

文章编号: 1672-8785(2018)10-0040-04

红外多光谱遥感数据的校准

方 正 林智乐 钟 硕 胡伟峰 陈思媛

(厦门大学仪器与电气系, 福建 厦门 361102)

摘 要: 主要介绍了一种对由特定红外图谱一体化系统采集的光谱数据进行校准与分析的方法。在实验部分, 首先对得到的设定已知特定温度的黑体红外光谱数据进行归类和整理, 然后在最大保留特征信号的条件利用不同的滤波算法与滤波窗宽尽量消除信号噪声, 并采用偏最小二乘算法对所得信号进行校准, 使之与理论值相符。最后利用校准结果构建数学模型, 并考虑主成分个数这一概念, 检验模型准确度。通过数据分析可知, 由这个特定红外图谱一体化系统采集的光谱数据经偏最小二乘算法校正后与理论值相近, 此时能够拟合出校正数列与波数之间的函数关系, 且受温度影响较小。

关键词: 红外光谱; 校正; 偏最小二乘

中图分类号: O433 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2018.10.007

Correction of Infrared Multispectral Remote Sensing Data

FANG Zheng, LIN Zhi-le, ZHONG Shuo, HU Wei-feng, CHEN Si-yuan

(Department of Instrumentation and Electrical Engineering, Xiamen University, Xiamen 361102, China)

Abstract: A correction and analysis method for the spectral data collected by a specific infrared spectrum integrated system is presented. In the experiment, the infrared spectral data of a blackbody at a known specific temperature are firstly classified and sorted. The signal noises are eliminated as much as possible with different filtering algorithms and filter window widths on condition that the characteristic signals are maximally reserved. Then, the signals obtained are corrected by a partial least square algorithm so as to let them agree with the theoretical value. Finally, the correction results are used to construct a mathematical model and the concept of confidence is taken into account to verify the accuracy of the model. It is found in data analysis that the spectral data collected by the integrated infrared spectrum integrated system are close to the theoretical value after being corrected by the partial least-squares algorithm and the function relationship between the corrected series and the wave number can be fitted and is less affected by temperature.

Key words: infrared spectroscopy; correction; partial least square

收稿日期: 2018-08-16

作者简介: 方正(1976-), 男, 湖北武汉人, 教授, 博士生导师, 主要从事 X 射线检测、基于 IMU 的智能传感与伺服、导航制导与控制、康复工程等方面的光机电算一体化装备研究。

E-mail: fangzheng@xmu.edu.cn

0 引言

红外光谱对样品具有非常普遍的实用性, 而且对样品本身的破坏少之又少; 无论是固态、液态或者气态的样品都能够应用, 无机物、有机物、高分子化合物都可检测。因为只要在空间中存在一定的物质, 就能对此空间内的光线造成一定影响。自从 1965 年光谱技术首次被用来分析土壤质地和水分含量以来, 很多学者利用土壤的近红外光谱对水分、有机质、碳、氮、铁、阳离子交换量等土壤特性进行了调查研究。结果表明, 利用近红外光谱技术尤其是光谱分辨率更高的高光谱技术能够成功地预测这些土壤特性^[1-5]。此外, 红外光谱测试技术还具有检测迅捷、操作简单、可靠性好、灵敏度高特性。因此, 该技术显然已经成为了当今结构化学和分析化学领域最常用亦或是不可缺少的一个角色。红外光谱技术在高分子聚合物的结构、特征和力学特性研究以及物理、遥感、气象、生物、天文、医学等范畴也有相当普遍的应用。

最近, 红外光谱技术已成为轻重工业各种领域用于无接触检测的重要技术。这种技术的实用性主要在于速度快、成本低, 不需要或者仅需要很少样品制作准备以及对分析样品没有破坏性^[6]。

1 理论与方法

1.1 中值滤波

中值滤波算法在应对脉冲型环境噪声时有着非常良好的去除效果, 而且这种滤波方式在滤除噪声的同时, 不会对信号的边缘造成太大破坏, 保护其不被模糊化^[7]。

1.2 小波滤波

小波变换的最大特性就是可描述试验信号中部分区域的频率特性, 因此我们可以把小波滤波看做是傅里叶变换的突破性进展^[8]。小波滤波由式(1)表述:

$$W_x(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (1)$$

式中, $a, b \in \mathbb{R}$; a 为尺度因子; b 为移位因子。显然尺度因子如果小于 0 的话, 式(1)将会缺乏所谓的物理意义。

1.3 偏最小二乘法

偏最小二乘法对多变量模型的拟合效果好, 最小二乘法对单变量模型的拟合效果好。本次实验中, 考虑到光的频率和环境温度对黑体辐射强度的影响, 于是选择偏最小二乘法。

近红外光谱数据常常包含比观测更多的变量, 变量共线性是典型的。也就是说, 一些变量可以表示为其他变量的线性组合。有时候, 由于训练集过度拟合, 使用其建立的 PLS 回归模型会产生较差的预测结果。共线性的存在限制了 PLS 的直接应用。从理论上讲, 一些变量可能包含噪声和背景等校准模型中无用或不相关的信息, 这会使 PLS 模型恶化^[9]。因此, 我们还需要做很多工作来处理变量选择问题。

2 实验过程

2.1 光谱数据采集

在本次研究中, 用于数据采集的光学仪器是厦门大学航空航天学院方正实验室提供支持的红外图谱一体化系统^[10]。如图 1 所示, 该原理系统由跟踪扫描转镜、红外窗口、分光镜、成像镜头、测谱镜头、红外光纤、傅里叶能谱测量单元、实时处理器以及红外面阵列探测器组成。

2.2 光谱校准

在本次实验中, 利用 Bruker OPUS 5.5 软件读取光谱文件, 并采用 Matlab 2016a 软件对光谱数据进行校正与分析。

在这里展示 600 °C/700 °C/800 °C/900 °C/1100 °C/1300 °C 高温黑体, 测量仪器不带窗口; 每个温度各 3 组, 每组 20 次, 共计 360 组光谱数据用于综合校正过程与效果。首先将不同温度的辐射曲线分为不同组别, 并将相同温度的归为一组。例如, 对于 600 °C 下的 60 组数据, 把相同波数的辐射值简单叠加以求平均值, 后文称之为平均曲线。为增加处理数据

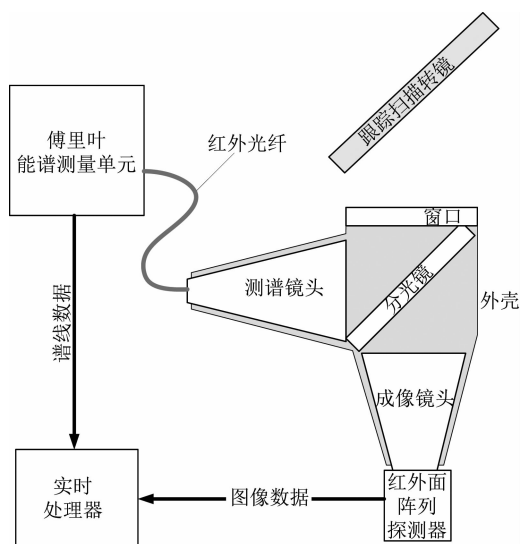


图1 原理系统的结构图

的稳定性,减少奇异点的出现。之后需分别对理论曲线和平均曲线进行归一化处理,使二者的最小值同为0、最大值同为1。接着用理论值当分子,用平均曲线当分母,得到各个不同波数下的数值即所要求的校正数列。当然这里的校正数列还仅仅只是一串简单的数列,从中并看不出它与波数及温度之间有什么关系。

在某个指定温度下,每个波数都有一个对应的校正数值,于是一段波数区间所对应的校正数值就构成了校正数列。为了探求校正数列与波数及温度之间的关系,需要用到600~1300℃之间6个温度档位的所有校正数列。

将波数设置为第一个自变量,将温度设置为第二个自变量,并将校正数值设定为它们的因变量。通过对其进行偏最小二乘拟合,试图找出波数、温度与校准数值之间的函数关系。在将上文提到的异常大的校正数列区间无效化后,能够得到较为良好的拟合效果(见图2)。

从图2中可以看出,数列的点大致都集中在斜线附近,稍微偏远一些的点也都比较对称地分布在斜线两侧。结果表明,校正数列与波数归一化后的相关系数为-0.7177,与温度归一化后的相关系数为-0.0770。再经反归一化还原后,可以归纳出一个用于计算校正数列的简单公式:

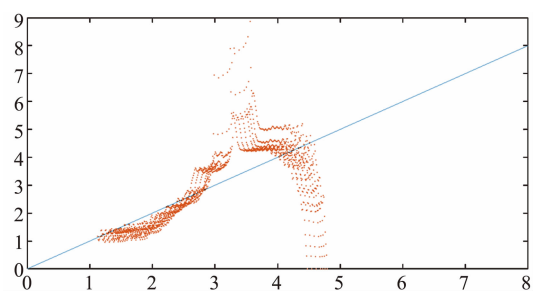


图2 校正数列的拟合效果图

$$J(i) = 10.1302 - 0.0028N - 0.0004T \quad (2)$$

由式(2)可知,温度对校正数列大小的影响相对于波数来说又非常小。因此,在精确度要求不是非常高的情况下,我们可将式(2)简化到只剩下前面两项。于是本实验尝试寻找出数值只与波数有关而与温度无关的校正数列。通过将其数值与各个波数的辐射强度相乘进行校正,使实验测量值尽可能地接近理论值。

图3所示为仅考虑波数影响时获得的校正效果。其中,红色曲线为理论值,蓝色曲线为校正后的实验测量值。

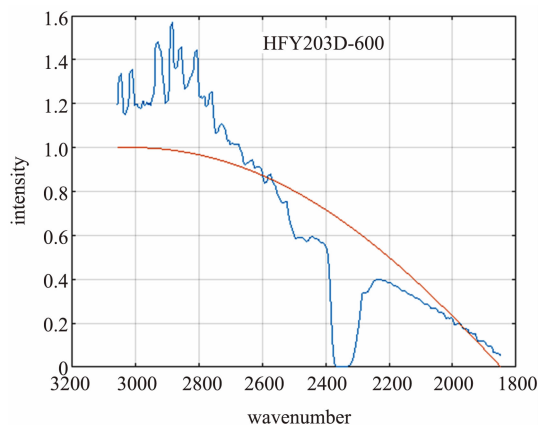


图3 873 K 黑体辐射的初步校正曲线

但是这次校正的效果并不太理想。通过仔细观察拟合效果散点图,发现校正数列数值激增的区间之前的散点分布大致符合二次函数曲线,而校正数列数值激增的区间之后的散点分布大致符合三次函数曲线。所以将1800~2400波数范围内的校正系数设为波数的一元三次函数,并将2401~3200波数范围内的校正系数设为波数的一元二次函数。最后,经拟合得出校正系数关于波数的分段函数:

$$J = \begin{cases} 1.23 \times 10^{-8} n^3 - 7.80 \times 10^{-5} n^2 + 0.164n - 115, \\ n \in (2400, 3200); \\ 8.83 \times 10^{-6} n^2 - 0.0526n + 79.7, \\ n \in (1800, 2400] \end{cases} \quad (3)$$

式中, J 为校正系数; n 为波数。

图 4 为分段校正数列的拟合效果图。可以看出, 其效果明显比一次函数拟合好很多。其中, 二次拟合曲线部分的采样点数为 175, 误差平方和为 12.8467; 三次拟合曲线部分的采样点数为 117, 误差平方和为 16.4148。

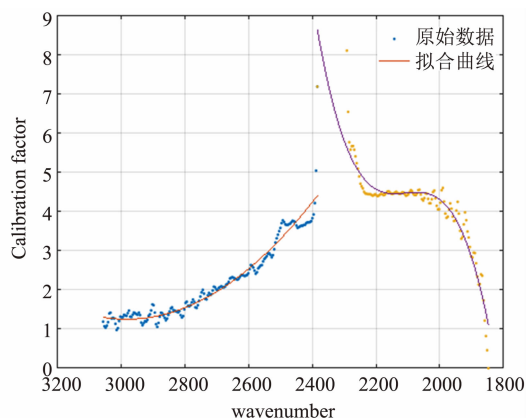


图 4 分段校正数列的拟合效果图

由于温度的影响较小, 拟合时使用的样本点数值为所有温度在该样本点上的平均值。

校正后的光谱曲线可以很接近理论值, 并且由于校正数列并不是关于温度的函数而是仅仅关于波数的函数, 这就增大了校正数列的适用性, 因此可在未知黑体温度的情况下使用。图 5 和图 6 所示分别为 873 K (600 °C) 和 1173 K (900 °C) 条件下高温黑体的辐射曲线。

通过观察校正后的辐射曲线可以发现, 在温度升高之后, 校正曲线会有上移趋势, 这是由于忽略温度影响所致。因为虽然温度对校正系数的相关性很小, 只有波数的十分之一左右, 但终究还是有的。如果想要获得更高的精确度, 那么可以引入关于温度的负反馈系数。当温度升高时, 校正系数会微微降低, 将本来要上移的校正曲线略微下压。

在不考虑温度的情况下, 相对于每一档温度, 该校正数列的效果都不太一样。表 1 列出了每个温度下平均每个采样点的校正值与理论

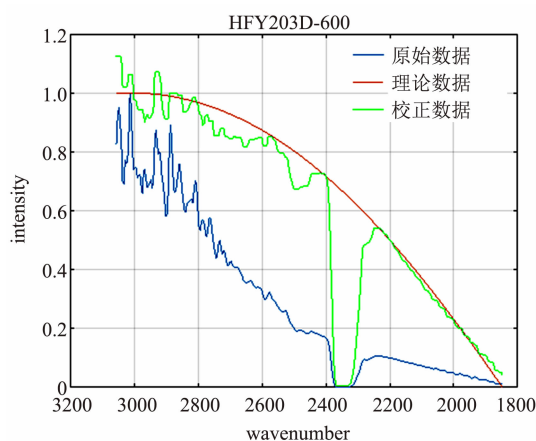


图 5 873 K 黑体辐射校正曲线

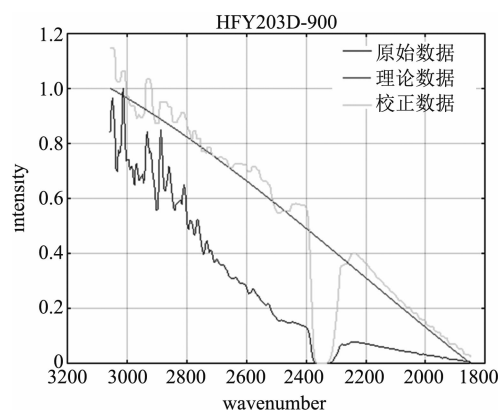


图 6 1173 K 黑体辐射校正曲线

值相减的绝对值与理论值的比值(以百分数显示)。

表 1 各个温度下的误差

温度(°C)	600	700	800	900	1100	1300
误差(%)	8.20	6.56	6.40	6.34	6.05	5.78

由表 1 可见, 温度高时误差相对较小, 这与中波红外波段对高温物体较为敏感的特性相符。

3 结束语

综上所述, 本文介绍的校正方法可以将不同波数段所对应的校正数归纳成不同的函数, 弥补了用同一函数求取不同波数的校正数时所带来的巨大误差, 从而更加贴合实际情况。但该方法的不足在于只探讨了本文所采集数据相应波数的校正函数, 覆盖范围还不够全面。

(下转第 48 页)