

12-12-2019

Multi-source Precipitation Data Fusion Method Based on Filtersim

Kunwei Li

Institute of Surveying and Mapping, Information Engineering University, Zhengzhou 450052, China;

You Xiong

Institute of Surveying and Mapping, Information Engineering University, Zhengzhou 450052, China;

Zhang Xin

Institute of Surveying and Mapping, Information Engineering University, Zhengzhou 450052, China;

Fen Tang

Institute of Surveying and Mapping, Information Engineering University, Zhengzhou 450052, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Multi-source Precipitation Data Fusion Method Based on Filtersim

Abstract

Abstract: Rainfall is an important factor affecting the vehicle off-road maneuver. High-precision rainfall data is a prerequisite for evaluating battlefield off-road traffic capacity quantitatively and developing a program of action reasonably. *In view of the shortcomings of traditional interpolation methods, the filtersim multi-point statistical method is used to improve the fusion accuracy of multi-source precipitation data. The rainfall is decomposed into the sum of the local mean and the local residual, the local residuals of the satellite precipitation data are used as the "training data", the local residual data of the weather stations are used as the "hard data", and the interpolation data of the local residual data of the satellite precipitation and the local residual data of the weather station are used as the "soft data" for filtersim multi-point statistical simulation to obtain 1 km resolution precipitation data.* By comparing the average absolute error, the root mean square error and the correlation coefficient of several sets of data, it is shown that filtersim is superior to the ordinary cooperative kriging interpolation, which can improve the accuracy of precipitation data effectively.

Keywords

precipitation data fusion, multi-point statistics, hard data, soft data

Recommended Citation

Li Kunwei, You Xiong, Zhang Xin, Tang Fen. Multi-source Precipitation Data Fusion Method Based on Filtersim[J]. Journal of System Simulation, 2019, 31(6): 1232-1238.

基于 Filtersim 的多源降水数据融合方法研究

李坤伟, 游雄, 张欣, 汤奋

(信息工程大学地理空间信息学院, 河南 郑州 450052)

摘要: 降雨是影响车辆越野机动的重要因素, 高精度的降雨数据是定量评估战场越野通行能力, 合理制定行动方案的前提。针对传统插值方法的不足, 论文采用 filtersim 多点统计法来提高降水数据的融合精度。将降雨量分解为局部均值与局部残差之和, 以卫星降水数据的局部残差作为“训练数据”, 以气象站点的局部残差数据作为“硬数据”, 以卫星降水的局部残差数据和气象站点的局部残差数据的插值数据作为“软数据”进行 filtersim 多点地统计模拟, 最终得到 1km 分辨率的降水数据。通过比较数据的平均绝对值误差、均方根误差和相关系数, 表明 filtersim 明显优于协同克里金插值, 能有效提高降水数据的精度。

关键词: 降水数据融合; 多点地统计; 硬数据; 软数据

中图分类号: E991 文献标识码: A 文章编号: 1004-731X (2019) 06-1232-07

DOI: 10.16182/j.issn1004731x.joss.17-0201

Multi-source Precipitation Data Fusion Method Based on Filtersim

Li Kunwei, You Xiong, Zhang Xin, Tang Fen

(Institute of Surveying and Mapping, Information Engineering University, Zhengzhou 450052, China)

Abstract: Rainfall is an important factor affecting the vehicle off-road maneuver. High-precision rainfall data is a prerequisite for evaluating battlefield off-road traffic capacity quantitatively and developing a program of action reasonably. In view of the shortcomings of traditional interpolation methods, the filtersim multi-point statistical method is used to improve the fusion accuracy of multi-source precipitation data. The rainfall is decomposed into the sum of the local mean and the local residual, the local residuals of the satellite precipitation data are used as the "training data", the local residual data of the weather stations are used as the "hard data", and the interpolation data of the local residual data of the satellite precipitation and the local residual data of the weather station are used as the "soft data" for filtersim multi-point statistical simulation to obtain 1 km resolution precipitation data. By comparing the average absolute error, the root mean square error and the correlation coefficient of several sets of data, it is shown that filtersim is superior to the ordinary cooperative kriging interpolation, which can improve the accuracy of precipitation data effectively.

Keywords: precipitation data fusion; multi-point statistics; hard data; soft data

引言

降雨是影响车辆越野机动的重要因素, 降雨改

变了地表的通行条件, 增加了越野通行的难度; 高精度的降雨数据是定量评估战场越野通行能力, 合理制定行动方案的前提^[1]。地面站观测能够准确地获取站点上的降水数据, 是获取降水信息最直接有效的方式。但是由于受到自然环境和成本的影响, 观测站的分布范围受到限制, 地面站点观测的降雨信息只能代表站点附近的降水, 无法准确反映降水



收稿日期: 2017-05-09 修回日期: 2017-07-07;
作者简介: 李坤伟(1988-), 男, 湖北仙桃, 博士生, 研究方向为战场环境仿真与作战模拟; 游雄(1962-), 男, 福建罗源, 博士, 教授, 研究方向为作战环境学; 张欣(1985-), 男, 河南夏邑, 博士, 讲师, 研究方向为作战环境建模。

<http://www.china-simulation.com>

• 1232 •

的空间分布。相比之下, 卫星遥感观测具有较大的空间覆盖范围, 能够连续、实时地对空间降水进行观测, 但是由于其是间接观测降雨量, 数据的精度还需要很大的改进。近年来将地面观测降水数据与卫星遥测降水产品融合成为估计降水空间分布的重要方向^[2]。

多源降雨数据融合的基本思想是首先构建降雨量的背景场, 然后用地面实测的雨量数据对背景场进行修正, 最终得到降雨估计场, 以此代表降雨的真实分布。常见的融合方法主要有克里金法^[3-4]、客观分析法^[5]、地理加权法^[6]、最优插值法^[7]、条件融合法^[8]等, 这些方法各有优势, 在不同的研究或业务工作中得到应用。克里金法是降水插值常用的方法, 在降水数据融合方面具有较大的优势^[9], 应用范围广泛。克里金法以变差函数为度量工具, 通过建立空间变量的变异性模型来推断空间变量的分布, 但是变差函数只能把握空间两点之间的相关性, 无法有效描述复杂的空间分布特征^[10]。同样, 最优插值法也只是描述两点之间的相关性, 无法整体描述多个点之间的相关性。

多点地统计法(Multiple-point Statistics, MPS)是地统计领域非常重要的方法^[11-12]。MPS 通过比较待估点周围的数据样式与训练图像中已存数据模式的相似性, 选择训练图像中最相似的数据模式整体或局部替换待估点周围的数据样式, 从而达到预测的目的; 相比克里金法, 多点地统计方法能够描述多点之间的相关性, 有效克服克里金法的不足, 广泛应用于储层建模领域; 近年来多点地统计在超分辨率图像重建领域也得到了一定的应用^[13-14]。降水在空间上复杂分布, 在时间上呈现非连续性特征, 鉴于现有插值方法的不足, 论文采用多点统计法来提高降水数据的插值精度。常用的多地统计方法主要有 snesim 和 flitersim, 由于 snesim 无法处理连续变量^[15], 因此, 论文主要采用 flitersim 对气象数据进行融合。

1 Flitersim 多点地统计方法原理

所谓多点地统计方法, 即模拟空间多个点联合分布的地统计方法。MPS 由 Farmer 和 Deutsch 等人在 1992 年提出, 由于受到计算机性能和算法的影响, 该方法一直停留在实验室阶段。随着搜索树概念的提出, MPS 得到快速发展, 自 2000 年斯坦福大学开发了第一个实用化的多点地统计 snesim 方法后, 国内外对多点地统计方法展开了广泛而深入的研究。

多点地统计方法主要由三个基本部分组成: 训练图像、数据模板和数据事件^[16]。训练图像是对所研究对象的一种抽象表达, 描述了对对象的基本几何形态和空间分布特征, 通过扫描训练图像可以获得对象的结构特征, 克服两点地统计学不能再现空间多点相关性的不足。数据模板是由 n 个向量确定的几何形态; 以一个 7×7 的二维模板为例, 设模板的中心位置为 u , 其它位置为 $u_\alpha = u + h_\alpha$ (h_α 是像素点相对模板中心点的位置), 那么数据模板由中心点 u 和 48 个向量 u_α 确定。数据模板中 n 个向量终点处的状态值称为数据事件 d_n , 定义为:

$$d_n = \{pat(u_1), pat(u_2), \dots, pat(u_n)\} \quad (1)$$

式中: $pat(u_\alpha)$ 为训练图像在 u_α 的状态值。

多点地统计的基本原理如图 1 所示: 获取待估点周围的数据事件 d_n ; 用数据模板扫描训练图像, 获取局部图案特征, 统计数据事件 d_n 出现的概率, 根据出现的概率对待估点的状态进行预测。

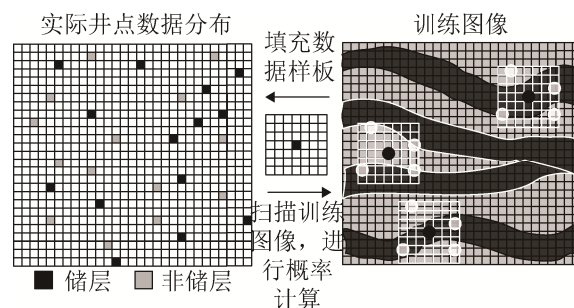


图 1 多点地统计原理^[17]

Fig. 1 Principle of multiple-point statistics^[17]

在具体实现多点地统计时, 不同的学者提出了不同的解决方案。flitersim 是多点地统计学中重要

的算法，能够同时处理连续变量和离散变量。filtersim 通过选定特定的过滤器来降低数据的维度，通过聚类来减少数据模式，提高计算效率^[18]。过滤器和得分是 filtersim 应用中非常重要的概念。过滤器是指与具有 n 个像素的数据模板 τ_n 相对应的 n 个权值的集合，描述为 $f(h_\alpha)(\alpha=1, \dots, n)$ 。过滤得分定义为过滤器与训练图像局部图案的加权值之和：

$$S(u) = \sum_{\alpha=1}^n f(h_\alpha) \cdot pat(u + h_\alpha) \quad (2)$$

过滤器得分最后被赋值到局部图案所在模板 τ_n 的中心位置 u 。一个过滤器得分只能描述该图案的某个模式特征，为了完整描述训练图案的特征，需要定义多个过滤器来反映图案的不同特征，filtersim 默认的过滤器有 6 个，分别是沿南北方向和东西方向的均值、梯度值和曲率值。

filtersim 支持两类数据的条件制约：硬数据和软数据，硬数据是指已知的点数据，软数据一般是精度较低的数据，用来描述局部变化均值或整体变化趋势。算法的流程如图 2 所示。

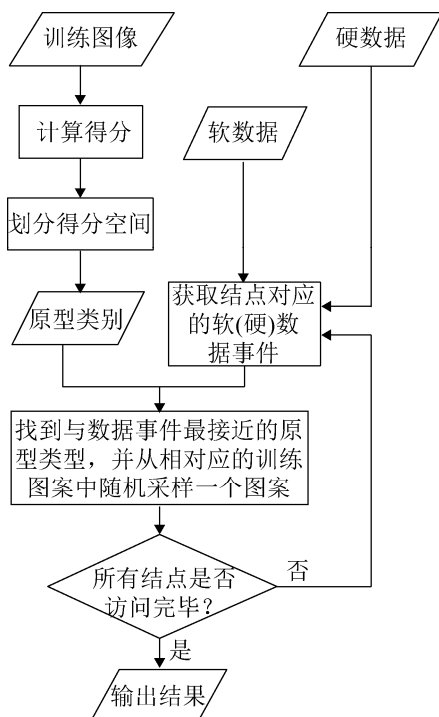


图 2 算法流程
Fig. 2 Flowchart of filtersim

(1) 将 K 个滤波器应用于训练图像，得到 K 个得分图。

(2) 相似的局部图像在得分空间中会有相近的滤波得分；对得分空间进行划分，将相似的局部图案分为一类，以局部图案的均值作为这一类的原型。

(3) 定义一条路径，保证路径访问所有未知点且只访问一次。

(4) 对于路径中的未知点 u ，获取以 u 为中心的数据事件，找到与数据事件最近的原型类别，并从原型类型对应的一系列训练图案中任意采样一个，拷贝到当前网格上。

(5) 继续路径中的下一个结点，重复步骤(4)，直到处理完所有的点。

2 降水融合实验

2.1 实验区域及数据来源

实验使用的气象数据主要有地面气象站逐时观测数据和 CMORPH 小时降水数据。CMORPH 是由美国气候预测中心开发的卫星降水产品，产品的原始时间分辨率为 30min，空间分辨率为 8km，产品范围覆盖了全球 60°S-60°N 的区域，相比其它卫星降水数据，CMORPH 资料能更好地描述东亚地区的降水分布^[19]。由于探测仪器或反演算法的精度等原因，卫星降水数据存在明显的系统误差，为了降低 CMORPH 降水产品的系统误差，国家气象信息中心研究人员基于我国的地面观测数据对系统误差进行了修订，修订后的降水数据时间分辨率为 1 小时，空间分辨率为 0.1°。论文主要以修订后的 CMORPH 降水数据为基础进行实验。

论文以 2017-03-05 江西省的降雨数据进行实验。自 07:00 时起(世界时间)江西大部分地区开始降雨，根据降雨的分布特点，论文最终选择如下区域进行实验，如图 3 所示。实验区域内共有 102 个气象观测站，气象观测站分布相对稀疏。

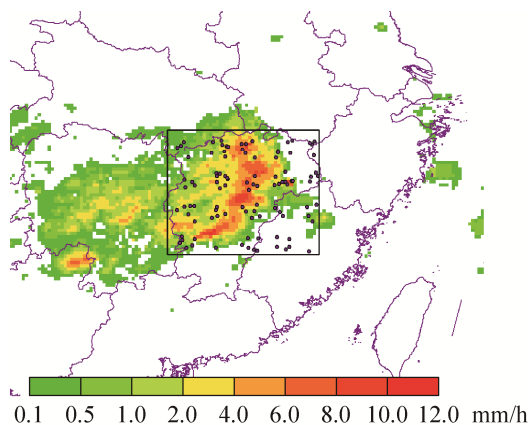


图 3 实验区域
Fig. 3 Experimental area

2.2 Flitersim 降水数据融合

训练图像、软数据和硬数据是进行多点地统计模拟的基础,合理地选择、构造输入数据能极大地提高模拟的精度。论文引入局部均值和局部残差对卫星降水数据和气象站观测数据进行处理,以此构造合适的输入数据。局部均值定义为以某点为中心的一定范围内的所有降雨量的平均值,局部残差定义为该点的真实降雨量与局部均值的差。设某位置某时刻内的降雨量为 $P(x, y)$, 该点的降水均值为 $L(x, y)$, 局部残差为 $R(x, y)$, 那么降雨量可以分解为局部均值与局部残差之和:

$$P(x, y) = L(x, y) + R(x, y) \quad (3)$$

式中: $L(x, y)$ 描述了降雨量的整体变化趋势; $R(x, y)$ 则反映了降雨在局部范围内的差异性。

由公式(3)可知,在降雨整体变化趋势已知的情况下,未知点的降雨量估算就可以转化为对局部残差的估算。卫星降雨数据虽然精度低,但是覆盖范围大,能够反映降雨的宏观分布,因此降雨量估算的关键是确定局部残差的分布,具体方法如下:

(1) 对于卫星降雨数据 $P_2(x, y)$, 用 3×3 的移动窗口对其进行平均运算,得到卫星降水数据的局部均值 $L_2(x, y)$; 从 $P_2(x, y)$ 中剔除局部均值得到卫星降雨数据局部残差 $R_2(x, y)$, 以 $R_2(x, y)$ 作为 flitersim 算法的训练数据。

(2) 对于气象站观测数据 $P_1(x, y)$, 由于 $P_1(x, y)$ 和 $P_2(x, y)$ 为同一区域内的降水数据,应该

具有相同的变化趋势,因此可以将 $L_2(x, y)$ 当作 $P_1(x, y)$ 的局部均值; 从 $P_1(x, y)$ 中剔除局部均值,就可以得到气象站处的局部残差 $R_1(x, y)$, 以 $R_1(x, y)$ 作为 flitersim 算法的“硬数据”。

(3) 由于观测数据相对稀疏,为了提高模拟精度,还需要使用“软数据”。论文以 $R_1(x, y)$ 为主变量,以 $R_2(x, y)$ 为辅助变量,采用协同克里金算法进行插值,得到 1 km 分辨率的“软数据” $S(x, y)$ 。

(4) 以上述三种数据为基础进行 flitersim 多点地统计模拟,得到 1 km 降水局部残差分布。采用线性插值法将 $L_2(x, y)$ 插值到 1 km 空间分辨率,然后与 1 km 降水局部残差相加,就可以得到 1 km 分辨率的降水数据,实验结果如图 7~8 所示。

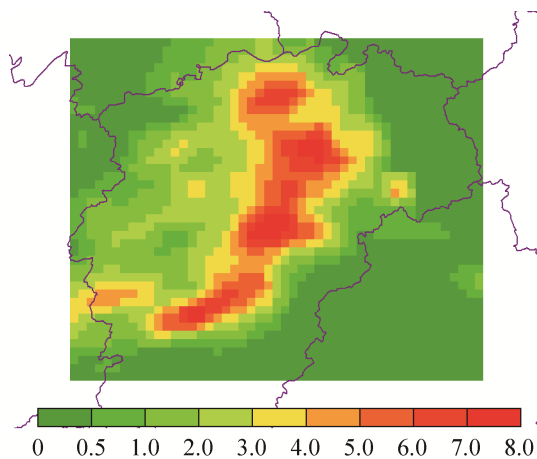


图 4 卫星降水局部均值
Fig. 4 Local mean of satellite precipitation data

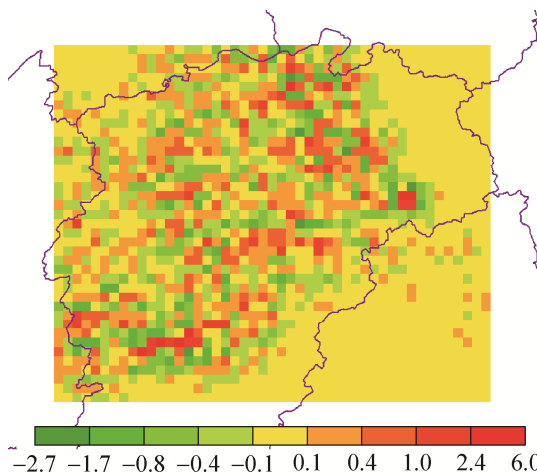


图 5 卫星降水局部残差
Fig. 5 Local residuals of satellite precipitation data

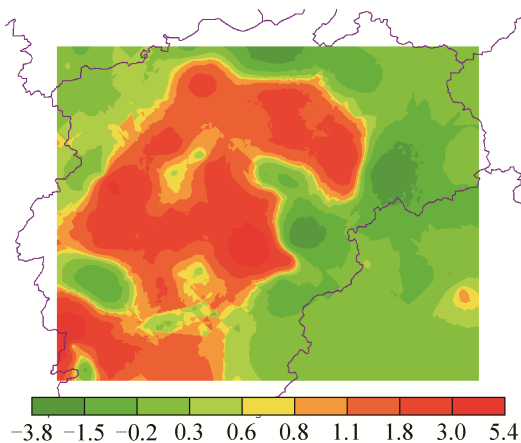


图 6 软数据
Fig. 6 Soft Data

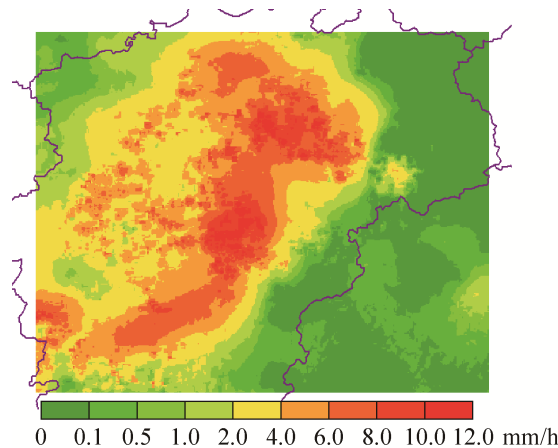


图 8 降水融合效果
Fig. 8 Result of precipitation fusion

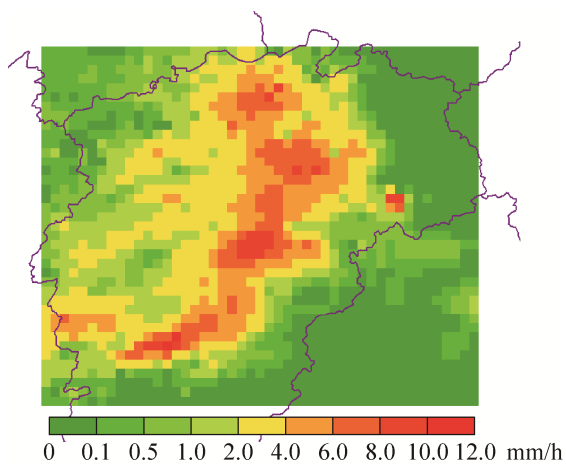


图 7 卫星降水数据
Fig. 7 Satellite precipitation data

2.3 融合效果检验

模拟得到的降雨数据与观测站数据相关,因此不能直接使用站点数据来评估模拟得到的降雨数据误差。论文采用交叉验证法来检验模拟得到的降水数据精度,依次假定某个站点的降水数据未知,用模拟得到的降水数据来估计该站点的降水值,通过计算站点的实际观测值与估计值之间的误差来评估模拟的精度。具体方法如下:将全部观测站随机分为 34 组,每次保留一组观测数据,用剩余的 33 组观测数据和卫星气象数据进行多点地统计分析,得到降水数据后再估算保留的观测站的降水量,以二者之间的误差作为该站点的插值误差。重复 34 次就可以得到全部观测站的误差。

论文使用相关系数(Correlation, 简称为 COR)、均方根误差(Root Mean Square Error, RMSE)和平均绝对误差(Mean Absolute Error, MAE)等 3 个统计量来对交叉检验的误差进行评估。COR、RMSE 和 MAE 的表达式分别为:

$$COR = \frac{\sum_{i=1}^n (P'_i - \bar{P}') (P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (P' - \bar{P}')^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P'_i - P_i)^2}{n}} \quad (5)$$

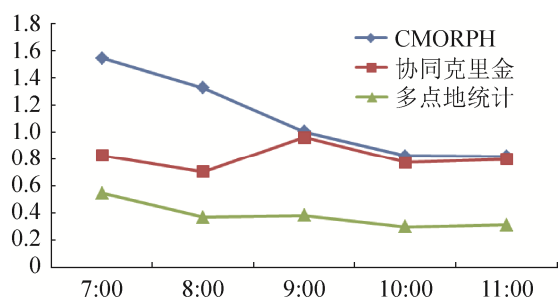
$$MAE = \frac{\sum_{i=1}^n |P'_i - P_i|}{n} \quad (6)$$

式中: n 为观测站的数量; P_i 为站点 i 的观测站; P'_i 为站点 i 交叉检验的估计值。

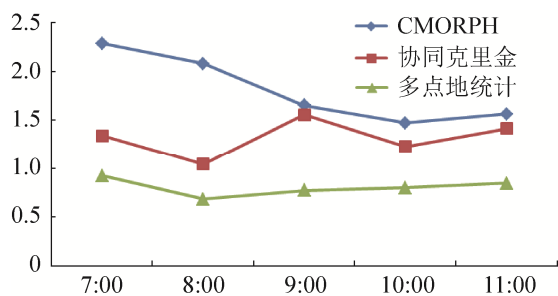
为了消除单次检验可能出现的误差,论文采用上述统计指标对 07:00-11:00 之间的 5 组卫星降水数据和气象观测站降水数据的融合结果进行检验,选用协同克里金方法进行对比分析。

如图 9(a)所示,未处理的卫星降水数据平均绝对误差最大为 1.544,最小为 0.821;采用协同克里金方法处理后最大平均绝对误差为 0.961,最小为 0.703;经过 flitersim 处理后的最大平均绝对误差为 0.545,最小误差为 0.3,相比之下 flitersim 能有

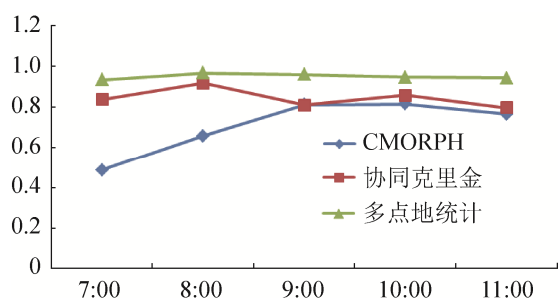
效降低数据的绝对平均误差。如图 9(b)所示, 未处理的卫星降水数据均方根误差最大为 2.28, 最小为 1.46; 采用协同克里金方法处理后最大均方根误差为 1.55, 最小为 1.03; 经过 flitersim 处理后的最大均方根误差为 0.92, 最小误差为 0.68。同样, 如图 9(c)所示, 经过 flitersim 处理后的数据相关系数明显高于其它两组。综上所述, flitersim 能有效提高降水数据的融合精度, 与协同克里金方法相比具有明显的优势。



(a) 平均绝对误差逐时变化



(b) 均方根误差逐时变化



(c) 相关系数逐时变化

图 9 降水融合精度分析

Fig. 9 Accuracy analysis of precipitation fusion

3 结论

针对已有的降水插值方法不能有效反映降雨

的复杂分布特征, 本文提出了基于 flitersim 多点地统计的多源降水数据融合方法。通过引入局部均值和局部残差对卫星降水数据和气象站观测数据进行处理, 以此构造合适的“训练数据”、“硬数据”和“软数据”。采用交叉验证方法和相关系数、均方根误差和平均绝对误差等统计指标对融合结果进行检验、分析, 结果表明: flitersim 能够有效降低数据的均方根误差和平均绝对误差, 提高数据的相关系数。

过滤器和数据模板的大小是影响 flitersim 模拟结果精度的重要因素, 论文在实验时使用了默认的过滤器和数据模板参数, 下一步将根据降水融合的特性, 确定合适的过滤器和数据模板大小。不同地区、不同季节的降雨分布特征不同, 接下来论文将对 flitersim 在不同地区、不同季节下的融合精度进行评估。

参考文献:

- [1] Åke Sivertun, Gumos A. Analysis of Cross Country Trafficability[C]. 12th International Symposium on Spatial Data Handling, Germany: Springer Verlag, 2006: 921-941.
- [2] 李哲. 多源降雨观测与融合及其在长江流域的水文应用[D]. 北京: 清华大学, 2015.
Li Zhe. Multi-source Precipitation Observations and Fusion for Hydrological Applications in the Yangtze River Basin[D]. Beijing: Tsinghua University, 2015.
- [3] Sideris I, Gabella M, Sassi M. Real-time spatiotemporal merging of radar and raingauge precipitation measurements in Switzerland[C]. International Workshop on Precipitation in Urban Areas, Switzerland: Urban Challenges in Rainfall Analysis, 2012.
- [4] Schiemann R, Erdin R. Geostatistical radar-raingauge combination with nonparametric correlograms: methodological considerations and application in Switzerland[J]. Hydrology & Earth System Sciences (S1027-5606), 2010, 15(5): 1515-1536.
- [5] Boushaki F I, Hsu K L, Sorooshian S. Bias adjustment of satellite precipitation estimation using ground-based measurement: a case study evaluation over the southwestern United States[J]. Journal of Hydrometeorology (S1525-755X), 2009, 10(5): 1231-1242.

- [6] 胡庆芳. 基于多源信息的降水空间估计及其水文应用研究[D]. 北京: 清华大学, 2013.
Hu Qingfang. Rainfall Spatial Estimation Using Multi-source Information and Its Hydrological Application[D]. Beijing: Tsinghua University, 2013.
- [7] 潘昉, 沈艳, 宇婧婧. 基于最优插值方法分析的中国区域地面观测与卫星反演逐时降水融合试验[J]. 气象学报, 2012, 70(6): 1381-1389.
Pan Yang, Shen Yan, Yu Jingjing. Analysis of the combined gauge-satellite hourly precipitation over China based on the OI technique[J]. Acta Meteorologica Sinica, 2012, 70(6): 1381-1389.
- [8] Todini E. A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements[J]. Hydrology & Earth System Sciences (S1027-5606), 2001, 5(2): 106.
- [9] Goudenhoofdt E, Delobbe L. Evaluation of radar-gauge merging methods for quantitative precipitation estimates[J]. Hydrology & Earth System Sciences (S1027-5606), 2008, 13(2): 195-203.
- [10] Sebastien Strebelle. Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics[J]. Mathematical Geosciences (S1874-8961), 2002, 34(1): 1-21.
- [11] 吴胜和, 李文克. 多点地质统计学——理论、应用与展望[J]. 古地理学报, 2005, 7(1): 137-144.
Wu Shenghe, Li Wenke. Multiple-point geostatistics: theory, application and perspective[J]. Journal of Palaeogeography, 2005, 7(1): 137-144.
- [12] 石书缘, 尹艳树, 冯文杰. 多点地质统计学建模的发展趋势[J]. 物探与化探, 2012, 36(4): 655-660.
Shi Shuyuan, Yin Yanshu, Feng Wenjie. The Development and Prospect of Multiple-Point Geostatistics Modeling[J]. Geophysical & Geochemical Exploration, 2012, 36(4): 655-660.
- [13] Boucher A. Sub-pixel Mapping of Coarse Satellite Remote Sensing Images with Stochastic Simulations from Training Images[J]. Mathematical Geosciences (S1874-8961), 2009, 41(3): 265-290.
- [14] 杜奕, 张挺. 基于多点信息统计法的土地覆盖图像超分辨率重建方法[J]. 电子学报, 2016, 44(11): 2576-2582.
Du Yi, Zhang Ting. A Super-Resolution Reconstruction Method for Land Cover Maps Using Multiple-Point Statistics[J]. Acta Electronica Sinica, 2016, 44(11): 2576-2582.
- [15] 孟欣然, 梁堰波, 孟宪海. 基于 FILTERSIM 算法的油藏沉积相模拟研究[J]. 计算机工程与设计, 2013, 34(2): 545-549.
Meng Xinran, Liang Yanbo, Meng Xianhai. Using FIMTERSIM for reservoir facies simulation [J]. Computer Engineering and Design, 2013, 34(2): 545-549.
- [16] 尹艳树, 张昌明, 李少华. 多点地质统计学原理、方法及应用[M]. 北京: 地质出版社, 2013: 1-3.
Yin Yanshu, Zhang Changming, Li Shaohua. Multiple Point Geostatistics Theory, Method and Application[M]. Beijing: Geological Publishing House, 2013: 1-3.
- [17] 科瓦列夫斯基. 基于地质统计学的地质建模[M]. 北京: 石油工业出版社, 2014: 86-87.
Evgeniy Kovalevskiy. Geological Modeling on the Base of Geostatistics[M]. Beijing: Petroleum Industry Press, 2014: 86-87.
- [18] Zhang T, Switzer P, Journel A. Filter-Based Classification of Training Image Patterns for Spatial Simulation[J]. Mathematical Geosciences (S1874-8961), 2006, 38(1): 63-80.
- [19] Shen Y, Xiong A Y, Wang Y, et al. Performance of high-resolution satellite precipitation products over China[J]. Journal of Geophysical Research Atmospheres (S0148-0277), 2010, 115(D2): 355-365.