锁标志位(loss of lock indicator, LLI)、几何精度衰减因子(geometric dilution of precision, GDOP)、卫星数、卫星利用率。其中,与观测值误差相关性越大的特征,对模型分类的贡献也越大。

上述特征中,高度角、信噪比、信号完整率、LLI标志位、卫星数可以直接通过与接收机无关的交换格式(receiver independent exchange format, RINEX)文件得到, GDOP、卫星数、卫星利用率则是一个历元所有样本所共享的特征,其他需要经过一定预处理才能得到的特征获取方法如下。

1) 信噪比衰减,用开阔环境下的理想观测信号拟合高度角-信噪比模板函数,相同高度角下标准信噪比与实际信噪比之差即为信噪比衰减,表示为:

$$\text{Att}_k = S_k^{\text{std}}(\theta) - S_k \quad (18)$$

式中: S_k 为卫星 k 的实际信噪比; $S_k^{\text{std}}(\theta)$ 为卫星 k 在高度角 θ 下的标准信噪比; Att_k 为卫星 k 的信噪比衰减。

2) 几何贡献,由式(8)可知, $\mathbf{R}$ 矩阵的第 k 个对角线元素越大,则卫星 k 的观测值对自身残差的影响越大,相反对其余观测值的残差影响越小,其代表该观测值在整个平差系统中所占的分量,表示为:

$$r_k = R_{kk} \quad (19)$$

式中: r_k 为卫星 k 对系统的几何贡献。若 $r_k = 0$, 则表示为观测值必要观测; 若 $r_k = 1$, 则表示观测值完全多余。

3) dMW 与 dGF, 即观测值无电离层组合与无几何组合分别的历元差, 在伪距不存在粗差且载波未发生周跳的情况下, 这两个特征应该维持在一个较为稳定的区间。无电离层组合与无几何组合的非差计算方法为:

$$MW = L_1 - L_2 - \frac{f_1 - f_2}{f_1 + f_2} \left(\frac{P_1}{\lambda_1} + \frac{P_2}{\lambda_2} \right) \quad (20)$$

$$GF = \lambda_1 L_1 - \lambda_2 L_2 \quad (21)$$

式中: L 为载波观测值; P 为伪距观测值; f 为载波频率; λ 为载波波长; 下标 1、2 代表该观测值的频点。

4) 载波一致性与多普勒一致性, 即卫星某一频点载波与伪距的一致性、多普勒与伪距的一致性, 同样在正常情况下两个特征不应该有太大的变动。其计算方法为:

$$LP = P(t) - P(t-1) - \lambda L(t) + \lambda L(t-1) \quad (22)$$

$$DP = P(t) - P(t-1) - dt\lambda \frac{D(t) + D(t-1)}{2} \quad (23)$$

式中: LP 为载波一致性; DP 为多普勒一致性; t 和 $t-1$ 代表两个相邻的历元; dt 为其时间差。

5) 单点定位残差, 将非差伪距观测值代入式(1)、(2)的模型后, 应用最小二乘法所求解出的残差即为单点定位残差。

传统的 QUAD 方法在确定了拟准观测值后, 通常遵循观测值的原始权重进行真误差的计算, 不会涉及重新定权, 因而丢失了确定拟准观测值过程中的中间信息。本文为了充分利用随机森林分类器的信息, 对于一个训练完成的分类器, 其将会为每一个观测值样本提供一个拟准得分。得分越高, 代表观测值为拟准的概率越大, 相应地误差分布也越小, 为了客观反映这种误差分布对真误差计算的影响, 本文对已确定的拟准观测值实施重新定权操作, 将式(12)替换为:

$$A^T \tilde{Q}_r^{-1} W \Delta = 0 \quad (24)$$

式中: $\tilde{Q}_r$ 为拟准得分定权的噪声矩阵, 其对角线第 k 个元素的取值如式(25)。

$$(\tilde{Q}_r)_{ii} = \left(a + \frac{b}{score_i} \right)^2 \quad (25)$$

式中: $score_k$ 为第 i 个观测值的拟准得分; a 、 b 为可根据实际情况调整的系数。本文经过实验, 建议合适的取值为 $a = -1$ 、 $b = 1.5$ 。

将式(24)代入式(14)计算得到观测值全集的真误差 Δ , 再用真误差代替最小二乘的验后残差进行 IGG3 抗差, 即:

$$\gamma_{ii} = \begin{cases} 1, & |\tilde{\delta}_i| \leq k_0 \\ \frac{k_0}{|\tilde{\delta}_i|} \left(\frac{k_1 - |\tilde{\delta}_i|}{k_1 - k_0} \right)^2, & k_0 < |\tilde{\delta}_i| \leq k_1 \\ 0, & |\tilde{\delta}_i| > k_1 \end{cases} \quad (26)$$

式中: γ_{ii} 为第 i 个观测值的噪声放大系数; $\tilde{\delta}_i$ 是第 i 个观测值经过标准化后的真误差; k_0 、 k_1 的建议取值是 $k_0 = 1.5$ 、 $k_1 = 2.5$ ^[32]。因为真误差是可信度高的拟准观测值计算出的, 理论上未受到粗差转移, 相比最小二乘的验后残差更接近实际的误差, 所以 IGG3 抗差模型对观测值噪声的重新标定也更符合真实的误差分布。

基于上述分析, 本文所提出的基于随机森林的拟准检定方法示意图如图3所示, 基本步骤为:

1) 采集训练集数据, 提取相应的观测值特征, 并以误差大小作为分类标签, 用式(17)表示特征与标签之间的相关性, 训练拟准观测值的随机森林分类模型。

2) 用另外一组和 1) 训练集数据不同的测试集数据, 通过相同的方法提取观测值特征, 用 1) 训练得到的分类模型挑选出每个历元的拟准观测值。

3) 用式(24)、(25)对拟准观测值定权, 式(14)计算出观测值全集的真误差。

4) 用式(15)计算真误差的检验指标, 重复 3)~4), 直到两次确定的拟准观测值没有变化或达到最大重复次数为止。

5) 用真误差替换验后残差, 代入式(26)计算噪声放大系数, 完成 IGG3 抗差。

6) 将抗差后的噪声矩阵求逆转换为权重矩阵, 替换式(5), 联合式(3)、(4)进行单历元伪距双差模型的最小二乘解算, 得到定位结果。

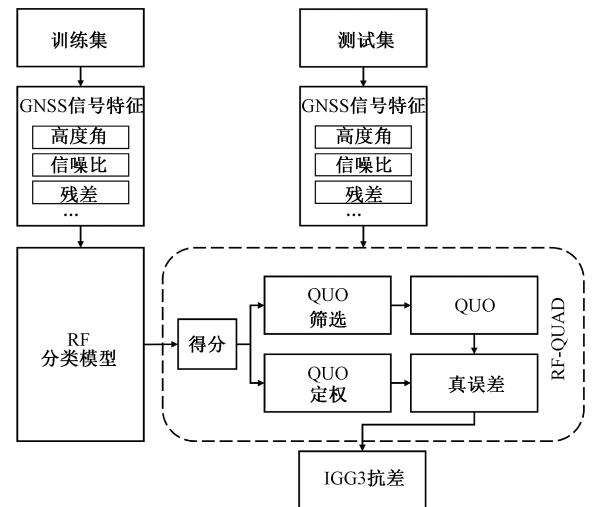


图3 基于随机森林的拟准检定方法

Fig. 3 Schematic diagram of the quasi-accurate detection method based on random forest

2 分场景实验验证

2.1 分类模型训练

本文所使用的车载数据采集平台,以 NovAtel 组合导航设备的定位结果作为行车轨迹参考真值,并用泰斗低成本接收机 N307 采集卫星观测信号。在南京市区内采集了约 1.4 h 的动态观测数据作为模型训练集,这组数据的观测值误差序列如图 4 所示。其中圆形样本点误差绝对值 < 1 m,被标记为拟准观测值;十字样本点误差绝对值 > 1 m,被标记为非拟准观测值(none quasi-accurate observations, NQUO)。

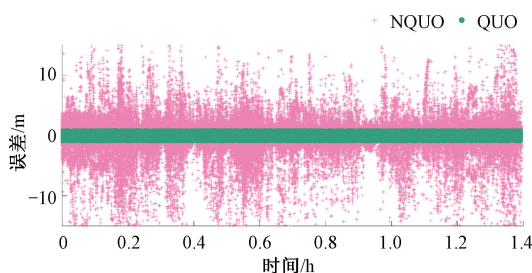


图4 训练集观测误差序列

Fig. 4 Training set observation error sequence diagram

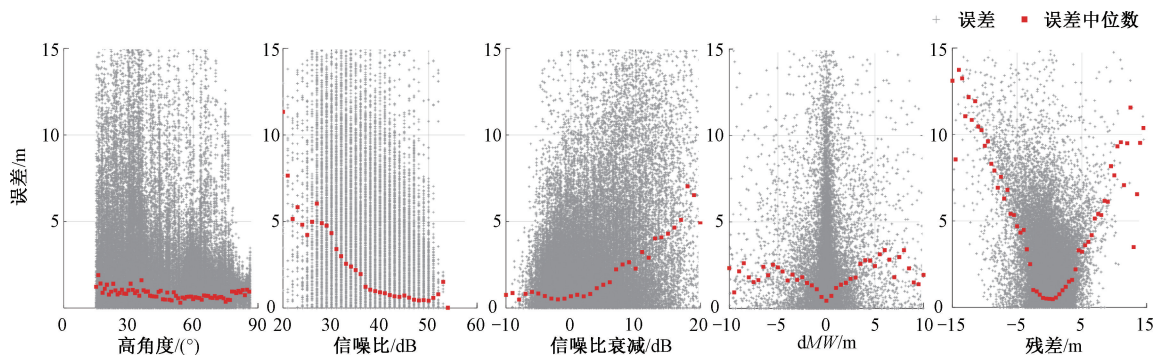


图5 训练集观测误差随不同特征的分布

Fig. 5 Distribution of training set observation error with different features

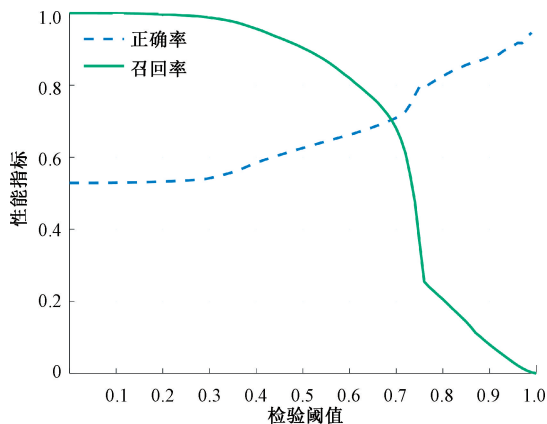


图6 模型指标在不同检验阈值下的变化情况

Fig. 6 Changes in model indicators at different test thresholds

由处理复杂环境下 GNSS 数据的经验可知,本文提出的 14 种特征中,高度角、信噪比、信噪比衰减、 dMW 和最小二乘残差是在抗差处理过程中较为关键的特征。这些特征在本文构建的随机森林分类模型中同样表现出较高的重要性,此外与文献[20-25]分类模型所采用的主要特征相吻合,从而进一步验证了实用价值与理论依据。误差在这些特征上的分布情况如图 5 所示,其中方形样本点为误差的中位数。观察到在动态场景下,信噪比、信噪比衰减、 dMW 、残差都存在较显著的误差相关性;相比之下,高度角的相关性较弱,主要是因为动态场景下,接收机周围的遮挡物也在快速变化,这种变化使得即使在相同高度,不同的时刻也可能面临完全不同的视野,这种快速变化的环境条件削弱了在动态场景中高度角与误差之间的相关性。

为了验证模型效果,选择与训练集不同的测试集样本以评估模型的泛化能力。重点考虑两个指标:正确率(模型识别的拟准观测值中,真正的拟准观测值所占的比例)、召回率(所有实际的拟准观测值中,模型正确识别的比例)。这两个指标在不同检验阈值的变化情况如图 6 所示,其中检验阈值是指用于决定样本是否被归类为拟准观测值的得分门槛。当该值为 0.7 时,模型的正确率和召回率均达到 0.7,表明在该阈值下模型表现较

为均衡和有效。未通过检验阈值的样本被归类为非拟准观测值,然而这并不意味着它们一定受到粗差影响,若经过适当处理这些样本仍然具有可利用的潜力。随得分变化的误差分布情况如图 7 所示,其中方形样本点代表误差的中位数,当得分超过 0.5 时,误差中位数保持在 2 m 以内,表明样本虽然未被归类为拟准观测值,但它们依然具有较高的可信度。

不同特征对随机森林分类模型的重要性排名如图 8 所示,其中重要性最高的特征依次为残差、信噪比、 dMW 和信噪比衰减;与之相比高度角的重要性较低,这一结果与前文对于动态场景的分析一致;与分类性能几乎无关的特征包括信号完整率和 LLI 标志位,表明该分类模型可以进一步简化。

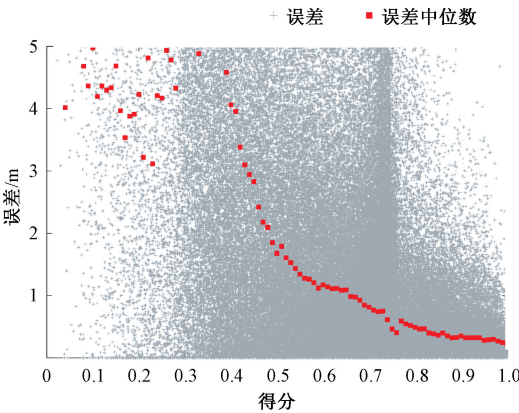


图 7 测试集观测误差随得分的分布

Fig. 7 Distribution of test set observation error with score

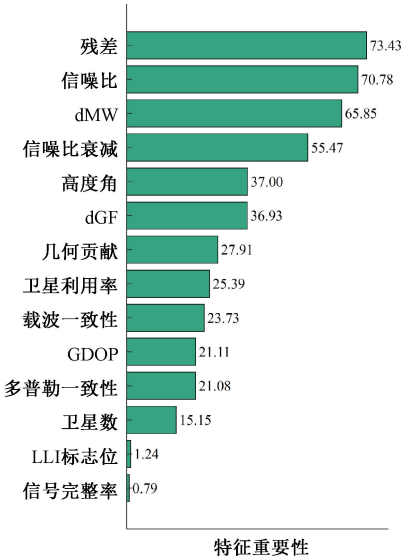


图 8 分类特征重要性

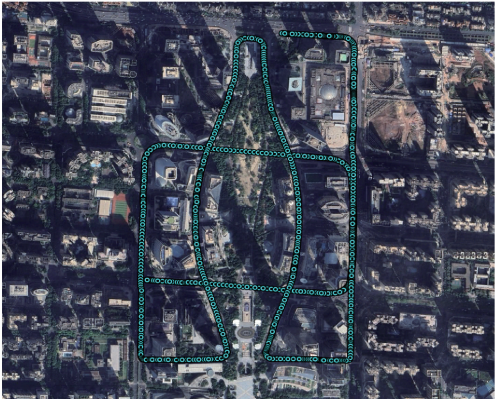
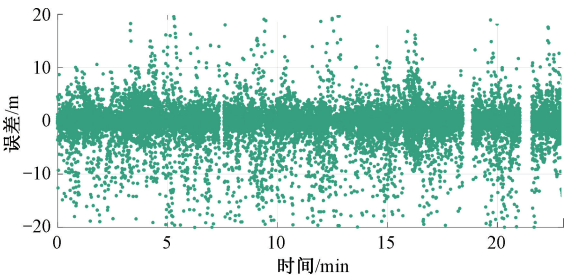
Fig. 8 Classification feature importance

2.2 模型分场景定位性能测试

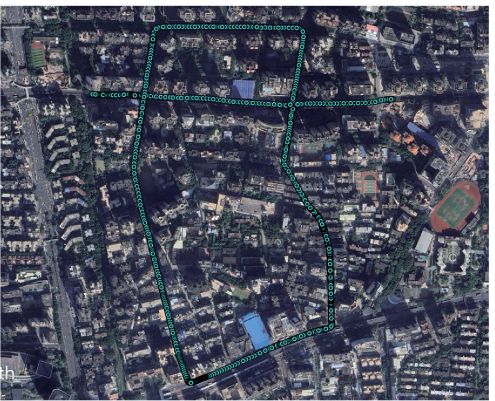
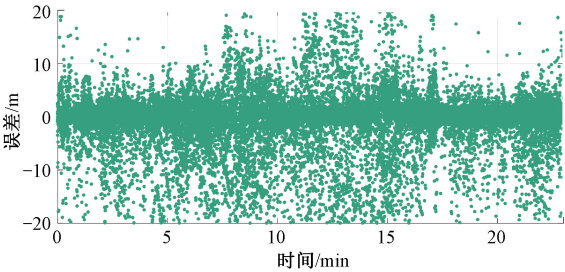
为了验证本文所提基于随机森林的拟准检定方法定位性能,通过与采集训练集时相同的低成本接收机泰斗 N307 采集了一组新的数据作为测试集。从测试集中截取两个场景观测数据进行测试,它们分别是时长约 23 min 的场景 A 和时长约 30 min 的场景 B。图 9 所示为通过两个场景测试路段的观测误差序列图和卫星俯视图展示两种场景的差异性,场景 A 为商业区,是人行广场、林荫步道、高层办公楼的混合路段,环境成分复杂,观测误差均方差为 6.33 m;场景 B 为居民区,道路两侧聚集大量高层住房,视野遮挡严重,观测环境较场景 A 更为恶劣,观测误差均方差为 9.21 m。

设计了如下 3 组不同抗差方案的对比实验。

1) 传统的抗差方法 (LS-IGG3), 利用最小二乘方法计算出后验残差,使用 IGG3 抗差模型进行抗差。



(a) 场景 A
(a) Scenario A



(b) 场景 B
(b) Scenario B

图 9 两个场景测试路段观测误差序列图和卫星俯视图
Fig. 9 Observation error sequence diagram and satellite bird's-eye view of the two test scenarios

2) 传统的拟准检定方法 (LS-QUAD), 利用最小二乘方法计算出的后验残差来筛选拟准观测值, 并计算真误差, 基于真误差使用 IGG3 抗差模型进行抗差。

3) 基于随机森林的拟准检定方法 (RF-QUAD), 收集观测值特征, 接着利用随机森林分类器评估观测值的拟准得分, 以此对拟准观测值进行筛选并确定真误差, 基于真误差使用 IGG3 抗差模型进行抗差处理。

3 种方法分别在场景 A 和 B 的定位轨迹如图 10 所示, 可以看到相比于 LS-IGG3 方法和 LS-QUAD 方法, RF-QUAD 方法的轨迹更为集中, 明显偏离轨迹的定位点也更少, 这说明该方法通过引入随机森林分类器, 可以更有效地识别粗差观测值, 定位精度显著提高、鲁棒性大幅增强。

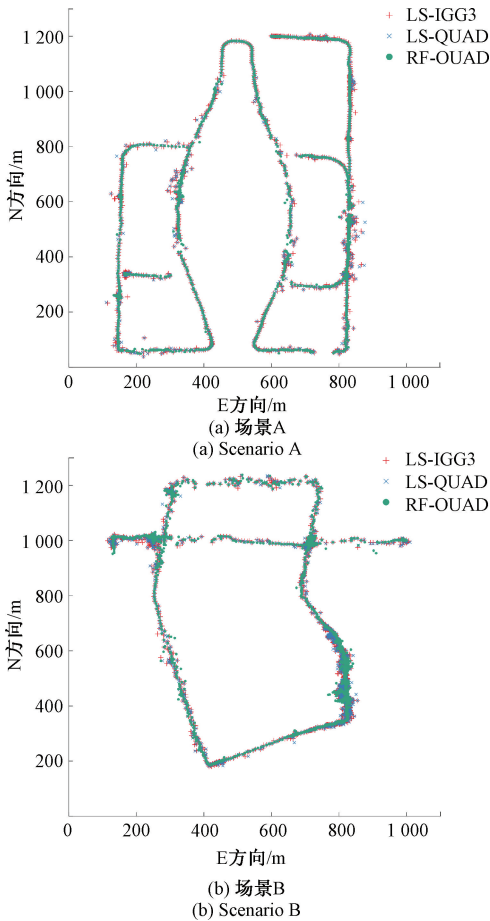


图 10 3 种方法的定位轨迹比较
Fig. 10 Comparison of positioning trajectories of three methods

在场景 A, LS-IGG3 方法在 E、N、U 这 3 个方向上的精度分别为 5.50、3.99、11.37 m, LS-QUAD 方法在 E、N、U 3 个方向上的精度分别为 5.64、3.83、11.97 m, 两者的精度差距很小。这是因为两种方法都处于同一个最小二乘的平差系统中, 传统的 QUAD 抗差方法必须有充分冗余的观测, 才能通过后验残差确定合适的拟准观测值, 而本测试场景处于复杂的城市环境中, 视野受限、多路径效应

显著, 不具备充分冗余的观测条件。RF-QUAD 方法在 3 个方向上的精度分别为 2.66、2.32、7.58 m, 较前两者提升明显, 这是因为该方法通过引入随机森林这个平差系统外的因素来对观测值进行客观评分, 而不会受到系统内的粗差影响, 抗差能力更强。3 种方法在场景 A 的误差序列如图 11 所示, 对应的具体精度数据与提升百分比如表 1 所示, 其中精度提升 1 为 RF-QUAD 方法相对于 LS-IGG3 方法的精度提升百分比, 精度提升 2 为 RF-QUAD 方法相对于 LS-QUAD 方法的精度提升百分比。

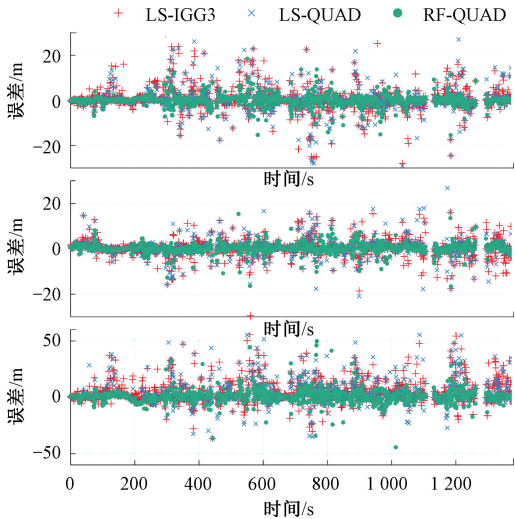


图 11 场景 A 3 种方法的误差序列
Fig. 11 Error sequence of three methods in scenario A

表 1 场景 A 3 种方法的精度以及相对精度提升					
Table 1 Accuracy and relative accuracy improvement of the three methods in scenario A					
方向	LS-IGG3 /m	LS-QUAD /m	RF-QUAD /m	精度提 升 1/%	精度提 升 2/%
E	5.50	5.64	2.66	51.6	52.8
N	3.99	3.83	2.32	41.9	39.4
U	11.37	11.97	7.58	33.3	36.7

在场景 B, LS-IGG3 方法在 E、N、U 3 个方向上的精度分别为 7.84、7.78、29.80 m, LS-QUAD 方法在 E、N、U 3 个方向上的精度分别为 7.96、7.89、28.00 m, 显然相比于场景 A, 场景 B 的环境更加复杂, 两种传统方法都无法很好地处理这些数据, 并且与场景 A 的情况相同, 两种方法的定位性能相差不大。RF-QUAD 方法在 3 个方向上的精度分别为 6.55、6.11、14.39 m, 和前两者相比, 水平方向精度有一定提升, 垂直方向精度提升明显。3 种方法在场景 B 的误差序列如图 12 所示, 对应的具体精度数

据与提升百分比如表 2 所示,其中精度提升 1 为 RF-QUAD 方法相对于 LS-IGG3 方法的精度提升百分比,精度提升 2 为 RF-QUAD 方法相对于 LS-QUAD 方法的精度提升百分比。

表 2 场景 B 3 种方法的精度以及相对精度提升
Table 2 Accuracy and relative accuracy improvement of the three methods in scenario B

方向	LS-IGG3 /m	LS-QUAD /m	RF-QUAD /m	精度提 升 1/%	精度提 升 2/%
E	7.84	7.96	6.55	16.5	17.7
N	7.78	7.89	6.11	21.5	22.6
U	29.80	28.00	14.39	51.7	48.6

为了直观解释 LS-IGG3 方法、LS-QUAD 方法和 RF-QUAD 方法在作用机理上的不同,图 13 所示为两个历元所有观测值的误差(残差)分布,图 13(a)、(b)属于同一个历元,图 13(c)、(d)属于另一个历元,横轴为观测值对应的卫星对,它们按照误差的大小排序,纵轴为观测值对应的误差大小。圆形虚线为通过坐标真值计算出的参考误差;三角形点划线为最小二乘后验残差,作为 LS-IGG3 方法的抗差标准,同时也是 LS-QUAD 方法确定初始拟准观测值的标准;方形实线为最终确定的拟准观测值所计算出的真误差,其中图 13(a)、(c)是通过后验残差

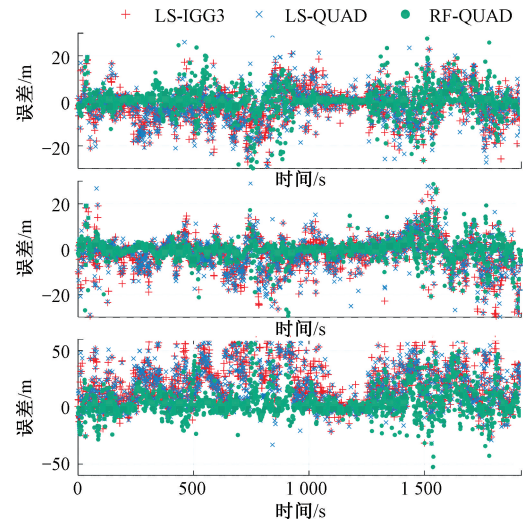
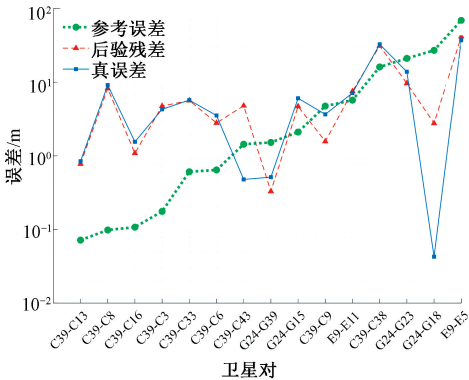
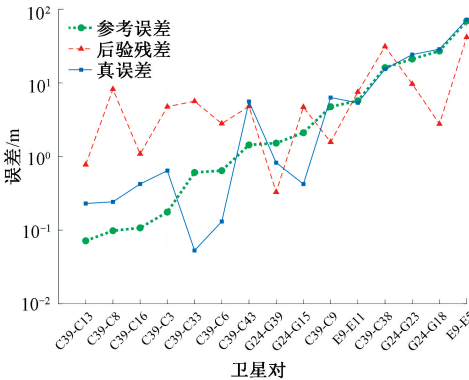


图 12 场景 B 3 种方法的误差序列

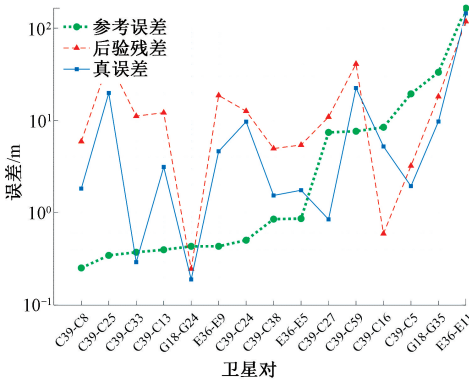
Fig. 12 Error sequence of three methods in scenario B



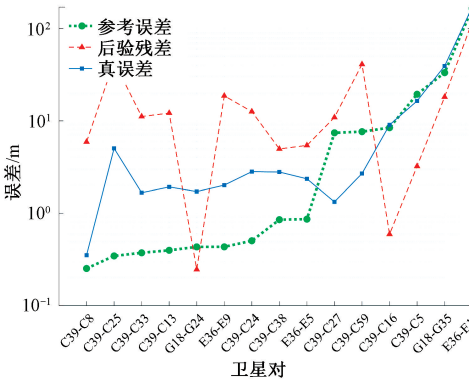
(a) 历元1: 后验残差确定拟准观测值
(a) Epoch 1: The quasi-accurate observations determined by posterior residuals



(b) 历元1: 随机森林确定拟准观测值
(b) Epoch 1: The quasi-accurate observations determined by random forest



(c) 历元2: 后验残差确定拟准观测值
(c) Epoch 2: The quasi-accurate observations determined by posterior residuals



(d) 历元2: 随机森林确定拟准观测值
(d) Epoch 2: The quasi-accurate observations determined by random forest

图 13 两个历元所有观测值的误差(残差)分布

Fig. 13 Error (residual) distribution of all observations in two epochs

确定的拟准观测值即 LS-QUAD 方法,图 13(b)、(d)是通过随机森林得分确定的拟准观测值即 RF-QUAD 方法。

理论上,QUAD 确定的真误差与参考误差越接近,利用 IGG3 模型抗差的效果就越显著。从图 13(a)可以看出,真误差和后验残差在左半段并没有明显的区别,而右半段的卫星对 G24~G18 参考误差为 27.14 m,真误差被缩小至 0.04 m,这表明此时 QUAD 并没有有效地分离粗差观测值,反而将粗差观测值误判为正常观测值;图 13(b)通过随机森林确定拟准观测值,与图 13(a)相比,真误差更符合参考误差从左到右逐渐增大的趋势,对于卫星对 G24~G18 也没有产生误判的情况,表明此时 QUAD 正确地将粗差观测值从正常观测值中分离出来。图 13(c)、(d)的情况与图 13(a)、(b)类似,当通过后验残差确定拟准观测值时,卫星对 C39~C25 参考误差为 0.35 m,真误差被放大到 19.85 m,卫星对 C39~C5 参考误差为 19.41 m,真误差被缩小至 1.94 m,这说明 QUAD 产生了相当程度的误判;而通过随机森林确定拟准观测值的方法则有效抑制了这种错误,真误差也更符合“左低右高”的走势。表 3 为上述两个历元后验残差、真误差与参考误差的平均差值,值越小说明 IGG3 模型的效果越显著。

表 3 观测值后验残差、真误差与参考误差的平均差值

Table 3 Average difference between the posterior residuals, true errors and reference errors (m)			
历元	LS-IGG3 后验残差	LS-QUAD 真误差	RF-QUAD 真误差
历元 1	7.47	7.65	1.26
历元 2	15.54	8.33	2.85

3 结 论

本文在传统的 QUAD 方法基础上,提出了基于随机森林识别的 QUAD 抗差滤波方法。实验分析了 GNSS 观测值误差与不同特征的关系并训练了一个随机森林分类模型,最终将模型应用到定位解算中。

1) 传统抗差方法基于最小二乘残差的大小对观测值进行评估,当出现误差转移时评估客观性会受到影响。本文利用观测数据的多种特征训练了一个随机森林分类器,引入了平差系统外的因素对观测值进行评价,解决了粗差容易干扰分类结果的问题。

2) 离散标签和连续误差之间的不对应,使得误差接近拟准阈值的特征不显著样本会影响分类模型的训练,降低分类模型的准确性。本文通过引入相关性因子以描述样本特征对于某一标签的显著性,介入分类模型的训

练过程,解决了离散标签与连续误差之间的映射问题。

3) 常规的 QUAD 方法在计算真误差时,通常使用 QUO 的原始权重,不会进行重新定权。本文为了充分利用随机森林分类器的信息,利用分类得分对 QUO 进行重新定权,确保每个观测值的随机模型符合其更真实的误差分布,提高了 QUAD 方法的可靠性。

采用实验对比了传统的抗差方法 (LS-IGG3 和 LS-QUAD) 与本文提出的抗差方法 (RF-QUAD), RF-QUAD 方法在不同场景的测试中均表现出了良好的自适应性,根据路况情况和方向在定位精度上有 16%~51% 的提升,误差明显小于传统方法。

本文提出的随机森林分类模型有效提高了拟准观测值的识别准确性,显著提升了抗差滤波的性能。研究结果表明,通过引入多特征信息,基于随机森林识别的 QUAD 抗差滤波方法可以克服传统抗差方法的局限性,实现更为可靠的定位。

参考文献

[1] WANG L, GROVES P D, ZIEBART M K. Multi-constellation GNSS performance evaluation for urban canyons using large virtual reality city models [J]. Journal of Navigation, 2012, 65(3):459-476.

[2] ZHANG ZH T, LI Y, HE X F, et al. A composite stochastic model considering the terrain topography for real-time GNSS monitoring in canyon environments [J]. Journal of Geodesy, 2022, 96:79.

[3] 李岚,朱锋,刘万科,等.城市分类场景的 GNSS 伪距随机模型构建及其定位性能分析[J/OL]. 武汉大学学报 (信息科学版),1-14[2024-12-19].
LI L, ZHU F, LIU W K, et al. GNSS pseudorange stochastic model for urban classification scenes and its positioning performance[J/OL]. Geomatics and Information Science of Wuhan University, 1-14[2024-12-19].

[4] 蔡亲青,朱锋,陈曦,等.高度角与信噪比混合的 GNSS 随机模型精化及其对 RTK 定位性能的影响[J]. 全球定位系统, 2023, 48(1):24-31.
CAI Q Q, ZHU F, CHEN X, et al. Refinement of GNSS stochastic model with mixed elevation angle and signal-to-noise ratio and its impact on RTK positioning performance[J]. GNSS World of China, 2023, 48(1):24-31.

[5] BAHADUR B, SCHÖN S. Improving the stochastic model for code pseudorange observations from Android smartphones[J]. GPS Solutions, 2024, 28(3):148.

[6] ZHANG P F, TU R, GAO Y P, et al. Improving the performance of multi-GNSS time and frequency transfer using robust helmert variance component estimation[J]. Sensors, 2018, 18(9): 2878.

- [7] YUAN H J, ZHANG ZH T, HE X F, et al. An Extended robust estimation method considering the multipath effects in GNSS real-time kinematic positioning [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71:8504509.
- [8] ZHAO J X, WANG J, HAN H Z, et al. A study on the model of robust fractional-order extended Kalman filtering with gross error[J]. GPS Solutions, 2024, 28(2):87.
- [9] ZHU H ZH, FAN J B, LI J, et al. Research on robust adaptive RTK positioning of low-cost smart terminals[J]. Sensors, 2024, 24(5):1477.
- [10] WANG R, BECKER D, HOBIGER T. Stochastic modeling with robust Kalman filter for real-time kinematic GPS single-frequency positioning [J]. GPS Solutions, 2023, 27(3):153.
- [11] YU X W, WANG J F. A method for separating unmodeled errors in GNSS based on the constraint of the prior variance of unit weight [J]. Measurement Science and Technology, 2023, 34(9):095109.
- [12] 陈熙源,周云川,钟雨露,等. 基于变分贝叶斯的鲁棒自适应因子图优化组合导航算法[J]. 仪器仪表学报, 2024, 45(1):120-129.
CHEN X Y, ZHOU Y CH, ZHONG Y L, et al. Robust adaptive factor graph optimization integrated navigation algorithm based on variational Bayesian [J]. Chinese Journal of Scientific Instrument, 2024, 45(1):120-129.
- [13] HSU L T, TOKURA H, KUBO N, et al. Multiple faulty GNSS measurement exclusion based on consistency check in urban canyons [J]. IEEE Sensors Journal, 2017, 17(6):1909-1917.
- [14] 张建,喻国荣,潘树国,等. 基于卡方检验的 GNSS 观测值部分粗差抗差滤波算法 [J]. 仪器仪表学报, 2019, 40(8):102-109.
ZHANG J, YU G R, PAN SH G, et al. Partial gross error robust filtering algorithm for GNSS observations based on chi-square test[J]. Chinese Journal of Scientific Instrument, 2019, 40(8):102-109.
- [15] 王鼎杰,孟德利,李朝阳,等. 抗野值自适应卫星/微惯性组合导航方法[J]. 仪器仪表学报, 2017, 38(12):2952-2958.
WANG D J, MENG D L, LI ZH Y, et al. Adaptively outlier-restrained GNSS/MEMS-INS integrated navigation method [J]. Chinese Journal of Scientific Instrument, 2017, 38(12):2952-2958.
- [16] 蒋晨. GNSS/INS 组合导航滤波算法及可靠性分析[J]. 测绘学报, 2020, 49(10):1376.
JIANG CH. Filtering algorithms and reliability analysis for GNSS/INS integrated navigation systems [J]. Acta Geodaetica et Cartographica Sinica, 2020, 49 (10): 1376.
- [17] 王东宇,张慧君,李孝辉,等. 遮挡环境下的基于 AKF 组合导航定位方法研究 [J]. 电子测量与仪器学报, 2023, 37(5):171-179.
WANG D Y, ZHANG H J, LI X H, et al. Research on positioning in covering environment with an AKF-based integrated navigation system [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37 (5): 171-179.
- [18] WANG H, PAN SH G, GAO W, et al. Multipath/NLOS detection based on k-means clustering for GNSS/INS tightly coupled system in urban areas [J]. Micro-machines, 2022, 13(7):1128.
- [19] 夏炎,潘树国,赵鹏飞,等. 基于无监督学习的卫星 NLOS 信号检测方法 [J]. 东南大学学报(自然科学版), 2019, 49(3):565-572.
XIA Y, PAN SH G, ZHAO P F, et al. Unsupervised learning based satellite NLOS signal detection method [J]. Journal of Southeast University (Natural Science Edition), 2019, 49(3):565-572.
- [20] YOZEVITCH R, MOSHE B B, WEISSMAN A. A robust GNSS los/nlos signal classifier [J]. Navigation: Journal of the Institute of Navigation, 2016, 63(4):429-442.
- [21] SOCHAROENTUM M, KARIMI H A, DENG Y. A machine learning approach to detect non-line-of-sight GNSS signals in Nav2Nav [C]. 23rd ITS World Congress, 2016:10-14.
- [22] HSU L T. GNSS multipath detection using a machine learning approach [C]. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2017:1-6.
- [23] ZHENG F, LI Q CH, WANG J L, et al. GNSS NLOS detection method based on stacking ensemble learning and applications in smartphones [J]. GPS Solutions, 2024, 28(3):129.
- [24] ZHU F, LUO K G, TAO X L, et al. Deep learning based vehicle-mounted environmental context awareness via GNSS signal [J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(8):9498-9511.
- [25] 王玉泽. 面向复杂城市环境的卫星导航信号特征统计分析研究 [D]. 上海:上海交通大学, 2019.
WANG Y Z. Statistical anlysis of GNSS signal characteristics for complex urban environment [D]. Shanghai:Shanghai Jiao Tong University, 2019.
- [26] OU J K. A new method of identifying and locating gross errors: Quasi-accurate detection [J]. Chinese Science Bulletin, 1999, 44(23):2200-2204.

- [27] 黄立人. 用于相对稳定点组判别的 QUAD 法[J]. 大地测量与地球动力学, 2002, 22(2):10-15.
HUANG L R. Quad method used for identifying relatively stable stations[J]. Journal of Geodesy and Geodynamics, 2002, 22(2):10-15.
- [28] LUO X W, WANG H T. Robust adaptive Kalman filtering-a method based on quasi-accurate detection and plant noise variance-covariance matrix tuning[J]. Journal of Navigation, 2017, 70(1):137-148.
- [29] 赵俊,郭飞霄. 利用 L1 范数和中位数选取拟准观测值[J]. 武汉大学学报(信息科学版), 2018, 43(8):1242-1248.
ZHAO J, GUO F X. Selection of quasi-observations using L1 norm and median[J]. Geomatics and Information Science of Wuhan University, 2018, 43(8):1242-1248.
- [30] QU W, CHEN H L, ZHANG Q, et al. A robust estimation algorithm for the increasing breakdown point based on quasi-accurate detection and its application to parameter estimation of the GNSS crustal deformation model[J]. Journal of Geodesy, 2021, 95(11):125.
- [31] ZOU ZH Y, JIANG Z S, WU Y Q, et al. Dynamic characteristics of crustal movement in north-south seismic belt from GPS velocity field before and after the Wenchuan Earthquake[J]. Chinese Journal of Geophysics, 2015, 58(5):1597-1609.
- [32] 周江文. 经典误差理论与抗差估计[J]. 测绘学报, 1989(2):115-120.
ZHOU J W. Classical error theory and robust estimation[J]. Acta Geodaetica et Cartographica Sinica, 1989(2):115-120.

作者简介



高旺(通信作者),2013 年于东南大学获得学士学位,2018 年于东南大学获得博士学位,现为东南大学副教授,主要研究方向为多传感器融合高精度导航定位。

E-mail:gaow@seu.edu.cn

Gao Wang(Corresponding author) received his B. Sc. degree from Southeast University in 2013, and Ph. D. degree from Southeast University in 2018. Now he is an associate professor at Southeast University. His main research interests include multi-sensors integrated precise navigation and positioning.



刘裕荐,2022 年于吉林大学获得学士学位,现为东南大学硕士研究生,主要研究方向为复杂环境 RTK 定位。

E-mail:501771460@qq.com

Liu Yujian received his B. Sc. degree from Jilin University in 2022. Now he is a master student at Southeast University. His main research interest includes RTK positioning in complex environments.



郭雅娟,1997 年于华北电力大学获得学士学位,2002 年于河海大学获得硕士学位,现为国网江苏省电力有限公司电力科学研究院研究员级高级工程师,主要研究方向为电力北斗技术及应用。

E-mail:mandy_guo@163.com

Guo Yajuan received her B. Sc. degree from North China Electric Power University in 1997, and M. Sc. degree from Hohai University in 2002. Now she is a senior researcher and principal engineer at the Research Institute, State Grid Jiangsu Electric Power Co., Ltd. Her main research interest includes power Beidou technology.