

文章编号: 1672-8785(2016)05-0043-06

苹果近红外光谱的预处理

张小英

(乐山师范学院物理与电子工程学院, 四川乐山 614000)

摘 要: 原始近红外光谱数据含有大量的噪声信号和较大的数据量, 所以在进行光谱数据分析之前对光谱数据进行预处理是非常必要的。近红外光谱数据的预处理主要有两个任务, 一是降噪, 提高模型的稳健性和预测结果的准确性; 二是数据压缩, 以便于数据的存储, 提高建模速度。传统的近红外光谱数据预处理方法各有局限, 很难在这两方面都得到令人满意的效果。将小波分析用于苹果近红外光谱数据的预处理, 并选取峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 和归一化相关系数 (Normalized Correlation, NC) 作为评价指标。与常用的 Savitzky-Golay 平滑滤波和多元散射校正相比, 小波方法不仅能有效地实现数据压缩, 而且在噪声去除和光谱细节保持等方面都具有优势。

关键词: 近红外光谱; 预处理; 小波分析; Savitzky-Golay 平滑; 多元散射校正

中图分类号: TH744.4 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2016.05.009

Pretreatment of Near Infrared Spectra of Apples

ZHANG Xiao-ying

(School of Physics & Electrical Engineer, Leshan Normal University, Leshan 614000, China)

Abstract: The original Near Infrared (NIR) spectral data inevitably contain a large amount of noise signals and data. So, before spectral analysis, it is necessary to preprocess the spectral data. The preprocessing of NIR spectral data mainly includes two tasks. One task is to de-noise so as to improve the robustness of the model and the accuracy of the prediction result. Another task is to compress the data for storage and modeling speed improvement. Traditional spectral preprocessing methods have their own limitations in these two aspects. The wavelet analysis method is used to preprocess the NIR data of apples. The Peak Signal to Noise Ratio (PSNR) and Normalized Correlation (NC) coefficient are selected as the evaluation index. Compared with the common Savitzky-Golay smoothing and multiple scattering correction, the wavelet method not only can compress spectral data effectively, but also has its superiority in noise removal and spectral detail preserving.

Key words: near infrared spectroscopy; data preprocessing; wavelet analysis; Savitzky-Golay smoothing; multiplicative scatter correction

0 引言

近红外光谱 (Near Infrared Reflectance Spectroscopy, NIRS) 分析技术能够快速、高效、准确

地对有机物样品进行物理性质、化学性质和力学性质等的无损检测, 是一项新的无损检测技术, 正在全球范围内受到广泛的重视和推广^[1-3]。

收稿日期: 2016-03-29

基金项目: 四川省教育厅重点项目 (12ZA070)

作者简介: 张小英 (1972-), 女, 四川仁寿人, 硕士, 主要从事电路与系统、近红外光谱分析和小波分析等研究。

E-mail: zhxy_417@163.com

由仪器采集的原始近红外光谱数据除了包含反映样品组分的信息外,也不可避免地包含由高频随机噪音、基线漂移、信号本底、样品不均匀和光散射等产生的噪音信号^[4]。这些噪音信号会对谱图信息产生干扰,影响校正模型的建立,从而降低未知样品组分和性质预测的准确性。另一方面,在近红外光谱分析中,光谱数据一般比较大,需要较大的存储空间和较长的建模时间^[4]。因此要通过近红外光谱分析样品的物质成分,应该先进行光谱预处理。预处理主要有两个任务,一是降噪,减少各种干扰的影响,提高模型的稳健性和预测结果的准确性;二是数据压缩,以便于数据存储和提高建模速度。

常用的近红外光谱数据预处理方法主要有光谱数据的 Savitzky-Golay 卷积平滑和多元散射校正 (Multiplicative Scatter Correction, MSC) 等。Savitzky-Golay 卷积平滑法是近红外光谱信号去噪中最常用的方法。该方法假设光谱中的噪声是白噪声 (均值 0), 通过平滑滤波可以提高信噪比, 同时也会导致信号失真, 例如使突变信号变得相对平缓^[5]。MSC 法由 Martens 等人提出, 该方法主要用于消除样品颗粒分布不均匀及颗粒大小不同产生的散射对近红外光谱的影响, 即主要消除理想的线性散射影响。在样品吸收度与浓度保持较好的线性关系情况下, MSC 的校正效果较好; 但在很多实际情况样品颗粒的散射对近红外光谱的影响并不成线性, 所以 MSC 方法有待改进^[6]。

小波分析具有多分辨率分析的特点, 在时域和频域都具有表征信号局部特征的能力, 被誉为“数学显微镜”, 是信号处理领域中继傅立叶分析之后的一种新方法^[5]。本文将小波分析用于苹果近红外光谱数据的预处理。与常用的 Savitzky-Golay 平滑滤波和多元散射校正相比较, 小波方法在噪声去除和光谱细节信号的保持以及数据

压缩方面都具有其独到之处。

1 用小波分析预处理近红外光谱数据

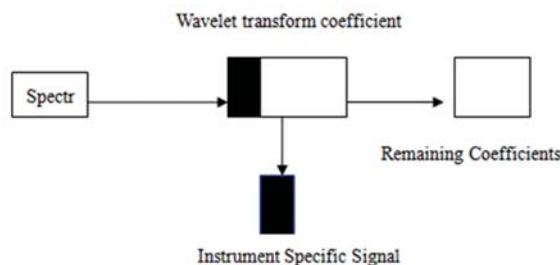


图 1 近红外光谱的小波变换

将小波变换用于近红外光谱数据预处理的主要作用是滤噪和数据压缩。通过小波变换, 近红外光谱数据可以被分解为背景信息 (低频)、组分信息 (中频) 和噪音信息 (高频) 部分。通过对低频和高频系数的适当处理, 可以有效地去除背景信息和噪音信息, 达到去噪和数据压缩的目的, 为后续分析提高建模的稳健性和预测的准确性。小波分析用于近红外光谱数据预处理的步骤如下:

(1) 小波分解: 通过实验反复筛选确定一种小波基函数和分解层数 N , 然后对原始光谱信号进行 N 层小波分解。

(2) 系数处理: 选择软阈值方式, 阈值由式 (1) 确定, 然后对第一层到第 N 层的高频系数进行软阈值量化处理。根据先验知识, 删除低频系数中足够小而被认为不代表有用信息的系数。

$$T(Y, t) = \begin{cases} \operatorname{sgn}(Y)(|Y| - t) & |Y| \geq t \\ 0 & |Y| < t \end{cases} \quad (1)$$

式中, sgn 是符号函数, t 是阈值, Y 是高频小波系数, T 是经软阈值处理后的小波系数。

(3) 小波重构: 根据处理后的第 N 层低频系数和经量化处理后的第一层到第 N 层的高频系数进行光谱信号的重构。

2 预处理评价指标

2.1 PSNR

本文引用 $PSNR$ 来评价预处理方法的去噪能力。 $PSNR$ 越大, 去噪能力越强, 预处理效果就越好, 其定义为

$$PSNR = 10 \times \log_{10} \left(\frac{(I_{max} - I_{min})^2}{MSE} \right)$$

$$= 10 \times \log_{10} \left(\frac{(255 - 0)^2}{MSE} \right) \quad (2)$$

$$MSE = \frac{1}{M \times M} \sum_{x=1}^M \sum_{y=1}^M [I'(x, y) - I(x, y)]^2 \quad (3)$$

式中, $I(x, y)$ 为原始光谱数据矩阵, $I'(x, y)$ 为预处理后的光谱数据矩阵, 大小都是 30×1557 ; MSE 称为均方误差; I_{max} 和 I_{min} 分别为光谱数据的最大值和最小值。

2.2 NC

本文采用 NC 来评价预处理前后光谱图像之间的相似程度。 NC 越大, 预处理后保留的光谱细节信息就越多, 预处理后的光谱和原始光谱就越相似。 NC 的定义为

$$NC = \frac{\sum_{x=1}^N \sum_{y=1}^N W(x, y) W'(x, y)}{\sum_{x=1}^N \sum_{y=1}^N [W(x, y)]^2} \quad (4)$$

式中, $W(x, y)$ 为原始光谱数据矩阵, $W'(x, y)$ 为预处理后的光谱数据矩阵。

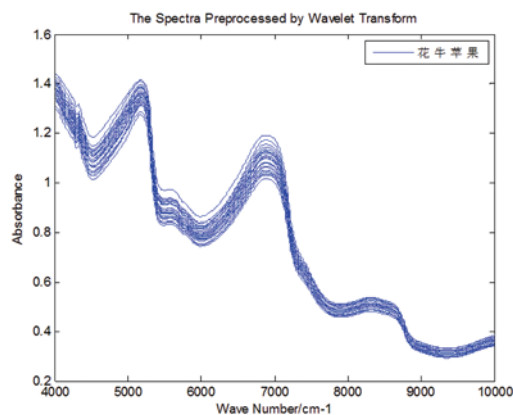
3 实验分析

选用大小均匀、无损伤的花牛苹果 (产地甘肃)、普通红富士苹果 (产地陕西) 和特级红富士苹果 (产地山东) 各 30 个组成实验样本。苹果样本在温度为 $20^\circ\text{C} \sim 25^\circ\text{C}$ 的实验室内存放 12 h。实验所用的光谱采集器是赛默飞世尔 (Thermo Fisher) 公司的 Antaris II 近红外光谱分析仪。Antaris II 近红外光谱分析仪开机时需预热 1 h, 采用反射积分球模式采集苹果的近红外光谱。每个样品被扫描 32 次, 以获取样品的漫反射光谱均值。光谱扫描的波数为 $4000 \sim 10000 \text{ cm}^{-1}$, 扫描间隔为 3.856 cm^{-1} , 对每个样品采集一个 1557 维的光谱数据。为减少误差, 每个苹果样本沿赤道轨迹采样 3 次, 取其平均值作为最终的试验数据。三类

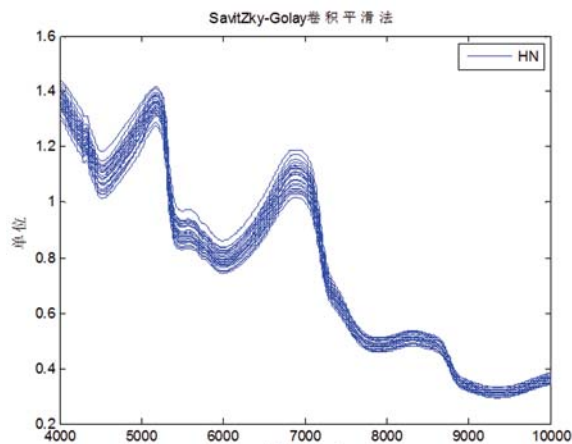
苹果的原始近红外光谱图如图 2(a)、图 3(a) 和图 4(a) 所示。

本文对三类苹果的近红外光谱数据分别采用 Savitzky-Golay 卷积平滑法、多元散射校正和小波分析方法进行预处理, 结果如图 2 和图 3 所示。Savitzky-Golay 卷积平滑法选用滤波的带宽为 9, 小波分析时选用 db3 小波对苹果光谱数据进行 3 层分解, 根据噪声强度采用软阈值方法分离信号和噪声。

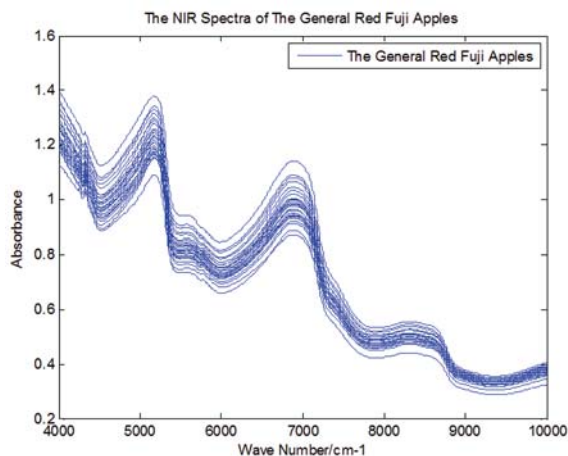
从图 2~图 4 可见, 经 Savitzky-Golay 卷积平滑法和小波分析法预处理后的光谱图像和原始光谱图像保持了很高的相似性, 而经 MSC 方法预处理后的光谱图像和原始光谱图像的差异较大。表 1 列出了三种预处理方法对三类苹果光谱预处理后的峰值信噪比 $PSNR$, 表 2 列出了三种预处理方法对三类苹果光谱进行预处理后的归一化相关系数 NC 。可见, 对三类苹果的光谱进行预处理后, 小波分析方法得到的 $PSNR$ 值都是最大的, Savitzky-Golay 平滑法的次之, MSC 方法的最小, 小波分析方法表现出最强的去噪能力。对三类苹果的光谱进行预处理后, Savitzky-Golay 平滑法和小波分析法都能保持 NC 值为 1, 表明这两种预处理方法能较好地保持光谱数据的相似性, 即都能很好地保留光谱的细节信息。而 MSC 方法在相似性保持方面要差一些。



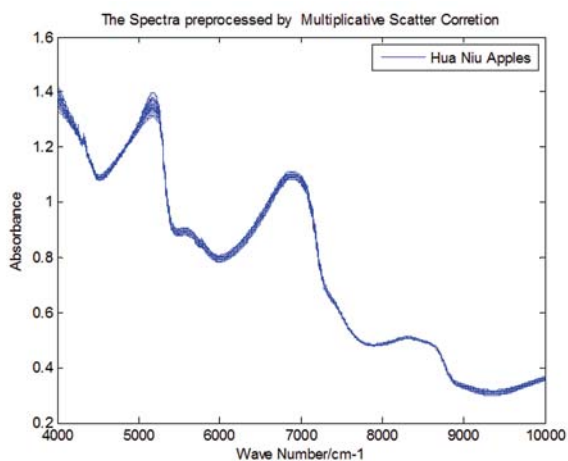
(a) 花牛苹果的近红外光谱



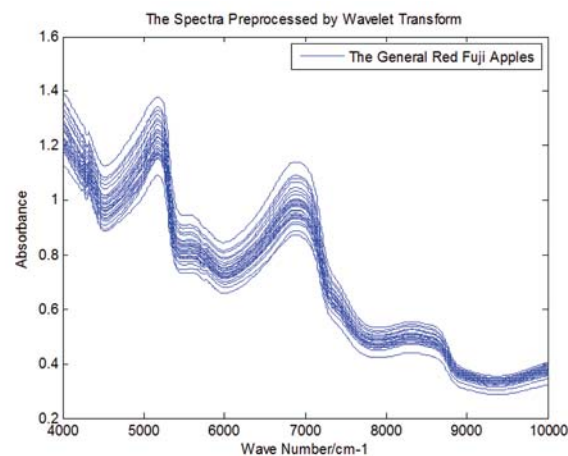
(b) Savitzky-Golay 卷积法预处理后的光谱



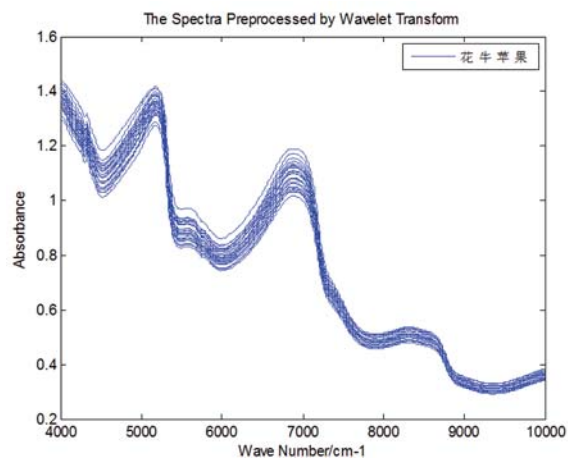
(a) 普通红富士苹果的近红外光谱



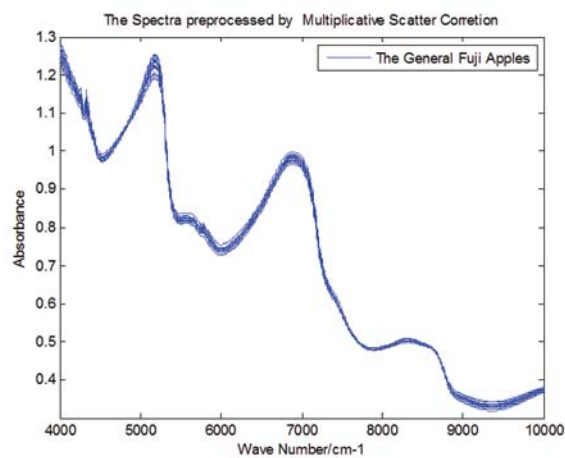
(c) 经 MSC 预处理的光谱



(b) 经 Savitzky-Golay 卷积法预处理后的光谱

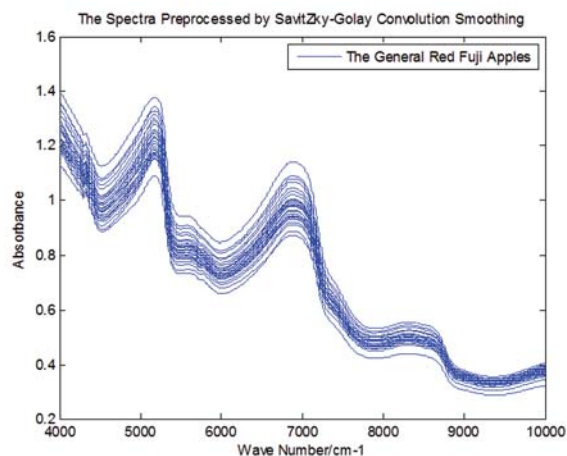


(d) 经小波分析法预处理后的光谱



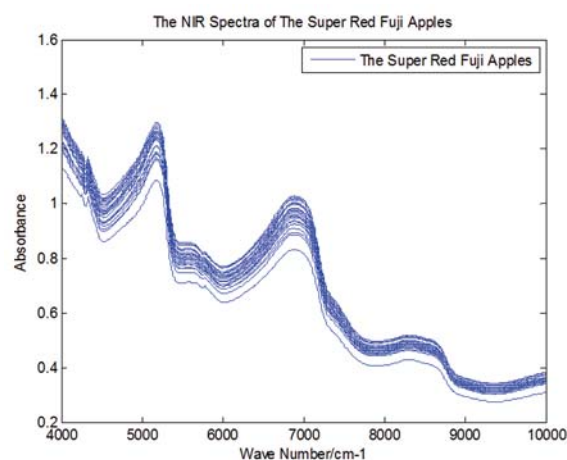
(c) 经 MSC 预处理后的光谱

图 2 花牛苹果近红外光谱的预处理

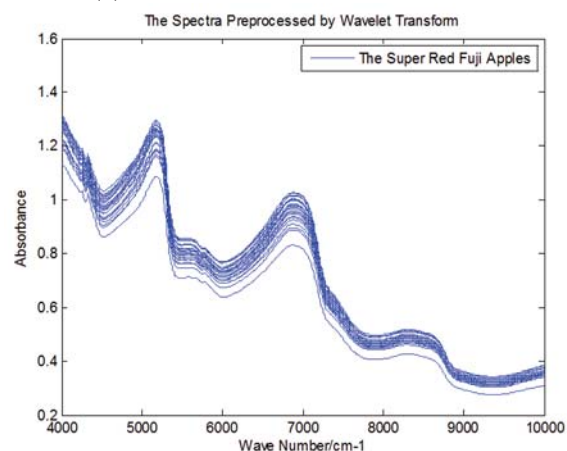


(d) 经小波分析法预处理后的光谱

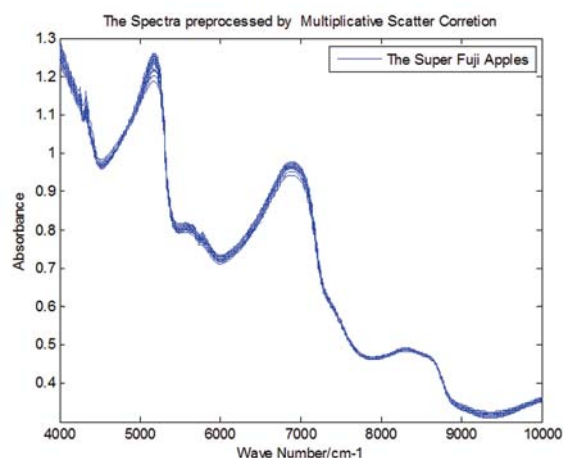
图 3 普通红富士苹果近红外光谱的预处理



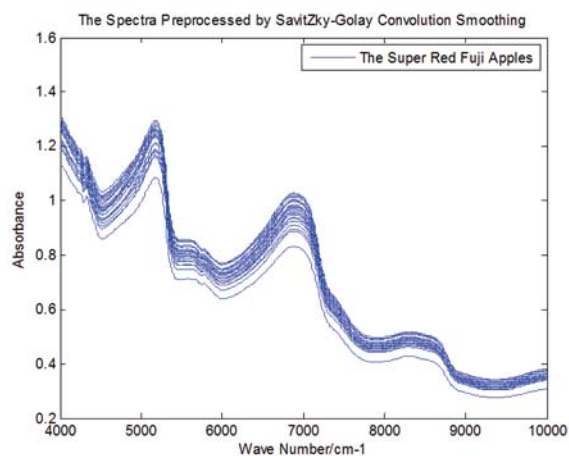
(a) 特级红富士苹果的近红外光谱



(b) 经 Savitzky-Golay 卷积法预处理后的光谱



(c) 经 MSC 法预处理后的光谱



(d) 经小波分析法预处理后的光谱

图 4 特级红富士苹果近红外光谱的预处理

表 1 三种预处理方法的 PSNR 比较

预处理 算法	SavitZky-Golay 平滑法	MSC 方法	小波分析 方法
花牛苹果	4.9486	3.0544	68.0728
普通红富士	68.4468	30.5774	78.8187
特级红富士	68.1974	32.5608	77.9999

表 2 三种预处理方法的 NC 比较

预处理 算法	SavitZky-Golay 平滑法	MSC 方法	小波分析 方法
花牛苹果	0.9995	0.4040	1
普通红富士	1	0.9986	1
特级红富士	1	0.9992	1

综上分析, 小波分析方法不仅能有效地实

现数据压缩,而且在噪声去除和光谱细节保持两方面也表现出优越性。总的来说,在苹果近红外光谱数据的预处理中,小波分析方法优于 Savitzky-Golay 平滑法和 MSC 方法。

4 结束语

将小波变换用于苹果近红外光谱的预处理,一方面能有效地去除噪声,提高后续建模的预测精度和稳健性;另一方面能有效地实现数据的压缩,减少数据存储空间和提高后续建模的速度。NIR 分析技术受样本散射性质、物理性质和共线性等多种因素的影响,具体应用时应根据不同的分析体系以及不同的分析目的选择不同的预处理方法,甚至可以结合选用几种不同的预处理方法。然而究竟选择哪些不同的预处理方法组合以及如何安排这些方法的先后运算顺序,还有待于我们在实践中不断探索和优化。

参考文献

- [1] 张银,周孟然. 近红外光谱分析技术的数据处理方法[J].**红外技术**,2007, **29**(06):345-348.
- [2] Via B K, Shupe T F, Stine M, et al. Tracheid Length Prediction in Pinus Palustris by Means of Near Infrared Spectroscopy: the Influence of Age [J].*European journal of wood and wood products*,2005,**63**(3):231-236.
- [3] Jones P D, Schimleck L R, Gary F P, et al. Non-destructive Estimation of Wood Chemical Composition of Radial Strips by Diffuse Reflectance Near Infrared Spectroscopy[J].*Wood Science and Technology*, 2005,**39**(7):529-534.
- [4] 高荣强,范世福,严衍禄,等. 近红外光谱的数据预处理研究 [J].**光谱学与光谱分析**,2004,**24**(12):1563-1565.
- [5] 王立琦,朱秀超,张礼勇. 基于小波分析的近红外光谱数据预处理 [J].**哈尔滨商业大学学报(自然科学版)**,2009, **25**(06):700-702.
- [6] 尼珍,胡昌勤,冯芳. 近红外光谱分析中光谱预处理方法的作用及其发展 [J].**药物分析杂志**,2008,**28**(5):824-829.
- [7] 田高友,袁洪福,刘慧颖,等. 小波变换在近红外光谱分析中得应用进展 [J].**光谱学与光谱分析**,2003, **23**(06):1111-1114.

(上接第 42 页)

层、本征层的厚度为 2 nm、背场的厚度为 5 nm、背场的掺杂浓度为 $3 \times 10^{21} \text{ cm}^{-3}$ 时,太阳能电池的参数 $V_{OC}=647.7 \text{ mV}$, $J_{SC}=42.29 \text{ mA} \cdot \text{cm}^{-2}$, $FF=75.32\%$, $EFF=20.63\%$ 。

参考文献

- [1] 侯国付. 新型薄膜窄带隙光伏材料 $\beta\text{-FeSi}_2$ 的研究进展 [J].**激光与光电子学进展**,2009(08):61-66.
- [2] 许文英,陈凤翔,王嘉赋. $n\text{-}\beta\text{-FeSi}_2/p\text{-Si}$ 异质结太阳能电池的计算机模拟 [J].**武汉理工大学学报**,2013,**35**(09):21-24.
- [3] 杨大洋,刘淑平,张棚,等. 不同工作温度下 $a\text{-Si(n)}/c\text{-Si(p)}/uc\text{-Si(p}^+)$ 异质结太阳能电池微晶硅背场的模拟计算与优化 [J].**红外**, 2013, **34**(8):40-46.
- [4] 靳瑞敏. **太阳能电池原理与应用**[M]. 北京大学出版社, 2011.
- [5] 王敏花,刘晓平,李彦林,等. AFORS-HET 软件模拟 N 型非晶硅 /p 型晶体硅异质结太阳能电池 [J].**太阳能学报**,2008,**29**(02):125-129.
- [6] 刘艳红,刘爱民. 带有本征薄层的异质结太阳能电池 [J].**半导体技术**, 2010, **35**(1):1-7.
- [7] 赵雷,周春兰,李海玲,等. $a\text{-Si(n)}/c\text{-Si(p)}$ 异质结太阳能电池薄膜硅背场的模拟优化 [J].**物理学报**, 2008,**57**(5):3212-3218.
- [8] 王东. **光伏电池原理及应用**[M]. 化学工业出版社, 2014.
- [9] 吴国盛,王振文,闻腾,等. $a\text{-Si(n)}/c\text{-Si(p)}$ 异质结电池非晶层的模拟优化 [J].**科技创新与生产力**,2012(10):87-90.
- [10] Wakisaka K, Taguchi M, Sawada T, et al. More than 16% Solar Cells with a New HIT (Doped $a\text{-Si}/$ non-doped $a\text{-Si}/$ crystalline Si) Structure [C].*Conference Record the 22th IEEE PVSC*,1991 :887-892.
- [11] Dwivedi N, Kumar S, Bisht A, et al. Simulation Approach for Optimization of Device Structure and Thickness of HIT Solar Cells to Achieve 27% Efficiency[J].*Solar Energy*, 2013, **88**:31-41.