第31卷，第2期
2011年2月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 31, No. 2, pp440-443
February, 2011

基于 Savizky-Golay 多项式的三维荧光光谱的曲面平滑方法

杜树新，杜阳锋，武晓莉

浙江大学工业控制技术国家重点实验室，工业控制研究所，浙江 杭州 310027

摘 要 平滑处理是光谱分析中常用的预处理方法。在二维光谱分析中应用广泛的 Savizky-Golay 多项式曲线平滑方法并不能直接应用于三维荧光光谱的曲面平滑。文章针对三维荧光光谱提出了多项式平滑方法，从而将 Savitzky-Golay 多项式平滑方法扩展到三维荧光光谱，以解决曲面平滑问题。对基于三维荧光光谱的水体有机污染物浓度检测进行了实验研究，实验结果表明采用三维荧光光谱平滑处理方法可有效提高检测模型的精度。

关键词 三维荧光光谱；Savitzky-Golay 多项式平滑；有机污染物浓度检测

中图分类号：O657.3 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2011)02-0440-04

引 言

三维荧光描述了荧光强度同时随激发波长和发射波长变化的关系，因此能完整地描述物质的荧光特征，是一种光谱指纹技术[1]，被广泛应用于水质检测[2,3]、油品检测、药物成分检测[4]、食品安全检测[5]以及化学成分检测[6]等领域。

在光谱测量中，由于外界干扰、仪器本身噪声等的存在，致使测得的光谱数据中含有随机噪声，谱图上有毛刺，影响谱图的质量，增加了后续分析和检测的误差。通过对测量得到的图谱进行平滑处理可降低噪声、提高信噪比，因此光谱平滑是光谱分析中常用的预处理方法。目前，二维光谱的平滑方法研究得比较成熟，常见的有均值法、中值法、多项式拟合法等[7]，其中 Savitzky-Golay 多项式拟合法效果较好，在二维光谱分析中应用最为广泛。但是，针对二维光谱的 Savitzky-Golay 多项式平滑方法并不能直接应用于三维荧光光谱中，其原因是三维荧光光谱本质上是曲面，而二维光谱仅仅是曲线。目前没有针对三维荧光光谱的 Savitzky-Golay 多项式平滑方法的研究报道。

本文将二维光谱中常用的 Savitzky-Golay 多项式平滑方法扩展到三维荧光光谱中，提出了针对三维荧光光谱的 Savitzky-Golay 多项式曲面平滑方法，从而有效地预处理三维荧光光谱，为后续的光谱分析奠定基础。

1 三维荧光光谱的平滑方法

在三维荧光中，荧光强度是激发光、发射光的函数，从数学的角度，本质上是一个类似于图像的二维矩阵，因此三维荧光光谱也称为激发发射矩阵(EEM)。光谱的平滑处理是基于移动窗口技术的，就是说选取包含奇数个数据点的二维窗口矩阵(如 5×5 矩阵)，在该计算窗口上进行平滑处理(计算平滑值)，并用该平滑值替代对应数据点的数值，再是往左或右或上或下移动一格数据，形成新的窗口。

Savitzky-Golay 多项式平滑方法的提出是针对二维光谱提出的，所拟合的是曲线。而在三维荧光光谱中，需要拟合的是曲面。本文将二维光谱的 Savitzky-Golay 多项式平滑方法扩展到三维光谱，从而进行曲面的拟合。

假设三维荧光光谱平滑矩阵窗口大小为 $(2n+1)\times(2p+1)$，$z(x_i,y_j)$ 为各个点的光谱数据(即荧光强度)，x_i 为第 i 个发射光谱波长，y_j 为第 j 个激发光谱波长，$\bar{z}(x_i,y_j)$ 为数据点 (x_i,y_j) 的平滑值。由于 x，y 值为等距分布，在这个拟合窗口中，可定义一个新的坐标系，坐标系的原点为数据的中心点，即数据的中心点为(0，0)，x 的值从 $-n$ 到 n，y 的值从 $-p$ 到 p，第 (i,j) 个数据点可记为 $z_{i,j}$。设拟合的二元 m 次多项式曲面为

$$z_{x,y}=\sum_{k=0}^{m}\sum_{l=0}^{m-k}a_{k,l}x^k y^l \tag{1}$$

上式的二元 m 次多项式也可写成如下形式的两个向量相乘

$$z_{x,y}=\boldsymbol{d}_{x,y}\boldsymbol{b} \tag{2}$$

收稿日期：2010-05-04，**修订日期**：2010-07-04

基金项目：国家自然科学基金项目(60974111)，国家(863 计划)项目(2009AA04Z123)资助

作者简介：杜树新，1967 年生，浙江大学副研究员 e-mail：shxdu@iipc.zju.edu.cn，shxdu@zju.edu.cn

其中

$$\boldsymbol{b}^{\mathrm{T}}=[a_{0,0}\quad a_{0,1}\quad\cdots\quad a_{0,m}\quad a_{1,0}\quad a_{1,1}\quad\cdots\quad a_{1,m-1}\quad a_{2,0}\quad\cdots\quad a_{m-1,0}\quad a_{m-1,1}\quad a_{m,0}]=[b_0\quad b_1\quad\cdots\quad b_{\frac{(m+1)(m+2)}{2}-2}\quad b_{\frac{(m+1)(m+2)}{2}-1}]\in R^{\frac{(m+1)(m+2)}{2}}\tag{3}$$

$$\boldsymbol{d}_{x,y}=[x^0y^0\quad x^0y^1\quad\cdots\quad x^0y^m\quad x^1y^0\quad x^1y^1\quad\cdots\quad x^1y^{m-1}\quad x^2y^0\quad\cdots\quad x^{m-1}y^0\quad x^{m-1}y^1\quad x^my^0]\in R^{\frac{(m+1)(m+2)}{2}}\tag{4}$$

在平滑窗口 $(2n+1)\times(2p+1)$ 中，共有 $(2n+1)(2p+1)$ 个数据点，将这 $(2n+1)(2p+1)$ 个数据点代入到式(2)得到方程组

$$z_{i,j}=d_{i,j}b,\ i=-n,-n+1,\cdots,0,\cdots,n-1,n;\quad j=-p,-p+1,\cdots,0,\cdots,p-1,p\tag{5}$$

其中

$$\boldsymbol{d}_{i,j}=[i^0j^0\quad i^0j^1\quad\cdots\quad i^0j^m\quad i^1j^0\quad i^1j^1\quad\cdots\quad i^1j^{m-1}\quad i^2j^0\quad\cdots\quad i^{m-1}j^0\quad i^{m-1}j^1\quad i^m]\in R^{\frac{(m+1)(m+2)}{2}}\tag{6}$$

式中 $0^0=1$。如果将 $(2n+1)(2p+1)$ 个数据点的值 $z_{i,j}$ 按行拉直而构成新的行向量

$$\boldsymbol{e}=[z_{-n,-p}\quad z_{-n,0}\quad\cdots\quad z_{-n,p}\quad z_{-n+1,-p}\quad\cdots\quad z_{-1,p}\quad z_{0,-p}\quad\cdots\quad z_{0,0}\quad\cdots\quad z_{0,p}\quad z_{1,-p}\quad\cdots\quad z_{n-1,p}\quad z_{n,-p}\quad\cdots\quad z_{n,0}\quad\cdots\quad z_{n,p}]\in R^{(2n+1)(2p+1)}\tag{7}$$

并由行向量 $\boldsymbol{d}_{i,j}$ 构造新的矩阵

$$\boldsymbol{G}^{\mathrm{T}}=[d^{\mathrm{T}}_{-n,-p}\quad d^{\mathrm{T}}_{-n,0}\quad\cdots\quad d^{\mathrm{T}}_{-n,p}\quad d^{\mathrm{T}}_{-n+1,-p}\quad\cdots d^{\mathrm{T}}_{-1,p}\quad d^{\mathrm{T}}_{0,-p}\quad\cdots\quad d^{\mathrm{T}}_{0,0}\quad\cdots\quad d^{\mathrm{T}}_{0,p}\quad d^{\mathrm{T}}_{1,-p}\quad\cdots\quad d^{\mathrm{T}}_{n-1,p}\quad d^{\mathrm{T}}_{n,-p}\quad\cdots\quad d^{\mathrm{T}}_{n,0}\quad\cdots\quad d^{\mathrm{T}}_{n,p}]\in R^{\frac{(m+1)(m+2)}{2}\times(2n+1)(2p+1)}\tag{8}$$

则式(5)写成如下矩阵方程

$$\boldsymbol{e}=\boldsymbol{G}\boldsymbol{b}\tag{9}$$

在上式中，需要求解的未知数为 $\frac{(m+1)(m+2)}{2}$，方程数量为 $(2n+1)(2p+1)$。一般地，方程数大于未知数，因此需要采用最小二乘算法求解 $\boldsymbol{b}$ 的最优值，即

$$\bar{\boldsymbol{b}}=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}e\tag{10}$$

则平滑窗口中各个点的平滑值为

$$\bar{z}\,|_{i,j}=\sum_{k=0}^{m}\sum_{l=0}^{m-k}\bar{b_{\frac{k(2m+3-k)}{2}+l}}i^kj^l\tag{11}$$

在实际计算中，一般只需要计算中心点的平滑值，由式(11)可得到中心点的平滑值为

$$\bar{z}\,|_{0,0}=\bar{b_0}=[1\quad 0\quad\cdots\quad 0]\bar{b}=[1\quad 0\quad\cdots\quad 0](\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}e\tag{12}$$

在上式计算中，应用了 $0^k=0(k\neq0)$，$0^0=1$。式(11)主要用于边界区域(非中心点)的平滑值计算，对于维数为 $N\times M$ 光谱矩阵，平滑窗口大小为 $(2n+1)\times(2p+1)$ 时，边界区域点的数量为 $NM-(N-2n)(M-2p)=2pN+2nM-4np$，如图 1 所示，其中深灰色区域为边界区域，灰色区域为非边界区域。在非边界区域，根据计算点的位置可得到以该计算点为中心点的平滑窗口，并根据式(12)计算得到该点的平滑值。

在平滑窗口计算平滑值过程中，如果只计算平滑窗口的中心点平滑值，将会丢弃掉边界区域上的所有数据点，这对于具有大量三维光谱数据情况，影响不大，但对于三维光谱数据不多的情况，显然不适宜，此时需要计算边界区域的平滑值。边界区域的任何一个点(称之为边界点)，可包含在许多平滑窗口中，考虑到边界点离中心点越近，平滑效果越好，在计算边界点平滑值时，选取中心点离边界点最近的平滑窗口中的一个，并根据该边界点与该平滑窗口中心点的位置情况，由式(11)计算该边界点的平滑值。也就是说，在计算某一数据点的平滑值时，首先判定能否构成以该点为中心点的平滑窗口，如果可构成平滑窗口，则在所构成的平滑窗口按式(12)计算该点的平滑值，如果不能构成这样的平滑窗口，则构造一个其中心点离该点最近的平滑窗口，并在所构成的平滑窗口根据两点的距离关系由式(11)计算该点的平滑值。

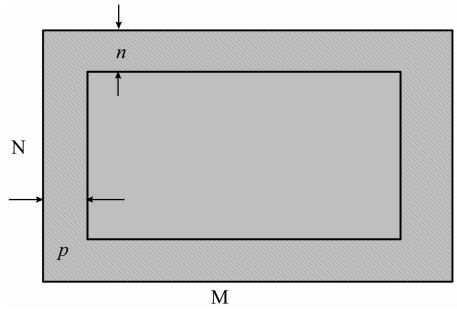

Fig. 1　Boundary area and no-boundary area in EEM

2　实验部分

应用三维荧光光谱分析技术进行水体有机污染物浓度(如化学耗氧量 COD、生物耗氧量 BOD、总有机碳 TOC 等)的检测是近几年的研究热点。本文以检测水体中 TOC 为实验对象研究三维荧光光谱的平滑方法。

2.1　实验仪器及主要参数

三维荧光光谱通过日立公司的 F-4500 型荧光光谱仪测量，激发波长为 225～400 nm，发射波长为 250～700 nm，采样波长间隔为 5 nm，扫描速度为 2 400 nm·min^{-1}。所测量的有机污染物浓度为总有机碳，采用日本岛津公司的 TOC-VCSH 总有机碳分析仪测量得到。

2.2　水样

用于实验的水样采集自某市地表水及生活排污水，滤除体积较大的杂物，共采集了 32 个水样，对每个水样都进行三维荧光光谱、总有机碳指标化学法参考值的测量。由于原始三维荧光光谱中既包含反映有机污染物组成信息的荧光光谱，也包含比较强的瑞利(Rayleigh)散射光，本文采用插值法来消除 Raleigh 散射光。

2.3　平滑效果的评价

在光谱分析研究中，目前还没有一种能直接评价平滑性能的方法，只能通过后续所建立的回归模型的预测性能来评价各种平滑方法。在实验中采用偏最小二乘方法[7]作为回归建模方法，同时采用留一交叉验证方珐评估不同平滑方法所得到的模型性能，所采用的性能评价指标为：(1)预测误差均方根(root mean square error of prediction，RMSEP)；

(2)模型预测值与化学分析值之间的复相关系数 R。计算公式为

$$\mathrm{RMSEP}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(o_i-p_i)^2}$$

$$R=\sqrt{1-\frac{\sum_{i=1}^{N}(o_i-p_i)^2}{\sum_{i=1}^{N}(o_i-\bar{o})^2}} \tag{8}$$

其中 o_i，p_i，$\bar{o}$ 分别为第 i 个水样的化学分析值、第 i 个水样的模型预测值、化学分析值的平均值，N 为样本的数目。RMSEP 值越小，R 值越大，说明模型的精度越高。

2.4 实验结果及其分析

(1)为了比较不同的平滑方法对回归模型预测精度的影响，平滑窗口取为 7×7(即 $n=m=3$)时分别对不采用滤波、采用中值平滑(即将平滑窗口内的数据点的数据进行排序，取中间值作为平滑值)、均值平滑(即将平滑窗口内的数据点的平均值作为平滑值)、本文提出的 Savitzky-Golay 多项式平滑方法(采用 2 次多项式，即 $m=2$)进行了实验，结果如表 1 所示，相对于不采用平滑算法来讲，均值平滑、中值平滑和 Savitzky-Golay 平滑使 RMSEP 分别降低了 1.8%，2.3%，11.1%，20.7%，相关系数分别增加了 0.3%，0.4%，1.7%，3.0%。由此可见，对于本组实验数据，均值平滑的效果最差，Savitzky-Golay 多项式平滑的效果最好，中值平滑居中。

Table 1 Effect of the surface smoothing methods on the modeling performance

采用的平滑方法	RMSEP	R
不采用平滑方法	11.055 3	0.926 0
均值平滑方法(包含中心点)	10.806 2	0.929 4
中值平滑方法	9.829 5	0.942 0
Savitzky-Golay 二次多项式平滑方法	8.763 6	0.954 1

(2)为了比较多项式不同阶次、不同平滑窗宽对预测精度的影响，采用 Savitzky-Golay 多项式平滑方法对二次多项式($m=2$)和三次多项式($m=3$)、不同的窗宽进行了实验，结果如表 2 所示。从表中可看出，在本实验中多项式的阶次对平滑效果影响不大。阶次高低的影响应该与光谱本身的特性有关，由于阶次越高，平滑计算的复杂度越大，在平滑效果没有得到显著提高的情况，一般应采用低次多项式拟合。另外，从表 2 中可发现发射光谱和激发光谱的窗宽对检测性能有一定的影响，以 2 次多项式拟合为例，RMSEP 从最差的 10.211 2 变化到最好的 7.418 3，r 从最差的 0.938 8 变化到最好的 0.967 3。因此在应用 Savitzky-Golay 多项式平滑方法对三维荧光光谱进行平滑处理时，需要适当选择平滑窗口的窗宽。窗宽过小，起不到平滑效果；窗宽过大，一方面增加了计算复杂度，同时会掩盖一些有用的光谱信息而导致过平滑，使性能变差。

Table 2 Effect of polynomial order and window width on modeling performance with Savitzky-Golay polynomial surface smoothing

发射波长窗宽	激发波长窗宽	二次多项式		三次多项式	
		RMSEP	r	RMSEP	r
5	5	9.966 4	0.940 3	10.452 3	0.931 2
5	7	9.346 0	0.947 7	9.912 0	0.941 0
5	9	9.084 1	0.950 7	10.373 4	0.935 1
5	11	8.996 7	0.951 6	12.651 1	0.901 9
5	13	9.324 1	0.948 0	16.818 8	0.818 7
7	5	9.564 9	0.945 1	9.397 3	0.947 1
7	7	9.829 5	0.942 0	12.095 3	0.910 7
7	9	8.476 9	0.957 2	8.791 6	0.953 9
7	11	8.274 9	0.959 3	8.791 2	0.953 9
7	13	8.208 2	0.959 9	9.219 0	0.949 2
9	5	9.466 1	0.946 3	9.779 1	0.942 6
9	7	8.260 6	0.959 4	8.425 5	0.957 7
9	9	8.209 3	0.959 9	8.044 2	0.961 5
9	11	7.694 9	0.964 9	7.790 5	0.964 0
9	13	7.745 6	0.964 4	7.680 7	0.965 0
11	5	10.211 2	0.947 2	12.840 4	0.898 7
11	7	8.201 6	0.960 0	9.587 6	0.944 9
11	9	7.882 9	0.963 1	7.311 4	0.968 3
11	11	7.787 8	0.964 0	7.492 4	0.966 7
11	13	7.418 3	0.967 3	7.122 2	0.969 9
13	5	10.091 2	0.938 8	13.801 6	0.882 0
13	7	8.417 2	0.957 8	9.431 5	0.946 7
13	9	7.922 6	0.962 7	7.652 3	0.965 2
13	11	7.676 9	0.965 0	7.839 8	0.963 5
13	13	7.510 6	0.966 5	6.995 1	0.971 0

3 结束语

本文将二维光谱的 Savitzky-Golay 多项式平滑方法扩展到三维光谱的曲面平滑。对 32 个水样进行三维荧光光谱检测水体总有机碳的实验说明了所提出的三维光谱 Savitzky-Golay 多项式曲面平滑方法的有效性。在本试验中，多项式的阶次对平滑效果影响并不显著，但平滑的窗宽大小对应平滑效果有一定的影响。

References

[1] SHANG Li-ping, YANG Ren-jie(尚丽平，杨仁杰). *In Situ* Fluorescence Spectral Analysis Technique and Its Applications(现场荧光光谱技术及其应用). Beijing: Science Press(北京：科学出版社), 2009.

[2] Henderson R K, Baker A, Murphy K R, et al. Water Research, 2009, 43: 863.

[3] Hudson N, Baker A, Reynolds D. River Research and Applications, 2007, 23: 631.

[4] JIANG Jun-duo, WU Hai-long, XIA A-lin, et al(江军朵, 吴海龙, 夏阿林, 等). Chemical Journal of Chinese Universities(高等学校化学学报), 2008, 29(1): 71.
[5] WANG Yu-tian, CUI Li-chao, LI Yan-chun, et al(王玉田, 崔立超, 李艳春, 等). Measurement Technique(计量技术), 2006, (3): 26.
[6] HAN Qing-juan, WU Hai-long, NIE Jin-fang, et al(韩清娟, 吴海龙, 聂瑾芳, 等). Chemical Journal of Chinese Universities(高等学校化学学报), 2007, 28(5): 827.
[7] XU Lu, SHAO Xue-guang(许　禄, 邵学广). Chemometrics(化学计量学方法). Beijing: Science Press(北京: 科学出版社), 2004.
[8] Zepp R G, Shelddon W M, Ann M M. Marine Chemistry, 2004, 89: 15.

The Surface Smoothing Methods for Three-Dimensional Fluorescence Spectrometry Based on Savitzky-Golay Polynomial Smoothing

DU Shu-xin, DU Yang-feng, WU Xiao-li
State Key Lab of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, China

Abstract Spectral smoothing is commonly used as effective pretreatment methods in the spectral analysis. However, the conventional Savitzky-Golay polynomial smoothing methods used in two-dimensional spectral analysis can not be applied to three-dimensional spectrometry directly. In the present paper, a polynomial surface smoothing method is proposed, and the two-dimensional Savitzky-Golay polynomial smoothing methods are extended to three-dimensional fluorescence spectra. Experiment for detecting dissolved organ matter in water using three-dimensional florescence spectrometry was carried out, and experimental results show that the smoothing method for the three-dimensional fluorescence spectrum can effectively improve the modeling accuracy.

Keywords Three-dimensional fluorescence spectrometry; Savitzky-Golay polynomial smoothing; Detection of dissolved organ matter

(Received May 4, 2010; accepted Jul. 4, 2010)