

蒋弥, 丁晓利, 李志伟. 2018. 时序 InSAR 同质样本选取算法研究. 地球物理学报, 61(12):4767-4776, doi:10.6038/cjg2018L0490.
Jiang M, Ding X L, Li Z W. 2018. Homogeneous pixel selection algorithm for multitemporal InSAR. *Chinese J. Geophys.* (in Chinese), 61(12):4767-4776, doi:10.6038/cjg2018L0490.

时序 InSAR 同质样本选取算法研究

蒋弥^{1,2}, 丁晓利³, 李志伟⁴

1 河海大学水文水资源与水利工程科学国家重点实验室, 南京 210098

2 河海大学地球科学与工程学院, 南京 211100

3 香港理工大学土地与地理资讯学系, 九龙 香港

4 中南大学地球科学与信息物理学院, 长沙 410083

摘要 分布式雷达目标时序 InSAR 技术是目前 InSAR 形变监测领域的主流方向, 其中同质样本选取是该技术的基础, 其估计精度直接影响 SAR 影像分辨率与后续参数解算精度. 本文在追踪最新研究进展之上, 系统回顾了当今统计同质选点算法的优缺点. 在参数与非参数两类统计方法的应用中, 采用蒙特卡罗方法和真实数据验证定量比较算法差异以及适用场景. 根据之前的研究结论, 提出一种改进的最优参数统计同质样本选择方法. 最后, 论文介绍了团队研发的 MATLAB 开源工具包, 涵盖了同质样本提取和时序 InSAR 协方差矩阵估计两部分内容, 为 InSAR 科研人员和后续数据处理提供高质量、全分辨率的观测源.

关键词 分布式目标; 时序 InSAR; 同质样本选择; 协方差矩阵估计; 开源工具包

doi:10.6038/cjg2018L0490

中图分类号 P228

收稿日期 2017-08-06, 2017-12-27 收修定稿

Homogeneous pixel selection algorithm for multitemporal InSAR

JIANG Mi^{1,2}, DING XiaoLi³, LI ZhiWei⁴

1 State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

2 School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China

3 The Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University,

Hung Hom, Hong Kong, China

4 School of Geoscience and Info-Physics, Central South University, Changsha 410083, China

Abstract Synthetic Aperture Radar Interferometry (InSAR) technique for distributed scatterers has become one of the current flavors. Inaccurate estimation of the covariance matrix is regarded as the most important source of error in such applications. The previous studies, named statistically homogeneous pixel selection algorithms, have demonstrated their values to refine the estimate accuracy for each target without loss of image resolution. In this paper, we review these methods globally under parametric and non-parametric statistical framework. The updated one will then be presented by integrating the advantages of the parametric statistics, and evaluated through synthetic and real data. Based on these algorithms, we finally introduce our SHPS-InSAR open-source toolbox, designed for homogeneous pixel selection and covariance matrix estimation. The toolbox includes almost all methods presented in this paper and tries to provide the optimal observable for the researchers in the field.

基金项目 国家自然科学基金(41774003), 江苏省自然科学基金(BK20171432), 中央高校基本科研业务费专项资金资助(2018B17714), 空间信息智能感知与服务深圳市重点实验室(深圳大学)开放基金资助项目(20165006212), 中欧龙计划项目第四期(32248_2)资助.

第一作者简介 蒋弥, 男, 1982 年生, 副教授, 主要从事 InSAR 研究. E-mail: mijiang@hhu.edu.cn

Keywords Distributed targets; Multi-temporal InSAR; Homogeneous pixel selection; Covariance matrix estimation; Open-source toolbox

0 引言

自 2011 年 Ferretti 等发布第二代永久散射体技术 SqueeSAR 以来,分布式散射目标(Distributed scatterer, DS)的数据处理便逐渐成为国际时序 InSAR 领域的研究热点之一. 相对于永久散射体(Permanent Scatterers, PS), DS 是指在雷达分辨率单元内没有任何散射体的后向散射占据统治地位的点目标(Hooper, 2008; Lee and Pottier, 2009). 由于地表特征如植被、水体、裸地、低强度不透水面等均同属分布式散射机制, DS 目标在雷达成像场景中一般占具主导地位, 因而采用 DS 点增加观测量是提高 InSAR 形变产品空间分辨率的有效途径之一.

与 PS 目标的物理属性不同, DS 易遭受时空去相关等因素影响, 干涉质量参层不齐. 因此在形变解算之前, 需要对 DS 点目标进行筛选和优化处理以提高观测信噪比、减小误差传播概率. DS 数据处理主要分为两个步骤: 同质点选取和时序 DS 优化. 前者是后者的保障, 也是 DS-InSAR 技术的核心(Ferretti et al., 2011; Jiang et al., 2015).

同质点选取, 又名同统计分布样本选择, 是以统计推断为手段度量邻域像素与中心像素相似度的一种算法(Jiang et al., 2015; Parizzi and Breic, 2011). 在同类地物目标具有相同后向散射属性和相同相位散射中心的假设前提之下, 将同类像素汇集进行参数估计, 可以实现信噪比的提升并兼顾图像分辨率,

特别适用于复杂场景的精细形变监测.

本文将着重讨论同质样本选择的数学原理, 并系统回顾目前国际上的主流算法. 用参数和非参数统计推断将这些算法进行划分, 定量评价它们在不同假设前提条件下的功效. 在改进现有方法基础上, 给出一种优化参数统计方法. 采用蒙特卡罗随机试验和真实数据集, 证明修正方法在收敛速度、综合类型 I/II 误差和方差三个方面均优于现有研究. 考虑到现行数据处理软件在 DS-InSAR 技术上的欠缺以及难以用于精细场景形变监测, 论文最后介绍了本团队开发的基于同质点选取算法的 SHPS-InSAR 开源工具包. 其设计主旨是在保持雷达图像分辨率的基础上最大程度提高多维复协方差矩阵的信噪比, 为 InSAR 科技工作者和后续数据处理提供高质量的 SAR 强度、干涉相干性和干涉相位观测源.

1 同质样本选取算法

从图像处理的角度来说, 同质样本选择属于二分类问题, 目的是将矩形窗口内的像素分开, 其中与参考像素具有相同值的一类称为前景, 剩余像素称为背景. 例如图 1 中, 交叉路口的红色点为参考像素, (b) 图中的绿色点是形成的矩形窗口, 算法需要对所有绿色样本进行筛选, 提取出路面上的点目标.

然而, 该问题不等价于彩色图像或光学遥感图像的分类问题(Li et al., 2016; Rother et al.,

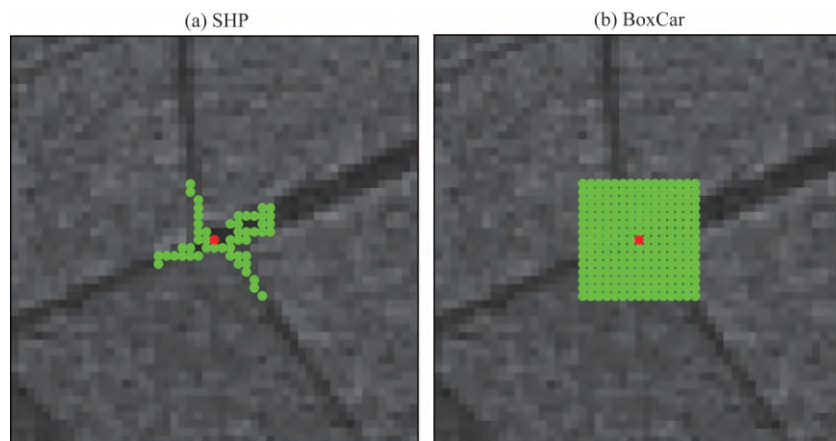


图 1 同质样本选取案例

(a) 同质点选取算法; (b) 传统规则窗方法.

Fig. 1 The case for pixels selection

(a) Homogeneous pixel selection algorithm (SHP); (b) Conventional boxcar Estimation.

2004). 由于 SAR 数据信噪比低, 噪声分布特征复杂, 要从单景 SAR 影像或直接移植计算视觉领域的算法提取出(a)中的样本十分困难, 因而学者们提出采用数据堆栈的方法, 从时序三维 SAR 数据中恢复同质点. 其基本思想是, 同时抽取参考像素与邻域内某位置像素的第三维时间样本进行比较, 度量它们的相似程度. 在双样本双边假设检验的框架下, 若检验统计量的估值落在拒绝域内, 则推翻假设, 否则接受假设, 即待估点与参考点在统计上无显著差异. 依照此法, 逐个比较中心像素与矩形窗口内所有空间位置的像素相似度, 最终确立同质目标.

这类方法的争论点在于选择何种统计检验方法可以达到最优估计. 所谓最优, 是指在给定显著水平下, 检验功效最大化(即类型 II 误差最小). 对于本文研究的问题, 就是保证选取的集合中同质样本最大化, 异质样本最小化.

然而(双样本)检验方法不下数十种, 根据对象不同判决差异很大. 例如, 在检验分布位置上(如均值), T 检验(Student's t test)可能获得最优; 在检验分布尺度上(如方差), F 检验可能胜出; 在检验分布形态上(如分布形状), 似然比检验是有力竞争者(Jiang et al., 2014; Kössler, 2010). 检验统计对分布特征十分敏感, 在非高斯条件下, T 检验和 F 检验的功效很低, 这时 Lepage 类型、Kolmogorov-Smirnov (KS) 等非参数检验更可取. 即便如此, 在非参数统计家族之中, 分布又因为偏度、峰度和尾重的不同, 检验功效差异也很大(Hogg et al., 1975; Kössler, 2010).

雷达单视复数数据常假设满足复圆高斯分布(Jiang et al., 2015), 因而其 SAR 强度服从指数分布, 幅度服从瑞利分布. 对于时间 SAR 数据集, 在斑点噪声不相关或时间轴不发生变异的前提下, 第三维样本也被假设满足上述分布. 依照功效最大化原则, 可将算法分为两类: 以 KS 检验为代表的非参数家族和以似然比检验为代表的参数家族. 检验问题被叙述成: 零假设 $H_0: F_{\text{ref}} = F_{\text{neig}}$, 备择假设 $H_1: F_{\text{ref}} \neq F_{\text{neig}}$. 其中零假设表示参考像素的时间样本分布特征 F_{ref} 与邻域内某像素的时间样本分布特征 F_{neig} 相等. 因为指数分布和瑞利分布均只包含一个参数, 所以对于参数统计推断, 参数相同等价于分布相同.

以下将着重介绍三种具有代表性的假设检验, 以及这类算法的备择方案置信区间. 最后提出一种改进的最优参数统计方法. 若无特殊指明, 对于 N 景 SAR 图像, 每个空间像素的时间样本数为 N .

1.1 非参数统计推断

1.1.1 Kolmogorov-Smirnov 检验

KS 是主流的优度拟合检验方法(Papoulis and Pillai, 2002). 在双样本条件下, KS 检验基于累计分布函数, 用以检验经验分布是否相同. 其检验统计 D 在量化两组样本的经验分布 \hat{F}_{ref} 和 \hat{F}_{neig} 的最大距离之后, 给出判决.

$$D(x) = \sup_x |\hat{F}_{\text{ref}}(x) - \hat{F}_{\text{neig}}(x)|, \quad (1)$$

其中 \sup 表示上确界(若极大值存在, 等效于上确界). 在给定显著水平 α 下, 拒绝域的形式为:

$$D > c_\alpha \sqrt{\frac{2}{N}}, \quad (2)$$

c_α 表示 Kolmogorov 分布的 α 分位点, 一般表示为:

$$c_\alpha = \sqrt{-\frac{1}{2} \ln\left(\frac{\alpha}{2}\right)}. \quad (3)$$

KS 检验的优点是统计量的抽样分布已知, 拒绝域可以用解析表达式给出, 因而广泛用于评价样本相似度. 主要缺点是统计差异倾向于经验分布的中心, 因而对于尾部的差异不敏感. 其次, KS 检验在小样本下的功效更低, 导致对于存在统计差异的样本, 易出现无法推翻零假设的情况. 相比之下, Anderson-Darling 检验、Cramer Von-Mises 检验和 Baumgartner-Wei-Schindler(BWS)检验等注重分布尾部差异, 因而备受瞩目(Anderson, 1962; Baumgartner et al., 1998).

1.1.2 Baumgartner-Wei-Schindler 检验

BWS 检验属于秩和检验方法, 是 KS 检验的加权版本. 其基本思想是对经验分布差的平方进行加权, 而权的确定依赖于真实分布. 考虑到真实分布是未知的, 检验统计量 B 用秩近似加权,

$$\begin{aligned} B_{\text{ref}} &= \frac{1}{2N^2} \sum_{i=1}^N \frac{(R_i - 2i)^2}{\frac{i}{N+1} \cdot \left(1 - \frac{i}{N+1}\right)}, \\ B_{\text{neig}} &= \frac{1}{2N^2} \sum_{i=1}^N \frac{(H_i - 2i)^2}{\frac{i}{N+1} \cdot \left(1 - \frac{i}{N+1}\right)}, \\ B &= \frac{B_{\text{ref}} + B_{\text{neig}}}{2}, \end{aligned} \quad (4)$$

其中 R_i 和 H_i 分别表示合并样本中第一组样本和第二组样本的秩(Baumgartner et al., 1998).

BWS 检验的优点是考虑样本分布的尾重, 进而对于尾部差异更加敏感, 在小样本条件下获得更高功效. 主要缺点是检验统计量的渐近分布无解析表达式, 进而导致拒绝域不能以函数形式给出. 一般需采用蒙特卡罗模拟实验获得数值解.

1.2 参数统计推断

1.2.1 似然比检验

假设 SAR 幅度服从瑞利分布(Lee and Pottier, 2009),

$$f(x) = \frac{2x}{\theta} e^{-\frac{x^2}{\theta}}, x \geq 0, \quad (5)$$

在似然比框架下,可以检验 $H_0: \theta_{\text{ref}} = \theta_{\text{neig}} = \theta$ 以及 $H_1: \theta_{\text{ref}} \neq \theta_{\text{neig}}$. 其中 θ_{ref} 和 θ_{neig} 分别表示参考像素和邻域某像素时间样本的总体分布参数. 前已述及,因为瑞利分布仅与一个参数有关,所以这里的假设叙述与上述分布是否相同一致.

奈曼-皮尔逊准则证明,在相同显著性水平下似然比检验是所有检验中功效最高的检验(Papoulis and Pillai, 2002),因而 H_0 为假拒绝 H_0 的概率最高,同质集合中的异质性最小.

在双样本条件下,检验统计 λ 可以写为(Papoulis and Pillai, 2002),

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Lambda})}, \quad (6)$$

其中 $L(\cdot)$ 表示两个分布的联合似然函数. 在零假设下, $\hat{\omega} = \{(\hat{\theta}, \hat{\theta}) : \hat{\theta} > 0\}$, $\hat{\theta}$ 是 θ 的似然估计. 在备择假设下, $\hat{\Lambda} = \{(\hat{\theta}_{\text{ref}}, \hat{\theta}_{\text{neig}}) : \hat{\theta}_{\text{ref}} > 0, \hat{\theta}_{\text{neig}} > 0\}$, $\hat{\theta}_{\text{ref}}$ 和 $\hat{\theta}_{\text{neig}}$ 分别是参考像素和邻域像素各自分布参数的似然估计. $L(\cdot)$ 在各自条件下取极值,略去推演过程,可以得到,

$$\Lambda = -2\ln(\lambda) = 4N\ln(\hat{\theta}) - 2\ln(\hat{\theta}_{\text{ref}}) - 2\ln(\hat{\theta}_{\text{neig}}), \quad (7)$$

$$\begin{aligned} \hat{\theta}_{\text{ref}} &= \frac{1}{N} \sum_{i=1}^N x_{\text{ref}}^2, \\ \hat{\theta}_{\text{neig}} &= \frac{1}{N} \sum_{i=1}^N x_{\text{neig}}^2, \\ \hat{\theta} &= \frac{\hat{\theta}_{\text{ref}} + \hat{\theta}_{\text{neig}}}{2}, \end{aligned} \quad (8)$$

其中 x_{ref} 和 x_{neig} 分别为各自样本的一个观测. 在大样本下, Λ 近似靠近卡方分布,因此可以获得单边临界值(Papoulis and Pillai, 2002). 然而在中小样本下, Λ 无分布描述,拒绝域需要通过蒙特卡罗模拟实验确定.

1.2.2 FaSHPS 算法

在之前的研究中,笔者研制了 FaSHPS 算法,在高斯假设下用置信区间代替假设检验提高运算速度(Jiang et al., 2015). 其基本思想是将三维数据在时间维进行平均,将平均后的参考像素值作为真值,邻域像素值作为待估值. 在给定的置信水平下,当这些待估值落入真值构筑的区间内,就认为它们是参考像素的同质样本.

这里运用了瑞利分布的变异系数是恒量的性质,可以获得区间解析式:

$$P\left\{\hat{\mu}_{\text{ref}} - z_{1-\alpha/2} \cdot 0.52 \cdot \hat{\mu}_{\text{ref}} / \sqrt{N \cdot L} < \hat{\mu}_{\text{neig}} < \hat{\mu}_{\text{ref}} + z_{1-\alpha/2} \cdot 0.52 \cdot \hat{\mu}_{\text{ref}} / \sqrt{N \cdot L}\right\} = 1 - \alpha, \quad (9)$$

其中 $P\{\cdot\}$ 表示概率, L 表示视数, $\hat{\mu}_{\text{ref}}$ 和 $\hat{\mu}_{\text{neig}}$ 分别表示参考像素和任意邻域像素的时间均值, $z_{1-\alpha/2}$ 表示标准正态分位点. 为减少异质点,算法先设置 $\alpha = 50\%$ 提炼中心像素均值,再设置 $\alpha = 5\%$ 进行选点.

FaSHPS 算法的优点是速度快,运用时间平均观测方差是原来的 $1/(NL)$ 倍的性质,将区间范围缩小,减小类型 II 误差,在小样本情形下估计更可靠. 然而,当时间平均 $\hat{\mu}_{\text{ref}}$ 靠近高斯分布的尾部,易引起区间偏离真实区间并无法收敛. 这种情况将导致类型 I 误差增加,真实 α 值大于给定水平,使得同质集合中同质样本的数量减少.

1.2.3 修正的参数统计方法

以上参数统计都有各自的优缺点,若将它们各自的优点融合,就能得到更好的估计量. 置信区间存在的问题是参考像素真值的估计是有偏的. 然而,似然比检验的估计是无偏的,能够控制类型 I 误差. 因此,可以将似然比检验的同质集合作为初值,提炼参考像素后再采用更窄的区间和高质量的观测迭代控制类型 II 误差,将样本异质性降至最小.

算法第一步根据 1.2.1 节中公式(7)和(8)比较参考像素和邻域像素的相似度. 为减小类型 II 误差,初始窗口应该尽量小,或者考虑邻域空间关系,只取与参考点直接或间接相邻的像素作为同质点. 在第 3 节介绍的 SHPS-InSAR 工具包中,定义了 7×7 的估计窗口,用于获取初始集合 Ω_{mit} .

接着,将集合内所有像素的平均得到参考像素的估计 $\hat{\mu}_{\text{ref}}$,再根据公式(9)就可以选取出同质样本. 注意在中小样本下,高斯假设不能逼近真实分布. 这里根据 SAR 强度服从指数分布的性质,其累加和服从伽玛分布 G ,去除高斯假设限制,因而可以获得更精确的区间(Jiang et al., 2017a).

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, x \geq 0 \quad (10)$$

指数分布的期望和标准偏差均为 θ . 假设一组来自指数分布的样本 $x_i, i = 1, \dots, N$, 其累加和服从 $\sum x_i \sim G(N, \theta)$, 则其 N 倍样本均值 $\hat{\mu}$ 也服从 $N\hat{\mu} \sim G(N, \theta)$, 令 $X = N\hat{\mu}/\theta$, 根据伽玛函数性质知 X 服从尺度参数为 N 的标准伽玛分布, $X \sim G(N, 1)$. 由于标准伽玛分布的 α 分位点 $g_{\alpha/2, N}$ 和样本数 N 已知,

则在给定置信水平下区间固定, 经过变换可以获得,

$$P\{g_{\alpha/2;N} \cdot \theta/N < \hat{\mu} < g_{1-\alpha/2;N} \cdot \theta/N\} = 1 - \alpha \quad (11)$$

伽玛分布是正偏态分布, 区间取值非对称, 注重了尾部. 在本文研究的问题中, 式(11)的样本均值 $\hat{\mu}$ 由邻域待估像素的时间均值 $\hat{\mu}_{neig}$ 代替, θ 由参考像素估值 $\hat{\mu}_{ref}$ 代替, 它是初始集合 Ω_{mit} 内所有三维样本的平均. 可以看出, 区间的研究对象是均值而不是原始样本, 因此能提高算法精度.

表 1 同质样本选取算法性能比较
Table 1 Capabilities of different methods for statistically homogeneous pixel selection

属性	非参数统计		参数统计		
	KS 检验	BWS 检验	似然比 检验	FaSHPS	新方法
计算效率	低	低	中	高	中
支持多视数据	是	是	否	是	否
类型 I 误差	小	小	小	大	小
类型 II 误差	大	较大	中	小	小
检验功效	低	较低	中	高	高
假设条件	无	无	复圆高斯 分布	高斯 分布	复圆高斯 分布
拒绝域解析式 (中小样本)	有	无	无	有	不完全

表 1 归纳了上述各类方法的性能, 这些指标仅表示相对关系, 将在下一节和 SHPS-InSAR 工具包中得到证明.

2 结果分析

2.1 模拟数据验证

算法验证从统计错误概率和假设条件出发, 本文采用蒙特卡罗随机试验检验上述五种方法. 在复

圆高斯分布假设前提下, 首先给定格网尺寸(设定格网尺寸为 11×11), 规定 11×6 格网像素的真值为 θ_1 , 其余 11×5 个真值为 θ_2 . 参考像素设为格网中心位置 $[6, 6]$, 其余 120 个像素作为待估像素. 接着, 在给定样本 $N=25$ 条件下, 分别从参数 θ_1 和 θ_2 的瑞利总体中模拟加噪幅度序列. 最后, 逐一用上述算法进行相似度比较, 统计出每种方法满足备择假设的百分比. 这一过程重复 10000 次, 可以获得各估计量的均值(偏差)和标准偏差. 随着 θ_1/θ_2 比值的增加, 最终得到不同对比度下的功效函数.

理论上, 被推翻的像素百分比等于 $(11 \times 5 + \alpha \times 11 \times 6) / 121$ (55 个参数为 θ_2 的异质像素加上给定水平 α 下 66 个同质像素的类型 I 误差). 取 $\alpha=0.05$, 得真实功效应为 48.18%.

从图 2a 的结果来看, 当总体参数对比度增加, 所有算法均能够收敛. 参数统计要明显优于非参数统计, 其中 FaSHPS 算法获得最高的功效, 但是与真实功效存在 5% 的偏差, 说明 FaSHPS 是有偏估计. 这是由于初值估计不稳定造成的 H_0 为真拒绝 H_0 , 导致类型 I 误差增加. 因此, 当 $\theta_1/\theta_2=1$ 时, $\alpha > 0.05$. 相比之下, 其余算法的第一类误差均维持在 $\alpha=0.05$ 附近. 其中新方法的效果最优, 它不但快速收敛至真实功效值, 而且在图 2b 中有最小标准偏差.

采用相同的模拟策略, 将复圆高斯假设替换为形状参数和尺度参数均为 1 的韦伯分布, 见图 3. 由于偏离原始假设, 参数统计量的功效不再具有实际价值, 因为任何估计的类型 I 误差都远大于 0.05. 以似然比检验为例, 在 $\theta_1/\theta_2=1$ 时, $\alpha=0.36$, 表示有 36% 的同质像素被排除, 导致同质集合中的样本

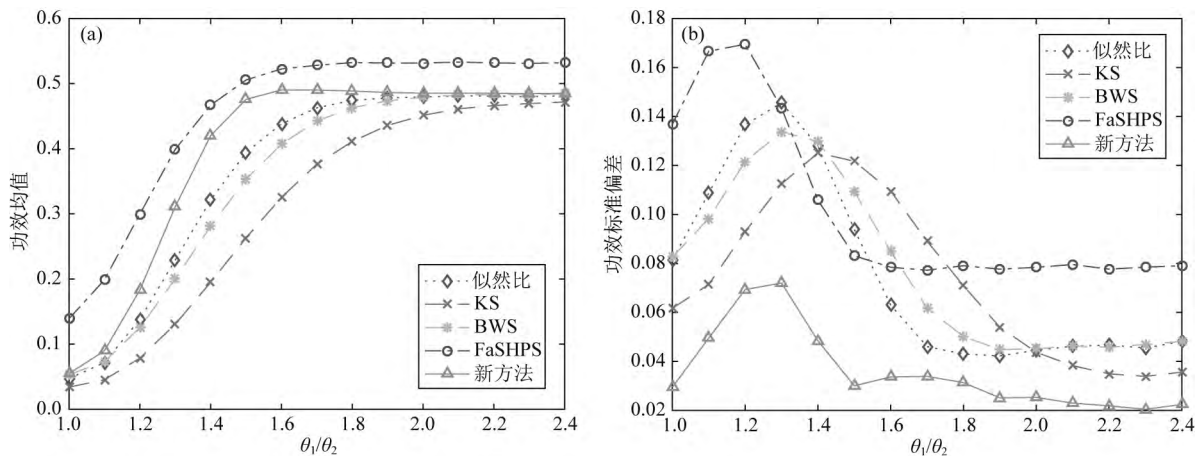


图 2 复圆高斯假设下, 同质样本选取算法模拟验证
(a) 功效均值; (b) 功效标准差.

Fig. 2 Monte Carlo simulation under complex (circular) Gaussian distribution
(a) The mean of power (SHP); (b) The stand deviation of power.

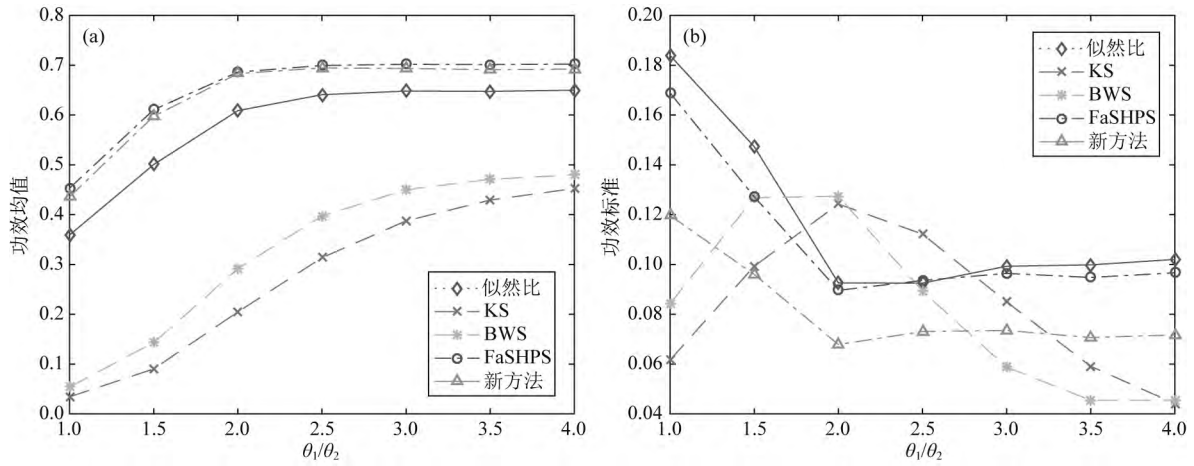


图 3 韦伯分布下,同质样本选取算法模拟验证

(a) 功效均值; (b) 功效标准差.

Fig. 3 Monte Carlo simulation under Weibull distribution

(a) The mean of power (SHP); (b) The stand deviation of power.

稀少. 样本数是后续 DS-InSAR 协方差估计的基础, 同质点的大量减少表明非稳健的参数估计. 相比而言, 非参数统计, 尤其是 BWS 检验能够准确收敛至真实功效值, 并且在图 3b 中表现出适中的标准偏差, 因而被认为是更好的估计量.

2.2 真实数据验证

真实数据源于 2006 年 12 月至 2011 年 2 月的 25 景 ALOS PALSAR-1 FBS/FBD 模式数据集 (Frame: 660, Path: 216). 研究区位于美国南加州洛杉矶的 Los Alamitos 军用机场, 覆盖面积约为 $2 \text{ km} \times 3.7 \text{ km}$. 在配准之后, 选取三类具有代表性的算法, BWS 检验、FaSHPS 和新方法进行测试, 定义邻域尺寸为 15×15 像素, $\alpha = 0.05$.

图 4 展示了每个像素邻域内选取的同质样本数量, 原始 SAR 特征详见图 6. 整体上, 机场周边的强散射体多属于 PS 目标, 因而选取的同质样本相对少. 反之, 中间的裸地是典型的 DS 目标, 同质区域大. 其中 BWS 检验选取的样本数最多, 新方法次之, FaSHPS 最少, 这与图 2a 的结论吻合. 由于 FaSHPS 拥有最高的功效, 更容易区分像元异质性, 但是需以同质样本减少为代价. BWS 功效低, 因此

同质集合中易包含异质样本, 样本数增加, 尤其是线状和点状地物包含了周围的异质样本, 使得整个机场框架不清晰. 比较 4(a) — (c) 也可以看出, 新方法最平滑, BWS 和 FaSHPS 显得更粗糙, 这也与图 2b 标准偏差的结论相吻合.

图 5 是图 4 更为直观的表述, 其中 (d) 和 (h) 来自图 4 中像素坐标 [81, 155] 和 [290, 140], 分别代表同质和异质区域. 以这两个点为参考点 (红), 评价不同算法选取的样本 (绿).

从目视上看, 当质地复杂时, BWS 检验 (a) 无法完全推翻那些来自于不同分布的样本, 因而在选择的同质集合中明显包含异质样本, 样本数增多. 前已述及, FaSHPS 和新方法在控制类型 II 误差上的优势, 因而 (b) 和 (c) 中能够排除异质样本, 结果更准确. 当研究区只包含一类样本时, BWS 检验和新方法类型 I 误差小, 因此在 (e) 和 (g) 中选择了大量样本, 而 FaSHPS 类型 I 误差更大, 所以在 (f) 中漏选了同质样本.

定量地, 将上述参考点所在估计窗口内 (15×15) 的所有像素先人工划分出同类样本, 然后根据估计结果统计不同方法的类型 I / II 误差于表 2. 在异

表 2 同质样本选取算法精度评估

Table 2 Accuracy evaluation of different methods for statistically homogeneous pixel selection

方法	异质区域 (参考点 [81 155])				同质区域 (参考点 [290 140])			
	类型 I 误差	类型 II 误差	探测概率	功效	类型 I 误差	类型 II 误差	探测概率	功效
BWS 检验	0	25%	100%	75%	12%	—	88%	—
FaSHPS	19%	1%	81%	99%	43%	—	57%	—
新方法	11%	2%	89%	98%	15%	—	85%	—

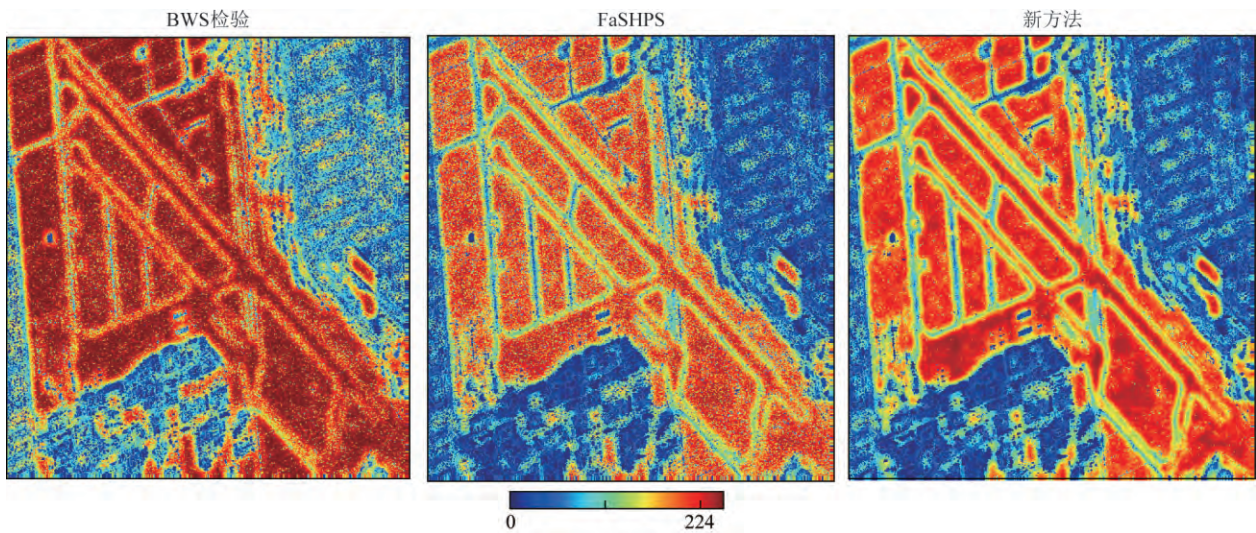


图 4 洛杉矶 Los Alamitos 军用机场不同算法下的同质样本选取数量图,其中颜色表示每个像素的同质样本数量.

Fig. 4 The number of homogeneous pixels over Los Alamitos Army Airfield, LA; The color denotes the SHP number for each pixel

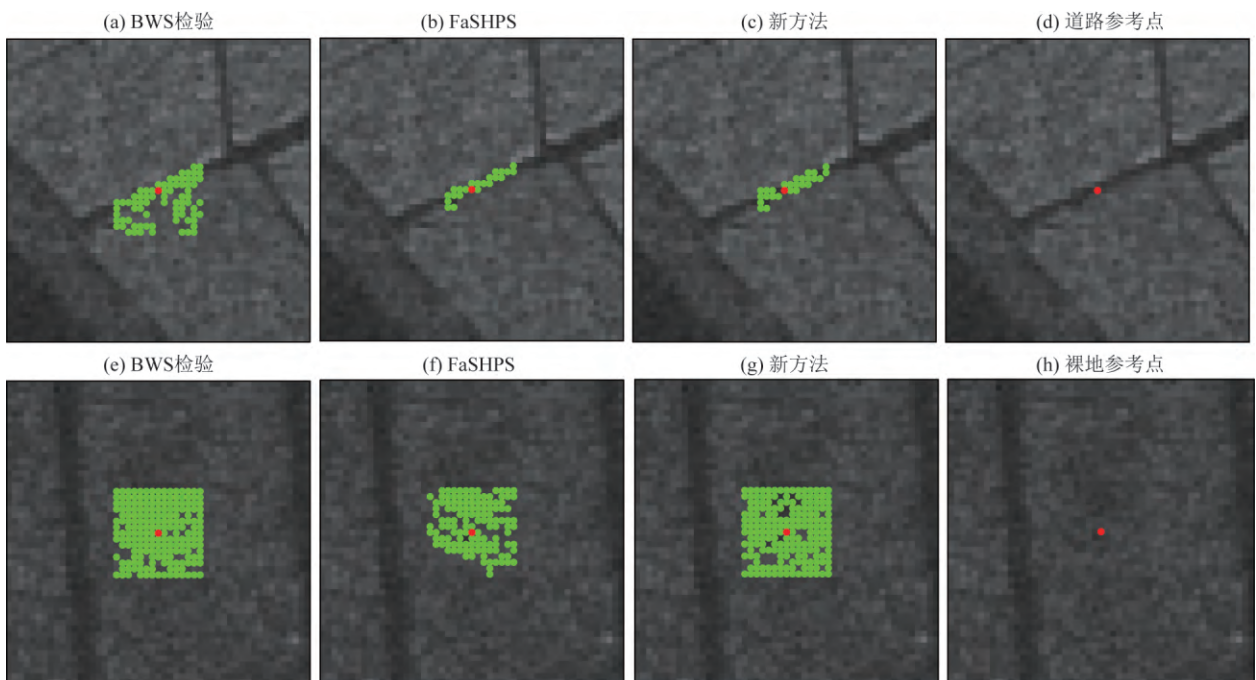


图 5 洛杉矶 Los Alamitos 军用机场子区域参考像素(红)及其同质样本(绿)示例

Fig. 5 The footprint of homogeneous pixels in Green for reference pixels in Red

质区域,虽然 BWS 的探测率达到 100%,但是以包含异质样本为代价,所以其功效最低,类型 II 误差最大.反之,FaSHPS 功效最高,但是以减少同质样本为代价,所以其探测率最低,类型 I 误差最大.同理,在同质区域,异质样本不存在,但是 FaSHPS 仍然排除许多样本,这些样本实际分布在高斯假设的尾部.相比之下,新方法在两个例子中的总误差分别为 13%和 15%,估计最为稳健.

3 SHPS-InSAR 开源工具包

根据前述理论和当今 DS-InSAR 发展的迫切需求,笔者开发了基于 MATLAB 平台的开源工具包,命名为 SHPS-InSAR (Statistically Homogeneous Pixel Selection InSAR).工具包主要由同质样本选取和协方差矩阵估计两个部分组成,其中前者是后

者的基础,后者是形变解算和地球物理应用的基础.与现行软件不同,工具包强调算法稳健性,经验参数少,无需人工干预和专家知识,可以从下列网址免费获取:<http://mijiang.org.cn/index.php/software/>. 本文的算法与实现,所使用的数据和模拟验证方法等均包含其中,供读者检验研究.

在采用本文算法获取同质集合之后,可以用协方差矩阵估计脚本获取 N 景 SAR 影像滤波的强度图、 $N(N-1)/2$ 景精确相干性估计图和干涉相位滤波图. 图 6—8 分别展示了工具包估计的高质量观测源,并与现行软件如 GAMMA 等处理结果进行了定性比较. 可以看出,基于同质选取算法的估计结果不仅能够维持原始 SAR 图像分辨率,还能同时抑制噪声,提升各同质区域的信噪比并确保点目标的信号不受周围 DS 信号污染. 相比之下,目前主流软件

多使用高斯加权或传统估计方法,造成分辨率损失、估计精度受限,因而难以用于精细场景的形变监测研究.

4 讨论与结论

(1) 本文在时序 DS-InSAR 技术背景下系统回顾了国际上主流的同质样本选取算法,并在修正前述算法的基础上给出了一种统计误差最小的参数统计方法. 从蒙特卡罗模拟实验和真实数据验证的结果来看,当 SAR 数据服从复圆高斯分布时,新方法在控制类型 I/II 误差、标准偏差和功效上能获得综合最优的结果. 基于同质样本选取算法,研究团队进而开发了 SHPS-InSAR 开源工具包,用于获取高精度时序 SAR 强度序列、干涉相位和相干性,为全分

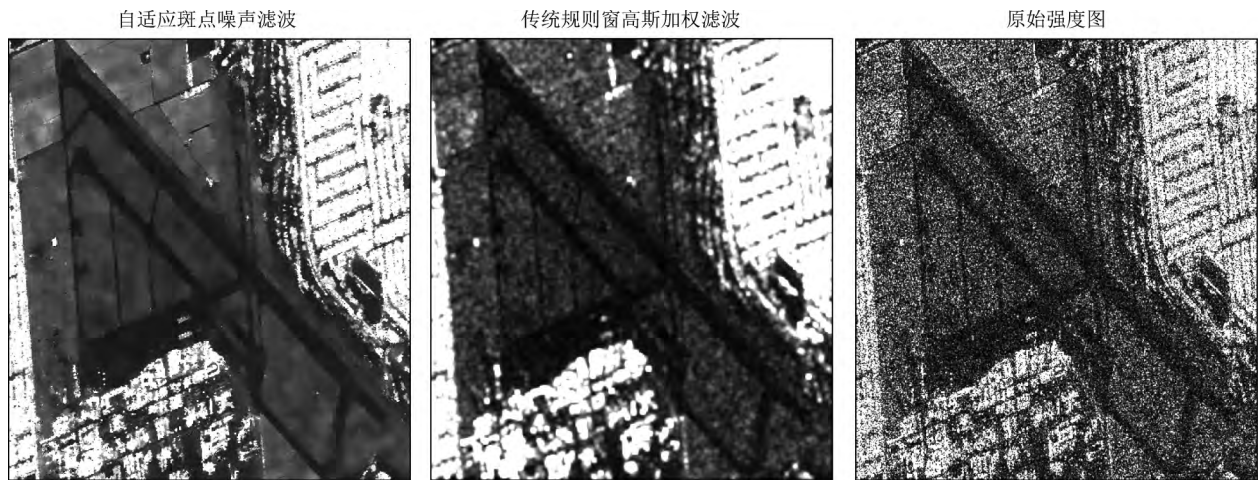


图 6 2006 年 12 月 31 日获取的 SAR 强度影像估计

Fig. 6 De-speckling for a SAR intensity image acquired from 20061231

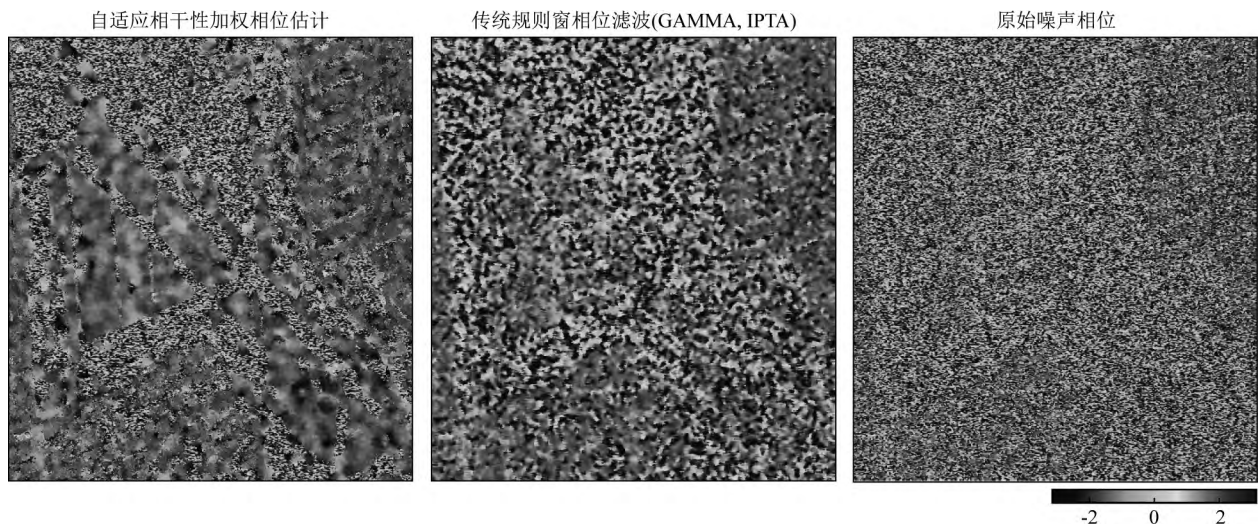


图 7 2009 年 8 月 23 日与 2009 年 10 月 8 日干涉对相位估计

Fig. 7 Phase filtering for interferometric pair 20090823 and 20091008

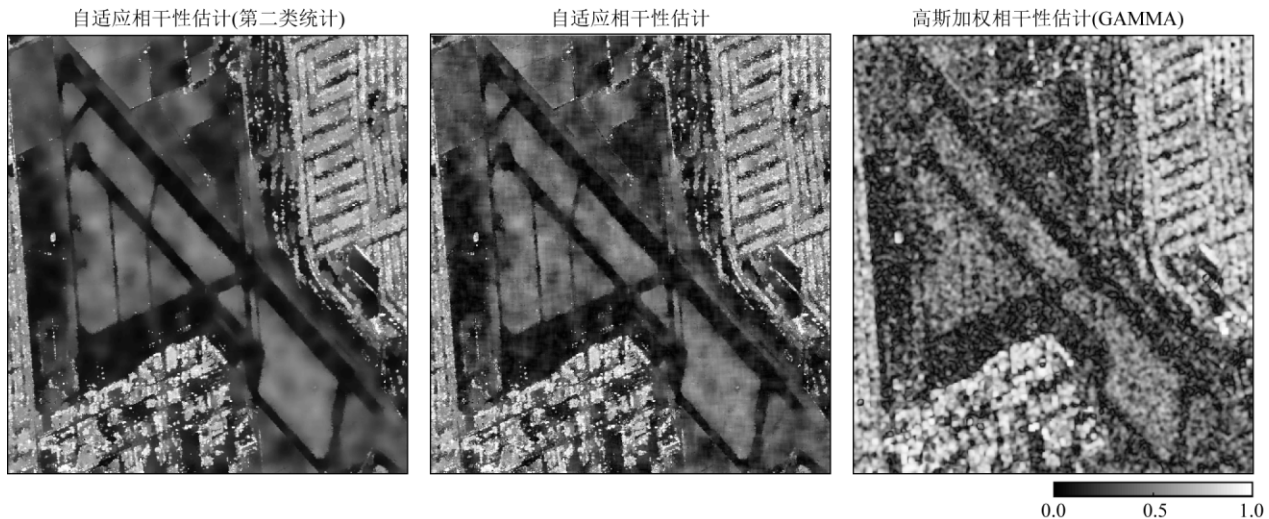


图 8 2010 年 2 月 23 日与 2010 年 4 月 10 日干涉对相干性估计

Fig. 8 Coherence estimation for interferometric pair 20100223 and 20100410

分辨率、精细化场景时序 InSAR 形变监测提供数据支持。

(2) 在计算效率方面,参数统计方法,如 FaSHPS 算法消耗内存最少,计算速度最快。以本文 $501 \times 457 \times 25$ 的三维数据为例, FaSHPS 仅耗时 13 s, 本文方法需 15 s, 而非参数统计方法因需要计算秩或经验分布,效率较低。以 BWS 检验为例,耗时约为 17 分钟。

在估计精度方面,算法选择需因地制宜,分以下三种情形讨论:(1) 当假设条件满足且在大中样本环境下,本文算法是最合适的方法;(2) 当假设条件不满足时,表示研究区有明显的时间变异,则应该采用非参数 BWS 检验方法;(3) 在小样本情形下,如 $N < 20$ 时,确保功率最大化是主要目标,因此应该推荐 FaSHPS 算法。

(3) 算法追求同质集合中同分布样本最大化,不同分布最小化的原则,在 InSAR 协方差矩阵估计时未必能获得最好的结果。受 SAR 传感器机理限制,汇集平均需以空间平均代替,所以过多的使用空间像素容易引起局部信号非平稳。FaSHPS 算法虽然更易推翻零假设,但可能回避了统计非平稳信号。一般地,只要一种算法保证同质集合中的样本数量足够获得精确的协方差矩阵估计,就不失为一种可取的方法。例如在相干性估计时,独立样本个数大于 30,就可得到近似无偏的估计结果(Jiang et al., 2014)。

(4) 最后,任何估计量在小样本条件下精度都会衰减,此时邻域对比度就显得更为重要了。从图 2a 中可知对比度高,异质样本可识别性就更强,算

法之间的性能差异就越小。然而,SAR 图像散射特征不一,所有算法在小样本下都有犯第二类错误的风险,因此同时考虑辐射值和邻域之间的上下文联系可能是解决该问题的最终方案。

致谢 感谢编辑部老师和审稿专家的帮助和支持。

References

- Anderson T W. 1962. On the distribution of the two-sample Cramer-von Mises criterion. *The Annals of Mathematical Statistics*, 33(3): 1148-1159.
- Baumgartner W, Weiß P, Schindler H. 1998. A nonparametric test for the general two-sample problem. *Biometrics*, 54(3): 1129-1135.
- Ferretti A, Fumagalli A, Novali F, et al. 2011. A new algorithm for processing interferometric data-stacks: SqueeSAR. *IEEE Transactions on Geoscience and Remote Sensing*, 49(9): 3460-3470.
- Goel K, Adam N. 2012. An advanced algorithm for deformation estimation in non-urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 73: 100-110.
- He Q, Yang H L, Fang Y, et al. 2017. Correlation analysis of surface deformation and groundwater level in Zhengzhou city. *Progress in Geophysics* (in Chinese), 32(6): 2636-2641, doi: 10.6038/pg20170646.
- Hogg R V, Fisher D M, Randles R H. 1975. A two-sample adaptive distribution-free test. *Journal of the American Statistical Association*, 70(351): 656-661.
- Hooper A. 2008. A multi-temporal InSAR method incorporating both persistent scatterer and small baseline approaches. *Geophysical Research Letters*, 35(16): L16302, doi: 10.1029/2008GL034654.
- Jiang M, Ding X L, Li Z W. 2014. Hybrid approach for unbiased coherence estimation for multitemporal InSAR. *IEEE Transactions*

- on *Geoscience and Remote Sensing*, 52(5): 2459-2473.
- Jiang M, Ding X L, Hanssen R F, et al. 2015. Fast statistically homogeneous pixel selection for covariance matrix estimation for multitemporal InSAR. *IEEE Transactions on Geoscience and Remote Sensing*, 53(3): 1213-1224.
- Jiang M, Miao Z L, Gamba P, et al. 2017a. Application of multitemporal InSAR covariance and information fusion to robust road extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 55(6): 3611-3622.
- Jiang M, Ding X L, He X F, et al. 2016. FaSHPS-InSAR technique for distributed scatterers: A case study over the lost hills oil field, California. *Chinese Journal of Geophysics* (in Chinese), 59(10): 3592-3603, doi: 10.6038/cjg20161007.
- Jiang M, Yong B, Tian X, et al. 2017b. The potential of more accurate InSAR covariance matrix estimation for land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 126: 120-128.
- Kössler W. 2010. Max-type rank tests, U -tests, and adaptive tests for the two-sample location problem—An asymptotic power study. *Computational Statistics & Data Analysis*, 54(9): 2053-2065.
- Lee J S, Pottier E. 2009. *Polarimetric Radar Imaging: From Basics to Applications*. Boca Raton: CRC Press.
- Li Z B, Shi W Z, Lu P, et al. 2016. Landslide mapping from aerial photographs using change detection-based Markov random field. *Remote Sensing of Environment*, 187: 76-90.
- Papoulis A, Pillai S U. 2002. *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill Europe.
- Parizzi A, Bricc R. 2011. Adaptive InSAR stack multilooking exploiting amplitude statistics: A comparison between different techniques and practical results. *IEEE Geoscience and Remote Sensing Letters*, 8(3): 441-445.
- Rother C, Kolmogorov V, Blake A. 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3): 309-314.
- Wang Y Y, Zhu X X. 2016. Robust estimators for multipass SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 54(2): 968-980.
- Zhu J J, Li Z W, Hu J. 2017. Research progress and methods of InSAR for deformation monitoring. *Acta Geodaetica et Cartographica Sinica* (in Chinese), 46(10): 1717-1733, doi: 10.11947/j. AGCS.2017.20170350.

附中文参考文献

- 何琪, 杨红磊, 方媛等. 2017. 郑州市地表形变与地下水位相关性分析. *地球物理学进展*, 32(6): 2636-2641, doi: 10.6038/pg20170646.
- 蒋弥, 丁晓利, 何秀凤等. 2016. 基于快速分布式目标探测的时序雷达干涉测量方法: 以 Lost Hills 油藏区为例. *地球物理学报*, 59(10): 3592-3603, doi: 10.6038/cjg20161007.
- 朱建军, 李志伟, 胡俊. 2017. InSAR 变形监测方法与研究进展. *测绘学报*, 46(10): 1717-1733, doi: 10.11947/j. AGCS.2017.20170350.