

Study on Soil Moisture of Heihe River Based on Spatial Kriging

Xitong Zhang

Chang'an University, Xi'an Shaanxi
Email: zhang_xitong@foxmail.com

Received: Nov. 20th, 2018; accepted: Dec. 6th, 2018; published: Dec. 13th, 2018

Abstract

Spatial kriging is one of the most important interpolation methods in geostatistics, which models the spatial autocorrelation of the interested variable in the region and provides the unbiased optimized estimation. This article studies the soil moisture in the upstream of the Heihe River, and analyzes the spatial variability of the observations using the R software packages. The spatial kriging model was constructed to calculate the estimation of the soil moisture in the whole region and provide the necessary index for further study of hydrological process in the region.

Keywords

Geostatistics, Regional Variable, Spatial Kriging, Soil Moisture

基于空间克里金插值的黑河流域土壤水分研究

张晞瞳

长安大学, 陕西 西安
Email: zhang_xitong@foxmail.com

收稿日期: 2018年11月20日; 录用日期: 2018年12月6日; 发布日期: 2018年12月13日

摘要

空间克里金是地统计学中重要的一类插值方法, 它通过对研究变量在研究区域内的空间自相关性进行建模, 对待估计的点给出无偏最优估计。本文以黑河上游的典型灌溉农田土壤水分为研究对象, 利用R软件中的相关扩展包, 分析观测数据的空间变异性, 构建研究区域内土壤水分含量的空间克里金插值模型, 得到该区域内土壤水分含量的空间估计, 为进一步研究该流域的生态水文过程提供了必要的指标。

关键词

地统计学, 区域化变量, 空间克里金, 土壤水分

Copyright © 2018 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

地统计学(又称地质统计学)是在传统的统计学基础上,以区域化变量为基础,采用变异函数为主要工具,研究在空间分布上具有自相关性的自然现象的科学[1]。克里金(Kriging)插值法,又称空间局部估计法,是地统计学的主要内容之一。它不同于传统基于距离的空间插值法,比如反距离加权法,只考虑了待测点周围的数据,直接通过插值公式得到结果[2]。克里金法是从研究变量本身出发,根据已有样本点和待测点之间的空间位置关系来对自相关性进行建模,从而对待测点进行线性无偏的估计,同时还能给出估计的方差,比传统方法能更好地描述所研究的地理现象[3]。

土壤是联系大气水、地表水、地下水 and 植物水并进行水分交换的核心地带,揭示土壤水文性质的空间分布特征是研究流域生态-水文过程的重要一环[4],土壤水分含量是水文过程分析中重要的一个指标。本文以流经青海省的黑河上游八宝河流域的土壤水分为研究对象,根据已有的无线传感器网络点收集的土壤水分数据,利用空间普通克里金法对研究区域内的土壤水分进行插值,得到整个地区土壤水分的分布估计,并给出估计方差,为进一步的水文环境分析提供了必要的指标。

2. 空间克里金方法简介

2.1. 区域化变量

区域化变量是指能用空间分布来表征一个自然现象的变量,也就是将空间位置作为随机函数的自变量,它是地统计学中重要的基础理论[5]。设研究区域的位置集合为 D ($D \subseteq R^2$ 或者 $D \subseteq R^3$),空间位置表示为 $x \in D$,所研究的变量在该处的属性值为 $Z(x)$ 。也就是说,将所研究的地质现象看作研究区域 D 内以空间位置 x 为参数的随机函数 $Z(x)$ 。

区域化变量在描述研究对象的空间结构性时,是从空间点对的角度出发的。设研究区域内任意两点为 $x_1 \in D, x_2 \in D$,两点之间的距离向量为 $h \in D$,则这两点之间属性的空间相关性通常用 $Z(x_1), Z(x_2)$ 之间的一阶矩和二阶矩来刻画。

区域化变量 $Z(x)$ 的自协方差函数定义为空间两点 x 和 $x+h$ 处的随机变量 $Z(x)$ 和 $Z(x+h)$ 的二阶混合中心矩,即

$$C(x, h) = \text{Cov}[Z(x), Z(x+h)] = E[Z(x)Z(x+h)] - E[Z(x)] \cdot E[Z(x+h)] \quad (1)$$

通常出于研究需要,对于区域化变量的性质需要做出一些假设,最常见的就是如下的平稳假设[5]。当区域化变量满足下列两个条件时,称该区域化变量满足二阶平稳假设:

- 1). 在研究区域内,区域化变量 $Z(x)$ 的数学期望存在且为常数:

$$E[Z(x)] = E[Z(x+h)] = m \quad (2)$$

2). 在研究区域内, 区域化变量 $Z(x)$ 的协方差存在且只依赖于两点之间空间距离 h :

$$\text{Cov}[Z(x), Z(x+h)] = E[Z(x)Z(x+h)] - m^2 \quad (3)$$

在二阶平稳假设下, 协方差函数 $C(x, h)$ 可以简记为 $C(h)$, 也就是只和点对之间的距离向量有关。

2.2. 变异函数

变异函数是地统计学特有的工具, 是许多地统计学计算的基础, 它定义为区域化变量 $Z(x)$ 在空间点 x 和 $x+h$ 处属性值之差 $Z(x)-Z(x+h)$ 的方差的一半, 即

$$\begin{aligned} \gamma(x, h) &= \frac{1}{2} \text{Var}[Z(x) - Z(x+h)] \\ &= \frac{1}{2} E\{[Z(x) - Z(x+h)]^2\} - \frac{1}{2} \{E[Z(x)] - E[Z(x)]\}^2 \end{aligned} \quad (4)$$

严格来说, 变异函数应当叫做半变异函数, 因为前面有系数, 但是这只是为了公式推导方便, 所以通常不做区分。

在二阶平稳的假设下, 变异函数和协方差函数之间通常有如下关系[6]:

$$\gamma(x, h) = \gamma(h) = C(0) - C(h) \quad (5)$$

变异函数是地统计学研究的出发点, 对于空间变量之间的自相关性的刻画都是从变异函数出发而不是从协方差函数出发。一般来说, 两点之间距离越远, 也就是 h 越大, 变差函数的值越大, 此时两点之间的相关性越小。

在二阶平稳假设下变异函数是关于距离 h 的函数, 该函数有指数模型、球状模型和高斯模型等[3]。实际应用中, 变异函数的具体表达式需要通过拟合得到。首先根据已有样本点数据计算每两个样本点之间的距离 h 和相应的实验变异函数值 $[Z(x) - Z(x+h)]^2$, n 个样本点两两之间总共可以得到 $n(n-1)/2$ 对值, 以 h 为横坐标、 $[Z(x) - Z(x+h)]^2$ 为纵坐标对这些值绘制散点图可以得到变异函数云图[7], 变异函数云图通常用于分析原始数据之间是否存在自相关性。然后对上述 $n(n-1)/2$ 对值进行聚合, 根据如下(6)式计算实验变异函数值[7],

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i+h)]^2 \quad (6)$$

这里 $N(h)$ 表示距离之差为 h 的点对数量。最后选定变异函数的模型, 使用上述得到的实验变异函数值对其中的参数进行拟合得到理论变异函数的表达式[6]。

2.3. 空间克里金插值

不同的克里金插值法在估计的条件和细节上略有不同, 本文以普通克里金(Ordinary Kriging)为例, 阐述克里金插值法的基本原理[3]。克里金插值的基本思想是利用空间上所有已知点的属性值, 进行加权求和来估计未知点处的值, 本质上也是一个线性估计量。设已知观测点集合为 $\{x_i | i = 1, 2, \dots, n\}$, 空间点 x_i 处的属性值为 $Z(x_i)$, 简记为 Z_i 。待估计的点为 x_0 , 待估计点的实际值为 $Z(x_0)$, 简记为 Z_0 , 估计值为 $Z(x_0)^*$, 简记为 Z_0^* 。

待测点 x_0 的估计值 Z_0^* 由如下的线性估计式给出

$$Z_0^* = \sum_{i=1}^n \lambda_i Z_i \quad (7)$$

关键在于如何确定这些权重 $\lambda_i (i = 1, 2, \dots, n)$ 的值。

克里金方法是一种在满足无偏性的条件下寻求估计方差最小的权重系数, 所以克里金插值问题可以

写成一个有约束条件的优化问题[3]:

$$\begin{cases} \min & \text{Var}[Z_0^* - Z_0] \\ \text{s.t} & E[Z_0^*] = Z_0 \end{cases} \quad (8)$$

上述优化问题可以展开成如下的优化问题模型:

$$\begin{cases} \min & C(Z_0, Z_0) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(Z_i, Z_j) - 2 \sum_{i=1}^n \lambda_i C(Z_0, Z_i) \\ \text{s.t} & \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad (9)$$

这里 $C(Z_i, Z_j)$ 表示 $C(|x_i - x_j|)$, 也就是根据点对 x_i 和 x_j 之间的距离计算得到的协方差值, 这个值通常是利用(5)式由已经拟合的理论变异函数计算得到。

(9)式的优化问题可以采用拉格朗日法转化为无约束优化问题求解出一系列的权重值, 从而得到待估计点 x_0 的最小方差无偏估计。

3. 黑河土壤水分的空间克里金插值

本文采用 R 软件中的 `gstat` 包来对黑河流域的土壤水分含量进行空间克里金插值。空间克里金插值一般分为三个步骤:

- 1) 根据已有样本点数据绘制变异函数云图和实验变异函数图, 进行初步分析;
- 2) 选择变异函数模型, 拟合实验变异函数;
- 3) 对待估计点进行空间克里金插值, 给出估计方差。

3.1. 数据来源

本文使用了来自于黑河数据管理计划寒区旱区科学数据中心(<http://westdc.westgis.ac.cn/>)的黑河生态水文遥感试验数据集[8], 该数据集由黑河流域上游生态水文无线传感器网络 WATERNET 于 2013 年收集, 这里使用了当年 8 月 24 号的观测数据。该数据集包括黑河上游八宝河流域内 40 个 WATERNET 节点的观测数据, 这 40 个节点的数采仪编号为 10~49, 但是其中第 46 和 48 号数采仪在 8 月 24 号这天的数据由于外界因素有缺失, 所以实际使用的是 38 个节点提供的观测数据, 这些节点的分布如图 1 所示[8]。

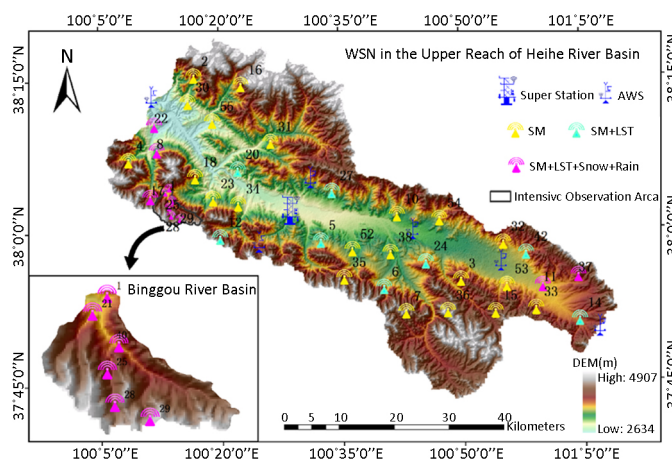


Figure 1. WSN nodes distribution in the Upper Reach of Heihe River
图 1. 黑河上游区域 WSN 节点分布

每个节点都含有 4 cm, 10 cm, 20 cm 三个深度的土壤水分数据, 这里以 4 cm 深度的土壤水分含量为对象。数采仪每 5 分钟采集一次数据, 这样的观测频率过于密集, 所以需要原始数据进行聚合, 这里以天为单位, 每个节点取 8 月 24 号观测数据的平均值。处理后的各节点分布以及土壤水分含量分布见图 2 (单位: cm^3/cm^3)。

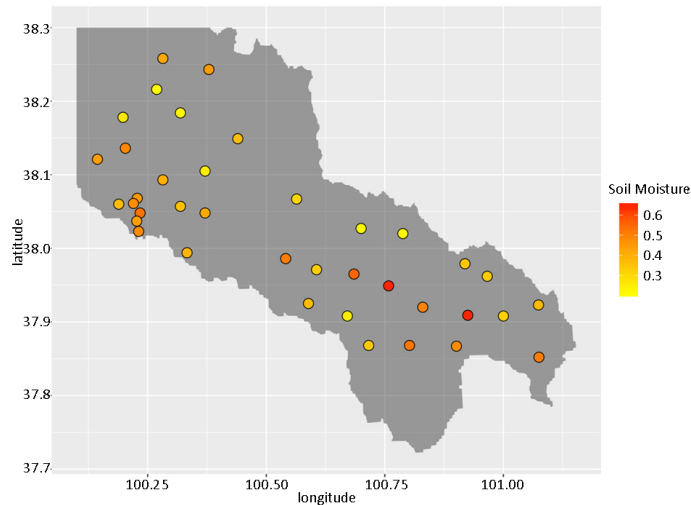


Figure 2. Soil moisture distribution
图 2. 土壤水分含量观测值分布(单位: cm^3/cm^3)

3.2. 绘制变异函数云图和实验变异函数图

首先根据已有 38 个节点的土壤水分含量数据绘制变异函数云图和实验变异函数图。由于不同方向上土壤水分的变化趋势可能不相同, 表现为各向异性[3], 这里选择了正北(0°)、东北(45°)、正东(90°)和东南(135°)四个方向分别绘制图像, 以判断变化趋势较为明显的方向。绘制的变异函数云图见图 3, 实验变异函数图见图 4。

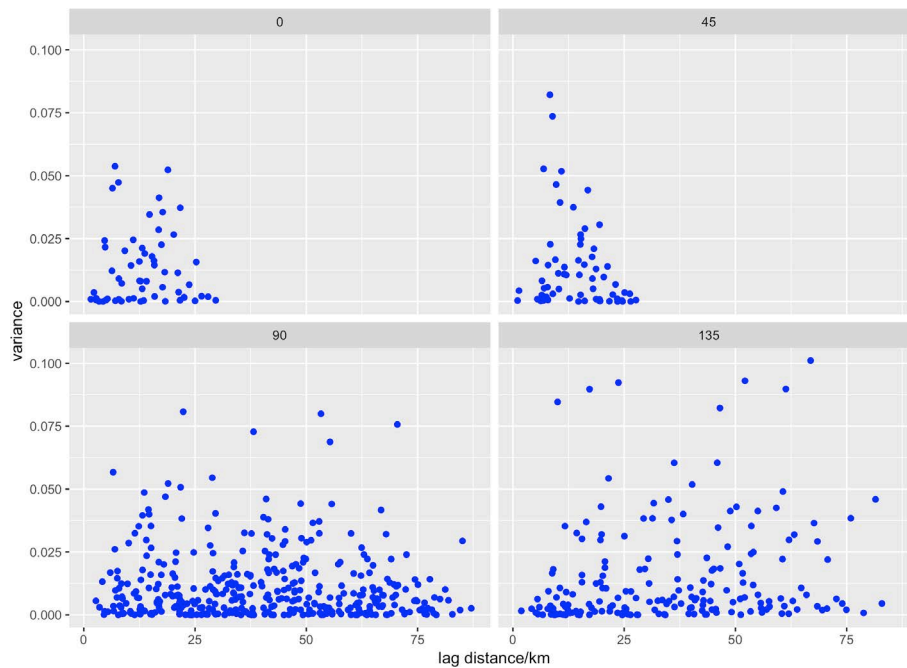


Figure 3. Variogram cloud maps of four directions
图 3. 四个方向的变异函数云图

由变异函数云图和实验变异函数图可以看出,正北和东北两个方向的点对数量较少,变异函数云图上表现为图3中点的数量较少,不利于变异函数的计算,同时由图4中也可以看出,这两个方向上的实验变异函数散点图过于稀疏,无法反映自相关性。正东和东南两个方向的点对数量较多,但是在图4的实验变异函数图像中,正东方向的变化程度没有东南方向的显著,东南方向的实验变异函数散点图分布趋势较为陡峭,而正东方向的实验变异函数散点图分布较为平缓。所以最终选择东南方向的实验变异函数来进行拟合。

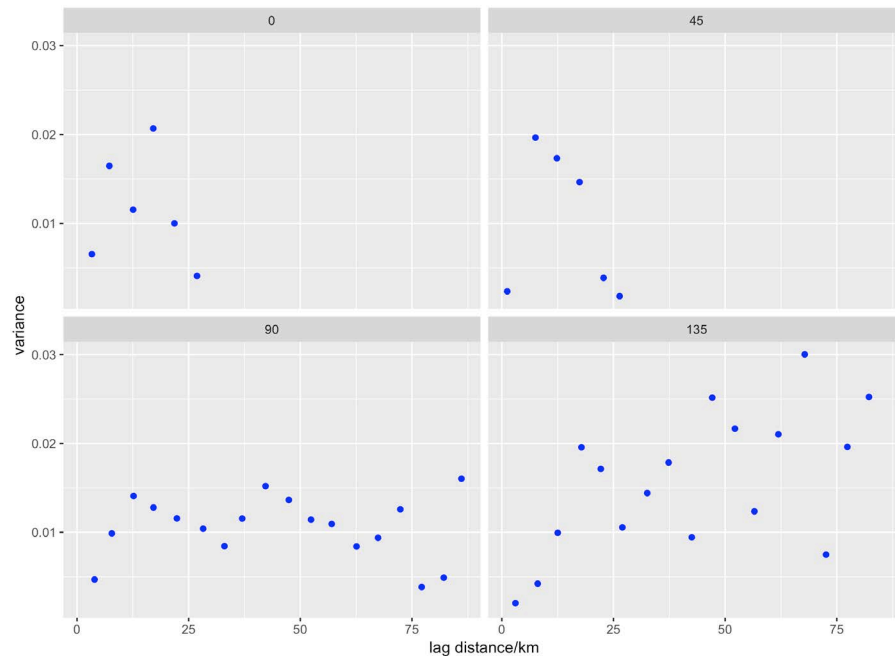


Figure 4. Experimental variogram maps of four directions
图4. 四个方向的实验变异函数图

同时,应当注意到,在滞后距离(lag distance)超过60 km的时候,图3中东南方向的变异函数云图中的点明显变少,这是导致图4中东南方向的实验变异函数图像忽然下降的原因,点的数目变少后,计算得到的方差也就降低了,也就是说,过少的点数计算得到的方差不具有代表性,所以这里只选择滞后距离在65 km以内的实验变异函数点进行拟合。

3.3. 拟合变异函数模型

得到上述的实验变异函数图像后,需要选择变异函数模型来进行拟合。常用的变异函数模型有球状模型和指数模型[3]。球状模型的表达式如(10)式所示:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ C_0 + C \left(\frac{3}{2} \cdot \frac{h}{a} - \frac{1}{2} \cdot \frac{h^3}{a^3} \right), & 0 < h \leq a \\ C_0 + C, & h > a \end{cases} \quad (10)$$

指数模型的表达式如(11)式所示:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ C_0 + C \left(1 - e^{-\frac{h}{a}} \right), & h > 0 \end{cases} \quad (11)$$

上述两式中, C_0 为块金常数, C 为偏基台值, a 为变程。

拟合后得到的两个模型参数见表 1:

Table 1. Fitted model parameters

表 1. 拟合的模型参数

参数	C_0	C	a	SSE
球状模型	0	0.01694676	31.22537	6.789126e-06
指数模型	0	0.02330491	25.1821	5.870586e-06

拟合后两个模型的图形见图 5:

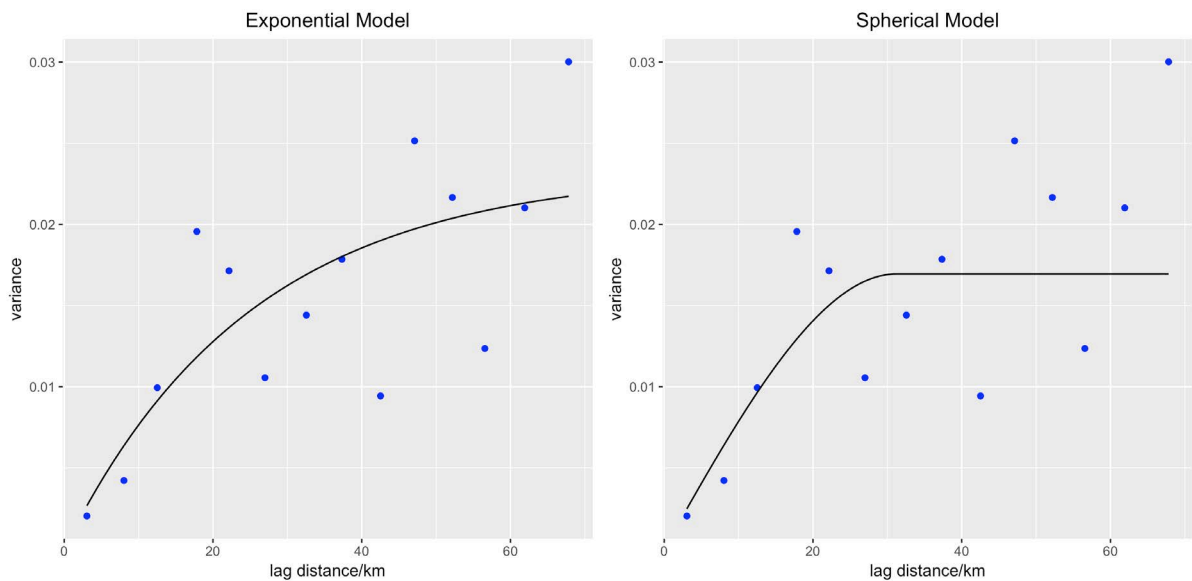


Figure 5. Fitted variogram models, left-Exponential model, right-Spherical model

图 5. 变异函数拟合模型, 左为指数模型, 右为球状模型

由表 1 可知, 指数模型的拟合误差 SSE 相比球状模型的较小, 所以这里最终采用的是指数模型的变异函数。

3.4. 空间克里金插值

得到了变异函数模型之后, 选择需要插值的空间点, 然后根据(9)式求解相应的优化问题, 即可得到空间克里金插值。本文对于整个黑河上游八宝河区域都进行了插值, 即先对八宝河流域网格化, 然后在每个网格中的点进行空间克里金插值, 得到整个八宝河流域的土壤水分含量估计分布图, 同时给出估计的方差。空间克里金插值的结果和方差分布如图 6 所示。

图 6 左边是空间克里金在整个区域的插值分布图, 右侧是相应的方差分布图。左图可以看出八宝河流域土壤水分含量的大致分布, 右图中标星号的点是已有的观测点, 可以看出, 在观测点附件的估计方差较小, 远离观测点的地方估计方差较大, 这与经验是相符合的。

4. 结论

本文利用地统计学中的空间克里金模型, 根据已有观测点收集的数据, 对黑河上游八宝河流域的土壤

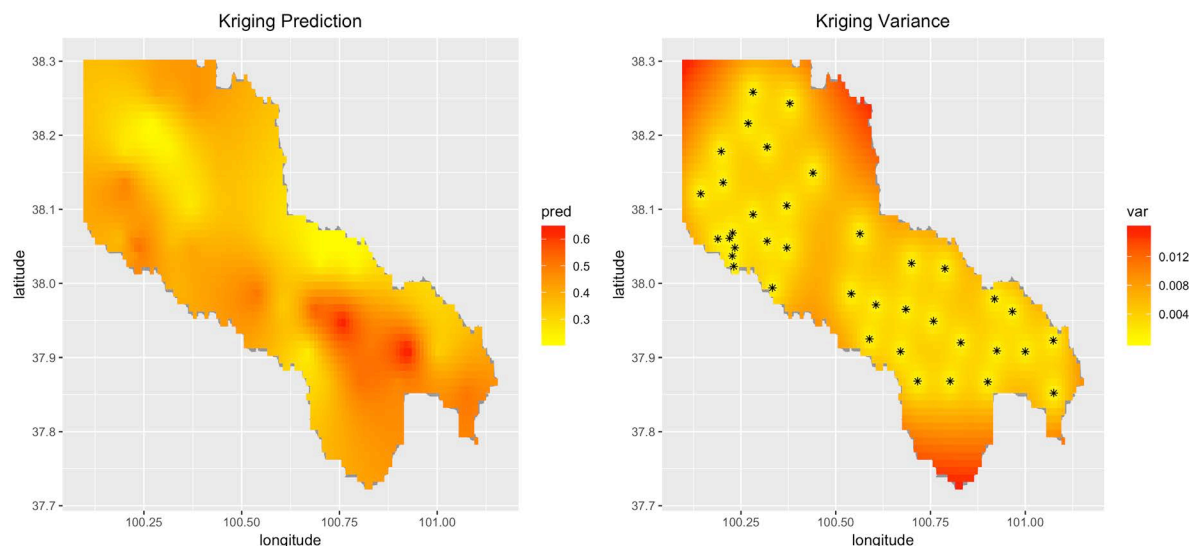


Figure 6. Spatial kriging results, left-Prediction, right-Variance
图 6. 空间克里金插值结果, 左为插值图, 右为方差图

水分进行了插值估计, 并给出了估计方差, 得到该流域土壤水分分布的一个估计, 为进一步的水文环境分析提供了必要的指标。空间克里金模型相对于反距离加权法等传统的空间插值方法, 不仅考虑了空间距离, 还考虑了空间中点与点之间的相对位置以及自相关性, 并且能给出估计精度, 具有较高的实用性和可靠性。

参考文献

- [1] 吕连宏, 张征, 迟志淼, 等. 地质统计学在环境科学领域的应用进展[J]. 地球科学与环境学报, 2006, 28(1): 105-109.
- [2] 杨永川, 杨轲, 王志浩, 等. 空间插值法在热环境流动观测中的应用[J]. 中南大学学报(自然科学版), 2012, 43(9): 422-429.
- [3] 刘爱利. 地统计学概论[M]. 北京: 科学出版社, 2012: 50-76.
- [4] 晋锐, 李新, 阎保平, 罗万明, 李秀红, 郭建文, 马明国, 亢健, 张艳林. 黑河流域生态水文传感器网络设计[J]. 地球科学进展, 2012, 27(9): 993-1005.
- [5] 李钟山. 地质统计学中的区域化变量理论[J]. 世界地质, 1997(2): 85-93.
- [6] Oliver, M.A. and Webster, R. (2014) A Tutorial Guide to Geostatistics: Computing and Modelling Variograms and Kriging. *Catena*, **113**, 56-69. <https://doi.org/10.1016/j.catena.2013.09.006>
- [7] Lichtenstern, A. (2013) Kriging Methods in Spatial Statistics.
- [8] 亢健, 晋锐, 李新, 马明国. 黑河生态水文遥感试验: 黑河流域上游生态水文无线传感器网络 WATERNET 2013 年观测数据集[Z]. 黑河计划科学数据管理中心, 2014. <https://doi.org/10.3972/hiwater.219.2014.db>

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2325-2251，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：sa@hanspub.org