

水面光谱数据的核回归平滑去干扰分析

韦玉春, 王国祥, 程春梅

(南京师范大学虚拟地理环境教育部重点实验室, 江苏 南京 210046)

[摘要] 水面光谱是利用遥感反演水体水质参数的数据基础, 去除光谱中的噪声干扰, 提高光谱的信号干扰比有助于改进水质参数的遥感反演的精度. 本文选择叶绿素 a 浓度相同而悬浮泥沙浓度差异较大的两个水面光谱为代表, 分析了核回归平滑方法对干扰的去除效果. 假定干扰类型为 4 类, 分别是正态分布、瑞利分布、指数分布和泊松分布, 设定的干扰强度分为 4 级. 利用蒙特卡洛模拟方法, 通过 500 轮次的模拟计算了核回归平滑前后水面光谱的信号干扰比, 并与多项式平滑、移动平均、局部回归和鲁棒性的局部回归平滑方法进行了比较. 结果表明, 不论干扰强度高或低, 核回归平滑后的光谱均具有最高的信号干扰比. 在四类干扰中, 核回归平滑对于正态分布的干扰去除效果较好. 与常用的多项式平滑方法相比, 核回归平滑方法比较完整地保留了水面光谱中的峰谷位置信息, 是一种值得推荐的提高水面光谱信号干扰比的方法.

[关键词] 核回归, 平滑, 信扰比, 蒙特卡洛模拟, 水面光谱, 遥感

[中图分类号] TP751.1 [文献标识码] A [文章编号] 1001-4616(2010)03-0097-06

Noise Removal in Spectrum Above Water Surface Using Kernel Regression Smoothing

Wei Yuchun, Wang Guoxiang, Cheng Chumei

(Key Lab of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210046, China)

Abstract Spectrum above water surface with high signal to interference ratio (SIR) is the key to estimate water quality parameters by remote sensing. To decrease interference and increase SIR of spectrum is the important content of spectrum analysis. In this paper, two spectrum above water surface, which chlorophyll a concentration is same and suspended substance concentration is different, was taken as examples to analysis the de-noising effect of the kernel regression smoothing. The paper uses the Monte-Carlo simulation method to estimate the average of SIR by 500 rounds, given four interference disturbance type, i.e. Normal distribute, Rayleigh distribute, Exponential distribute and Poisson distribute, and four interference intensity. The SIR of kernel regression smoothing were also compared with that of Savitzky-Golay smooth filter, moving average, local regression and robust local regression. The result shows that kernel regression smoothing not only increases the SIR of spectrum, but also has the highest SIR than other four methods whether the interference intensity is higher or lower. SIR is the highest when interference is of the normal distribute. Compare with Savitzky-Golay smooth filter, the spectrum by kernel regression smoothing was more smoother and kept more information on spectrum's peak and valley position. The paper concludes that kernel regression smoothing is a better method to decrease interference influence in the spectrum above water surface.

Key words kernel regression smoothing, signal to interference ratio, monte-carlo simulation, spectrum above water surface, remote sensing

水面光谱是野外光谱仪在水面获得的光谱数据, 是水体中各种光学活性成分的综合反映. 水面光谱可以用来估算水体中的水质参数如叶绿素 a 浓度、悬浮物浓度^[1], 是建立水质参数遥感反演模型、进行水体富营养化遥感评价的基础数据.

收稿日期: 2010-04-16

基金项目: 国家自然科学基金 (40771152)、江苏省普通高校自然科学研究计划资助项目 (07KJB420062).

通讯联系人: 韦玉春, 博士, 教授, 研究方向: 环境遥感. E-mail: weiyuchun@njnu.edu.cn

受测量平台的稳定性、风浪、水面波动、光照等影响,光谱仪测量得到的水面光谱数据中不可避免的混杂着噪声干扰.在测量过程中,一般通过重复测量来压抑噪声.例如,ASD 光谱仪手册中建议水面光谱测量时取重复测量的中值作为结果光谱^[2].然而,即使如此,与恒定光源、无风、稳定平台下测量的光谱相比,野外环境下受风浪、天气、水面水汽、载体(船)的稳定性等因素的影响,重复测量处理后的光谱仍可能具有较多的“毛刺”.因此,需要对光谱进行平滑处理,以进一步消除干扰的影响,增强光谱中的有用信号.

数据平滑是提高光谱信噪比的有效方法,使用平滑处理后的光谱数据进行回归预测,其效果普遍较好^[3].在某些领域,一些平滑方法为大家所公认,例如分析化学中经常使用 Savitzky 和 Golay 提出的多项式平滑方法进行光谱预处理^[4].但是,对于水面光谱的平滑处理,尚没有公认的方法.

核回归是统计学中一种用来估计随机变量条件期望的非参数技术,可用来进行数据的拟合,也可以用作平滑方法去除数据中的噪声干扰,自产生之后,在图像处理、信号处理中应用较多^[5-6],并取得了较好的效果.

本文以野外现场测量的水面光谱为例,通过设定不同分布类型干扰和不同干扰强度,利用蒙特卡洛模拟方法,对光谱进行了核回归平滑,并与常用的多项式平滑方法、移动平均方法等进行了比较,通过信号干扰比(信扰比)参数和平滑处理前后光谱的变化比较,得出了核回归更适用于水面光谱平滑处理的结论.本文的研究可为水面光谱预处理方法的选择提供参考依据,进而有助于提高水质参数遥感反演模型的稳健性.

1 数据来源

选择两个典型的水面光谱数据进行比较分析.数据 A 为 2009 年 7 月 20 日南京师范大学仙林校区月亮湾的水面光谱;数据 B 为 2004 年 7 月 5 日太湖平台山样点的水面光谱,该样点是太湖水质的国家常规监测样点之一.水面光谱均使用 ASD FieldPro 按照倾斜方式进行测量,测量时间为 14 点左右,晴天,取重复测量 10 次的中值为样点的光谱.数据 A 在岸边测量,光谱波长间隔为 1 nm,选用测量噪声比较小的波长范围 400~760 nm 进行计算,共 361 个波段.数据 B 在船上测量,光谱波长间隔平均为 1.5 nm,选用的光谱波长范围为 421.3~761.4 nm,共 216 个波段.

两个样点的叶绿素 a 浓度均为 12 ug/L,不同的是,数据 A 的水体中几乎没有悬浮泥沙.数据 B 的位置在太湖中心的平台山,水体悬浮泥沙含量较高,1996~2002 年该样点 7 月平均悬浮物浓度为 69 mg/L.

按照唐军武的计算公式将测量的光谱反射率转换为遥感反射率^[7],在此基础上进行计算.

2 数据处理

2.1 干扰类型和强度

本文考虑 4 类干扰,其数学分布分别为正态分布(Normal)、指数分布(Exponential)、瑞利分布(Rayleigh)和泊松分布(Poisson).干扰强度从高到低设定 4 个级别,其幅度为理论光谱标准差的 1/5、1/10、1/20、1/30,并分别称为干扰 1、干扰 2、干扰 3 和干扰 4.使用各个数学分布的随机函数产生受干扰影响的光谱.

2.2 信号干扰比

采用信号干扰比来度量平滑去干扰的效果.信号干扰比越高,保留的信号成分越多,去除干扰的能力越强.信号干扰比(Signal to Interference ratio, SIR, 本文简称为信扰比)公式来自于 Boscolo^[8],单位为 db.一般认为,当信号干扰比低于 8~10 db 时,信号是无效的.

2.3 平滑方法

核回归平滑(Kr)算法使用高斯核函数,利用 Nadaraya-Watson 方法进行计算^[9,10],窗口宽度可以人为指定或利用 Bowman 的算法通过自动优化产生^[5].

用于对比的 4 种常用平滑方法分别为多项式平滑(sgolay)、移动平均(moving)、局部回归(less)和稳健局部回归(rlless).其中,sgolay 是 Savitzky 和 Golay 提出的多项式平滑方法,在化学谱分析中经常使用.这些方法均包括在 MATLAB 的平滑函数 smooth 中.

2.4 数据处理

使用 MATLAB R2009b 软件进行本文的数据计算和处理。

首先使用傅里叶函数对原始光谱数据进行拟合, 并将拟合的结果作为理论光谱 (并不是所有的光谱数据都能用傅里叶函数进行良好的拟合, 所以, 使用傅里叶函数来去除干扰不具有通用性)。然后, 设定不同窗口宽度和干扰分布类型, 使用数学分布的随机函数产生干扰并叠加到理论光谱中, 产生具有噪声的模拟光谱。对模拟光谱进行平滑处理, 并对照理论光谱计算信扰比。以 500 轮次信扰比的统计结果来确定平滑方法对干扰的去除能力。

在平滑处理中, 除窗口宽度外均使用默认参数。窗口宽度影响着平滑的效果, 太宽会导致信号过于平滑而丢失局部的变化信息, 太窄则会过分强调局部变化而难以消除干扰的影响。本文首先选用 5 个波长间隔为窗口宽度进行处理, 然后用其他波长间隔作为窗口宽度进行了对比。

首先对数据 A 进行平滑处理, 然后给出数据 B 的平滑处理结果。

3 结果与讨论

3.1 原始光谱数据的特征

原始光谱和拟合后的光谱见图 1。将拟合后的光谱作为理论光谱。

决定系数 R^2 和均方根误差 RMSE 参数可用来表明拟合程度。数据 A 的 R^2 和 RMSE 为 0.998 5 和 0.023 93。数据 B 分别为 0.998 6 和 0.026 3。拟合的光谱曲线高度逼近了原始的光谱数据。

数据 A 的光谱具有典型的藻类控制水体的光谱特征。从光谱变化来看, 675 nm 附近藻类叶绿素 a 吸收谷的特征明显。数据 B 的光谱具有典型的悬浮泥沙控制的浑浊水体光谱特征。在视觉上, 该处水面为浊黄色。从光谱变化来看, 670 nm 附近的叶绿素 a 吸收谷的特征不明显。

3.2 干扰类型和强度

正态分布使用理论光谱的均值为干扰光谱的均值, 不同比例的理论光谱的标准差为干扰强度。瑞利分布和指数分布使用不同比例的理论光谱的标准差为参数。泊松分布使用不同比例的理论光谱的标准差为 λ 参数, 计算结果除以 1000 作为干扰。

为便于结果的统一比较, 所有的干扰均以正态分布干扰的幅度为基准进行归一化。

3.3 窗口宽度为 5 时平滑效果比较

设定窗口宽度为 5, 通过 500 轮模拟计算得到的平滑前后的信扰比均值见表 1。

从表 1 可以看出, 给定窗口宽度 5 时, 在干扰强度高 (干扰 1) 的情况下, kr 平滑方法的信扰比高于其他方法。当干扰强度低的情况下, kr 方法的信扰比低于 $sgolay$ 、 $moving$ 和 $loess$ 方法。对于不同数学分布的干扰, 受正态分布干扰的光谱平滑后的信扰比高于其他分布, 表明正态分布的干扰容易通过平滑进行抑制。除了 $rlloess$ 外, 受指数分布干扰的光谱平滑后的信扰比最低, 表明指数分布干扰不容易通过平滑抑制。在选用比较的 5 个平滑方法中, $rbess$ 的信扰比最低, 干扰压抑效果最差。

从平滑前后的信扰比来看, $loess$ 和 $rlloess$ 几乎无法提高信扰比。对于其他三种方法, 在干扰 1 下, 信扰比 $kr > moving > sgolay$, 在干扰 2 下, 信扰比 $moving > kr > sgolay$, 在干扰 3 和 4 下, 信扰比 $moving > sgolay$, kr 无法提高信扰比。这表明, 对于指定窗口宽度 5 的情况下, kr 更适用于平滑干扰比较强的光谱。

3.4 不同窗口宽度的影响

窗口宽度影响着平滑的效果。为此, 比较了干扰强度为 1 和 4 两种情况下不同平滑方法的平均信扰比。表 2 给出了窗口宽度为 3 的情况下各个平滑方法的信扰比均值。

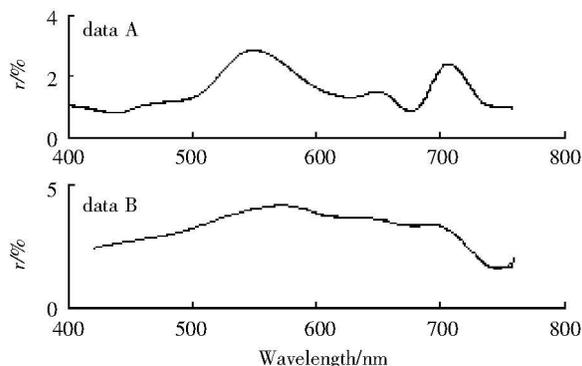


图 1 水面光谱曲线

Fig.1 Spectrum above water surface

表 1 窗口宽度为 5 时的 SIR 变化

Table 1 SIR change when window size is 5

干扰强度	干扰类型	平滑前	平滑后				
			kr	sgolay	moving	bess	rbess
1	Normal	22.7	32.2	25.8	29.6	22.7	4.4
	Rayleigh	19.5	23.1	21.1	22.4	19.5	3.4
	Exponential	15.9	16.8	16.4	16.7	15.9	5.2
	Poisson	20.8	26.1	22.9	24.9	20.8	4.0
2	Normal	28.7	34.2	31.8	35.5	28.7	10.5
	Rayleigh	25.4	28.1	27.0	28.3	25.4	9.8
	Exponential	21.9	22.5	22.3	22.6	21.9	11.3
	Poisson	27.6	31.6	30.0	32.4	27.6	10.3
3	Normal	34.7	34.9	37.8	41.5	34.7	16.3
	Rayleigh	31.4	32.0	33.0	34.3	31.4	15.6
	Exponential	27.8	27.8	28.3	28.5	27.8	17.5
	Poisson	33.9	34.2	36.5	39.1	33.9	16.3
4	Normal	38.2	35.1	41.4	44.8	38.2	19.9
	Rayleigh	35.0	33.5	36.6	37.9	35.0	19.1
	Exponential	31.4	30.5	31.9	32.2	31.4	20.8
	Poisson	37.5	34.7	40.2	42.7	37.5	19.8

表 2 窗口宽度为 3 时 SIR 的变化

Table 2 SIR change when window size is 3

干扰强度	干扰类型	平滑前	平滑后				
			kr	sgolay	moving	bess	rbess
1	Normal	22.7	32.5	22.7	27.4	22.7	22.7
	Rayleigh	19.4	23.0	19.4	21.7	19.4	19.4
	Exponential	15.9	16.7	15.9	16.5	15.9	15.9
	Poisson	20.7	25.7	20.7	23.6	20.7	20.7
4	Normal	38.2	42.4	38.2	43.0	38.2	38.2
	Rayleigh	35.0	37.2	35.0	37.2	35.0	35.0
	Exponential	31.4	31.9	31.4	32.0	31.4	31.4
	Poisson	37.5	41.0	37.5	41.4	37.5	37.5

从平滑前后的信扰比来看, Sgolay, loess 和 rbess 方法没有提高光谱的信扰比. kr 和 moving 方法提高了光谱的信扰比. 与窗口宽度为 5 的结果相比, 窗口宽度为 3 时 moving 和 sgolay 处理后光谱信扰比均有所下降. 与平滑前相比, Sgolay 方法的信扰比没有提高. kr 在干扰 4 的情况下, 信扰比有了提高; 对于正态分布的干扰, 其信扰比从 35.1 提高到了 42.4.

选择干扰类型为正态分布, 干扰强度为干扰 1 和 4, 设定窗口宽度为 3, 5, 7, 9, 11, 选择 kr, sgolay 和 moving 三种方法进行光谱的平滑处理, 其信扰比变化如表 3 所示.

表 3 正态分布干扰情况下窗口宽度与 SIR

Table 3 Relationship between SIR and window size for normal distribution interference

干扰强度	平滑前	平滑方法	窗口宽度				
			3	5	7	9	11
1	22.7	kr	32.6	32.2	29.1	26.0	23.5
		sgolay	22.7	25.8	27.5	28.6	29.5
		moving	27.4	29.5	30.9	31.8	32.4
4	38.2	kr	42.4	35.1	29.9	26.3	23.6
		sgolay	38.3	41.4	43.0	44.2	45.1
		moving	43.0	44.9	45.1	43.8	41.7

在强干扰的情况下(干扰 1), 随着窗口宽度的增加, moving 和 sgolay 方法处理后的光谱信扰比增加, 其中, moving 的略高于 sgolay, kr 的信扰比减少. 在弱干扰的情况下(干扰 4), 随着窗口宽度的增加, sgolay 的信扰比增加, moving 的信扰比出现波动, kr 的信扰比减少. 窗口宽度小时, 信扰比 $kr > moving > sgolay$, kr 均有较高的信扰比. 窗口宽度大时, 信扰比 $sgolay > moving > kr$.

核回归方法中的窗口宽度可以自动优化产生. 设定干扰为正态分布、干扰强度 1 和干扰 4 的情况下,

使用自动优化产生窗口宽度时, 其平均信扰比分别为 32.3 和 43.3 对应的 moving 的信扰比分别为 27.3 和 42.9 对于另外三种分布的干扰, 使用自动窗口 kr 处理后光谱的信扰比也略高一些.

图 2 给出了正态分布干扰、干扰强度 1 情况下 kr 与 sgolay 结果的比较. 其中, kr 的窗口宽度自动优化产生, sgolay 的窗口宽度为 11. 通过典型峰谷位置放大显示的图可以看到, kr 平滑后的光谱与 sgolay 平滑后的光谱相比, 峰谷的位置偏移小, 能够更好地保留峰谷的形状, 更接近理论光谱.

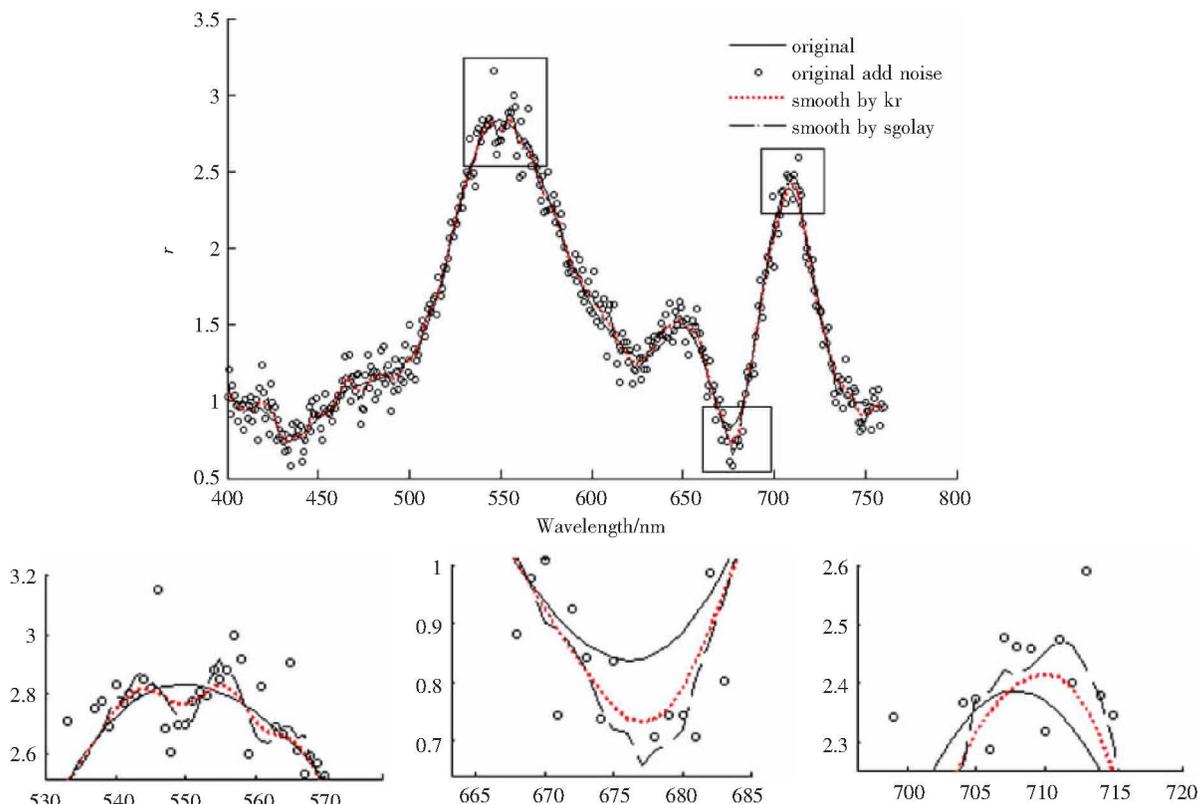


图 2 强干扰情况下核回归平滑与多项式平滑的结果

Fig.2 Spectrum smoothed by kr and sgolay under the higher interference intensity

总体上看, kr 方法对于不同的干扰具有较好的去除效果, 化学光谱分析中常用的 sgolay 方法用较大的窗口宽度平滑弱干扰的水面光谱效果较好.

3.5 数据 B 不同平滑结果的比较

设定不同的干扰分布, 干扰强度为 1 和 4 使用 kr sgolay moving 方法平滑处理数据 B. 其中, kr 通过自动优化产生窗口宽度, 其余方法的窗口宽度为 3. 其平均的信扰比结果如表 4 所示. 通过对比可以看到, 与数据 A 平滑的效果类似, kr 处理后光谱的信扰比均比 sgolay 和 moving 处理后的更高. 对于不同的干扰类型, 指数分布干扰的去除效果最差, 正态分布干扰的去除效果最好.

表 4 数据 B 平滑处理前后的 SIR 变化

Tab 4 SIR change before and after smooth in data B

干扰强度	干扰类型	平滑前	平滑后				
			kr	sgolay	moving	bess	rbess
1	Normal	27.5	36.2	27.5	32.2	27.5	27.5
	Rayleigh	24.7	28.6	24.7	27.3	24.7	24.7
	Exponential	21.3	22.3	21.3	22.1	21.3	21.3
	Poisson	25.8	30.9	25.8	28.9	25.8	25.8
4	Normal	43.0	48.7	43.0	47.7	43.0	43.0
	Rayleigh	40.3	43.3	40.3	42.8	40.3	40.3
	Exponential	36.9	37.6	36.9	37.6	36.9	36.9
	Poisson	42.3	46.7	42.3	46.0	42.3	42.3

综合来看, 核回归平滑能有效地提高水面光谱的信扰比. 与常用的 sgolay 和 moving 方法相比, 核回归

平滑需要的参数少,自动化程度高,自适应能力强,处理后的光谱能够更好地保留峰谷特征,不论干扰强弱,其信扰比均更高。Sgolay 和 moving 方法的平滑效果受窗口宽度的影响较大。窗口大,干扰弱,sgolay 方法处理的效果较好。Moving 方法的平滑效果与窗口宽度和干扰强度相关,干扰强,窗口大,效果较好;干扰弱时,受窗口影响大,需要通过试验确定最佳处理效果。

不同分布类型的干扰比较表明,正态分布的干扰更容易通过平滑提高信扰比,指数分布的干扰使用本文平滑方法提高的信扰比有限。

本文的结果表明,核回归平滑可比较有效地降低水面光谱数据中的正态分布噪声干扰。这种平滑处理对于保证水面光谱数据的可比性、基于水面光谱建立稳健的估算水体水质参数浓度的模型是有意义的。

本文是基于水面光谱数据进行比较分析的结果,其他类型的光谱数据是否具有相同的结论还需要做进一步的研究。

[参考文献]

- [1] Arst H. Optical Properties and Remote Sensing of Multicomponent Water Bodies[M]. Chichester: Springer Praxis Publishing, 2003.
- [2] Analytical Spectral Devices Inc. FieldSpec Pro User's Guide[M]. Boulder: Analytical Spectral Devices Inc, 2002.
- [3] 郑咏梅,张铁强.平滑、导数、基线校正对近红外光谱 PLS 定量分析的影响研究[J].光谱学与光谱分析,2004,24(12): 1546-1548.
- [4] 倪永年.化学计量学在分析化学中的应用[M].北京:科学出版社,2005.
- [5] Bowman A W, Azzalini A. Applied Smoothing Techniques for Data Analysis[M]. New York: Oxford University Press, 1997.
- [6] 郭波涛,易东,王文昌.核平滑半参数回归模型在重复测量资料中的应用[J].中国卫生统计,2007,24(5): 456-458.
- [7] 唐军武,田国良,汪小勇,等.水体光谱测量与分析 I:水面以上测量法[J].遥感学报,2004,8(1): 37-44.
- [8] Boscolo Pan R H, Roychowdhury V P. Independent Component Analysis Based on Nonparametric Density Estimation[J]. IEEE Trans on Neural Networks, 2004, 15(1): 55-65.
- [9] Yi Cao. Kernel Smoothing Regression, 2008[EB/OL]. [2009-03-28]. <http://www.mathworks.com/matlabcentral/fileexchange/19195-kernel-smoothing-regression>.
- [10] Nadaraya E A. On estimating regression[J]. Theory of Probability and its Applications, 1964, 9(1): 141-142.

[责任编辑:丁蓉]