



(12) 发明专利申请

(10) 申请公布号 CN 103854305 A

(43) 申请公布日 2014. 06. 11

(21) 申请号 201410104520. 7

(22) 申请日 2014. 03. 19

(71) 申请人 天津大学

地址 300072 天津市南开区卫津路 92 号

(72) 发明人 陈达 李奇峰 卢帆

(74) 专利代理机构 天津市北洋有限责任专利代

理事务所 12201

代理人 温国林

(51) Int. Cl.

G06T 17/00(2006. 01)

G01N 21/25(2006. 01)

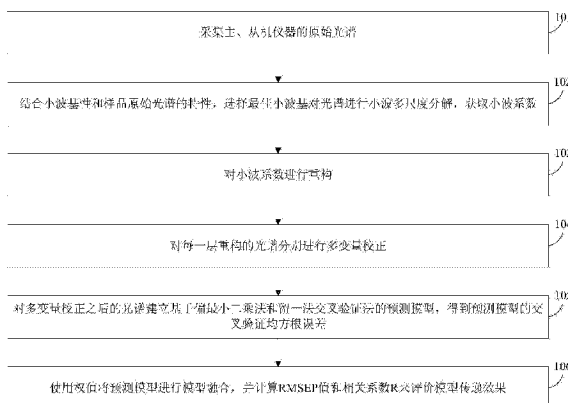
权利要求书2页 说明书5页 附图8页

(54) 发明名称

一种基于多尺度建模的模型传递方法

(57) 摘要

一种基于多尺度建模的模型传递方法,该方法包括以下步骤:采集主、从机仪器的原始光谱;结合小波基性和样品原始光谱的特性,选择最佳小波基对光谱进行小波多尺度分解,获取小波系数;对小波系数进行重构;对每一层重构的光谱分别进行多变量校正;对多变量校正之后的光谱建立基于偏最小二乘法和留一法交叉验证法的预测模型,得到预测模型的交叉验证均方根误差;使用权值将预测模型进行模型融合,并计算预测均方根误差和相关系数来评价模型传递效果。本方法相较于常规模型传递方法,极大的提高了模型传递的效率和性能,能广泛应用于近红外和拉曼光谱等领域。



1. 一种基于多尺度建模的模型传递方法,其特征在于,所述方法包括以下步骤:

- (1) 采集主、从仪器的原始光谱;
- (2) 结合小波基性和样品原始光谱的特性,选择最佳小波基对光谱进行小波多尺度分解,获取小波系数;
- (3) 对小波系数进行重构;对每一层重构的光谱分别进行多变量校正;
- (4) 对多变量校正之后的光谱建立基于偏最小二乘法和留一法交叉验证法的预测模型,得到预测模型的交叉验证均方根误差;
- (5) 使用权值将预测模型进行模型融合,并计算预测均方根误差和相关系数来评价模型传递效果。

2. 根据权利要求1所述的一种基于多尺度建模的模型传递方法,其特征在于,所述交叉验证均方根误差具体为:

$$RMSECV = \sqrt{\frac{\sum_1^n (C_{NIR} - C_{REF})^2}{\sum_1^n C_{REF}^2}}$$

式中:RMSECV为交叉验证均方根误差;n为样品数; C_{NIR} 是样品的某一实际属性; C_{REF} 为预测出的样品属性。

3. 根据权利要求1所述的一种基于多尺度建模的模型传递方法,其特征在于,所述预测均方根误差具体为:

$$RMSEP = \sqrt{\frac{\sum_1^n (C_{NIR} - C_{REF})^2}{n}}$$

式中:RMSEP为预测均方根误差;n是样品数, C_{NIR} 是样品的某一实际属性; C_{REF} 为预测出的样品属性。

4. 根据权利要求1所述的一种基于多尺度建模的模型传递方法,其特征在于,所述相关系数具体为:

$$R = \frac{\sum_1^n C_{NIR} C_{REF} - \frac{\sum_1^n C_{NIR} \sum_1^n C_{REF}}{n}}{\sqrt{\left(\sum_1^n C_{NIR}^2 - \frac{(\sum_1^n C_{NIR})^2}{n}\right) \left(\sum_1^n C_{REF}^2 - \frac{(\sum_1^n C_{REF})^2}{n}\right)}}$$

式中:R为相关系数;n是样品数, C_{NIR} 是样品的某一实际属性; C_{REF} 为预测出的样品属性。

5. 根据权利要求1所述的一种基于多尺度建模的模型传递方法,其特征在于,所述使用权值将预测模型进行模型融合的操作具体为:

$$W_i = \frac{\frac{1}{(RMSECV_i)^2}}{\sum_1^m \frac{1}{(RMSECV_i)^2}}$$

根据权值 W_i 将预测模型融合在一起,公式如下:

$$C = \sum_i^m W_i \cdot C_{iREF}$$

其中, C_{iREF} 是预测模型的预测结果, m 是分解的尺度, C 为预测模型融合后的预测结果, 即最终的模型传递结果。

一种基于多尺度建模的模型传递方法

技术领域

[0001] 本发明涉及化学计量学领域,尤其涉及一种基于多尺度建模的模型传递方法。

背景技术

[0002] 多元校正模型是近红外光谱定量分析的基础,建立一个长期稳健、可靠、准确的数学模型是非常复杂的过程,需要消耗大量的时间、人力、物力和财力。由于模型中包含着建模样品的理化性质、装样条件、仪器状态以及环境因素等信息,因此必须对该模型能否适用于新样品或在新仪器上测定的光谱进行评价和研究,如果模型不适用,也就是说,模型在预测过程中所遇到的样本信息并非在校正过程中所包含的变异,导致预测结果偏差较大,那么就需要进行模型的传递。

[0003] 模型传递,也称为仪器标准化,广义的模型传递泛指在原条件下建立的模型能够应用于新的测试条件,包括在不同的样品状态、环境条件和仪器状态等测试系统之间进行传递;狭义的模型传递一般是指使用数学方法对不同光谱仪之间的差异进行修正。本申请研究的模型传递通常指的是狭义的模型传递。目前,常用的模型传递方法主要有对模型系数 b 的校正、斜率/偏差(S/B)算法、直接校正(DS)算法、分段直接标准化(PDS)算法、Shenk 专利算法和目标因子法等,其中应用最为普遍和最成功的是多变量校正方法,它包括 DS 算法和 PDS 算法等。

[0004] 众所周知,光谱本身就是多尺度的,其包含的信息也是多尺度的。不同光谱仪的光谱之间的差异,可能只存在于一部分频率域上,其他部分的信息则是相同的。而常规的多变量校正方法是在对整个光谱层面上进行校正,没有在频率域对光谱进行划分,这样的模型传递很笼统,不够精细,精度也不够,性能自然也不好。

发明内容

[0005] 本发明提供了一种基于多尺度建模的模型传递方法,本发明通过小波多尺度建模和多变量校正相结合,对光谱的频率域进行精细划分,对有差异的频率域就、进行多变量校正,从而大幅度的提高模型传递的效率和性能,详见下文描述:

[0006] 一种基于多尺度建模的模型传递方法,所述方法包括以下步骤:

[0007] (1) 采集主、从机仪器的原始光谱;

[0008] (2) 结合小波基性和样品原始光谱的特性,选择最佳小波基对光谱进行小波多尺度分解,获取小波系数;

[0009] (3) 对小波系数进行重构;对每一层重构的光谱分别进行多变量校正;

[0010] (4) 对多变量校正之后的光谱建立基于偏最小二乘法和留一法交叉验证法的预测模型,得到预测模型的交叉验证均方根误差;

[0011] (5) 使用权值将预测模型进行模型融合,并计算预测均方根误差和相关系数来评价模型传递效果。

[0012] 所述交叉验证均方根误差具体为:

$$[0013] \quad \text{RMSECV} = \sqrt{\frac{\sum_1^n (C_{\text{NIR}} - C_{\text{REF}})^2}{\sum_1^n C_{\text{REF}}^2}}$$

[0014] 式中:RMSECV 为交叉验证均方根误差;n 为样品数, C_{NIR} 是样品的某一实际属性; C_{REF} 为预测出的样品属性。

[0015] 所述预测均方根误差具体为:

$$[0016] \quad \text{RMSEP} = \sqrt{\frac{\sum_1^n (C_{\text{NIR}} - C_{\text{REF}})^2}{n}}$$

[0017] 式中:RMSEP 为预测均方根误差;n 是样品数, C_{NIR} 是样品的某一实际属性; C_{REF} 为预测出的样品属性。

[0018] 所述相关系数具体为:

$$[0019] \quad R = \frac{\sum_1^n C_{\text{NIR}} C_{\text{REF}} - \frac{\sum_1^n C_{\text{NIR}} \sum_1^n C_{\text{REF}}}{n}}{\sqrt{\left(\sum_1^n C_{\text{NIR}}^2 - \frac{(\sum_1^n C_{\text{NIR}})^2}{n}\right) \left(\sum_1^n C_{\text{REF}}^2 - \frac{(\sum_1^n C_{\text{REF}})^2}{n}\right)}}$$

[0020] 式中:R 为相关系数;n 是样品数, C_{NIR} 是样品的某一实际属性; C_{REF} 为预测出的样品属性。

[0021] 所述使用权值将预测模型进行模型融合的操作具体为:

$$[0022] \quad W_i = \frac{\frac{1}{(\text{RMSECV}_i)^2}}{\sum_i^m \frac{1}{(\text{RMSECV}_i)^2}}$$

[0023] 根据权值 W_i 将预测模型融合在一起,公式如下:

$$[0024] \quad C = \sum_i^m W_i \cdot C_{i\text{REF}}$$

[0025] 其中, $C_{i\text{REF}}$ 是预测模型的预测结果,m 是分解的尺度,C 为预测模型融合后的预测结果,即最终的模型传递结果。

[0026] 本发明提供的技术方案的有益效果是:首先选择最佳小波基对光谱进行多尺度分解并重构;其次应用多变量校正对重构的每一层光谱进行校正;接着通过计算出的权值进行子模型融合;最后通过预测均方根误差和相关系数对模型传递的结果和性能进行评价。本方法相较于常规模型传递方法,极大的提高了模型传递的效率和性能,能广泛应用于近红外和拉曼光谱等领域。

附图说明

[0027] 图 1 (a) 玉米样品的 m5 仪器光谱;

[0028] 图 1 (b) 玉米样品的 mp5 仪器光谱;

[0029] 图 2 为 m5 和 mp5 仪器光谱二者的差异;

[0030] 图 3 (a) 不同小波基在不同分解尺度下的 RMSEP;

- [0031] 图 3 (b) 不同小波基在不同分解尺度下的相关系数；
 [0032] 图 4 (a) 第一层的模型传递后预测结果；
 [0033] 图 4 (b) 第二层的模型传递后预测结果；
 [0034] 图 4 (c) 第三层的模型传递后预测结果；
 [0035] 图 5 为子模型的交叉验证均方根误差；
 [0036] 图 6 为子模型的决策融合权重；
 [0037] 图 7 为一种基于多尺度建模的模型传递方法的流程图；
 [0038] 图 8 (a) 常规 DS 方法模型传递的结果；
 [0039] 图 8 (b) 基于多尺度建模的模型传递方法的结果。

具体实施方式

[0040] 为使本发明的目的、技术方案和优点更加清楚，下面对本发明实施方式作进一步地详细描述。

[0041] 为了对光谱的频率域进行精细划分，对有差异的频率域进行多变量校正，从而大幅度的提高模型传递的效率和性能，本发明实施例提供了一种基于多尺度建模的模型传递方法，详见下文描述：

[0042] 101 :采集主、从机仪器的原始光谱；

[0043] 其中，采集原始光谱即：对同一样品在同一测量条件下采集不同仪器的光谱，主、从机光谱仪器需保证是同一类型号仪器。

[0044] 102 :结合小波基性和样品原始光谱的特性，选择最佳小波基对光谱进行小波多尺度分解，获取小波系数；

[0045] 其中，选择最佳小波基，是对小波基数学特性进行分析，得到具有对称性、紧支性、正交性及高阶消失矩的小波基函数，有 Daubechies 小波、Symlets 小波、Coiflets 小波等；通过计算信号重构的均方根误差来衡量小波基重构信号的能力。

[0046] 103 :对小波系数进行重构；

[0047] 其中，重构是对不同仪器的光谱分解后的每一层小波系数分别进行重构。

[0048] 104 :对每一层重构的光谱分别进行多变量校正；

[0049] 其中，多变量校正方法，是常规的模型传递方法，其中最常用的是直接标准化算法 (DS)，分段直接标准化算法 (PDS) 等。

[0050] 105 :对多变量校正之后的光谱建立基于偏最小二乘法和留一法交叉验证法的预测模型，得到预测模型的交叉验证均方根误差；

[0051] 其中，对多变量校正后的光谱进行 PLS 建模指的是用校正后的光谱和化学法测得的浓度矩阵，建立基于偏最小二乘法和留一法交叉验证法的预测模型，交叉验证均方根误差为 RMSECV，公式如下：

$$[0052] \quad \text{RMSECV} = \sqrt{\frac{\sum_1^n (C_{\text{NIR}} - C_{\text{REF}})^2}{\sum_1^n C_{\text{REF}}^2}}$$

[0053] 式中：n 是样品数， C_{NIR} 是样品的某一实际属性，一般为浓度； C_{REF} 为预测出的样品属性；sum 为求和。

[0054] 106 :使用权值将预测模型进行模型融合,并计算 RMSEP 值和相关系数 R 来评价模型传递效果。

[0055] 其中, RMSEP 为预测均方根误差,其公式如下:

$$[0056] \quad \text{RMSEP} = \sqrt{\frac{\sum_1^n (C_{\text{NIR}} - C_{\text{REF}})^2}{n}}$$

[0057] 相关系数为 R,其公式如下:

$$[0058] \quad R = \frac{\sum_1^n C_{\text{NIR}} C_{\text{REF}} - \frac{\sum_1^n C_{\text{NIR}} \sum_1^n C_{\text{REF}}}{n}}{\sqrt{\left(\sum_1^n C_{\text{NIR}}^2 - \frac{(\sum_1^n C_{\text{NIR}})^2}{n}\right) \left(\sum_1^n C_{\text{REF}}^2 - \frac{(\sum_1^n C_{\text{REF}})^2}{n}\right)}}$$

[0059] 其中,模型融合采取如下的方式:

[0060] 在每一层上对新旧仪器光谱进行多变量校正后,得到预测模型和预测均方根误差。根据每个预测模型的交叉验证均方根误差计算出融合的权值,公式如下:

$$[0061] \quad W_i = \frac{\frac{1}{(\text{RMSECV}_i)^2}}{\sum_1^m \frac{1}{(\text{RMSECV}_i)^2}}$$

[0062] 然后根据权值将子模型融合在一起,公式如下:

$$[0063] \quad C = \sum_i^m W_i \cdot C_{i\text{REF}}$$

[0064] 其中, $C_{i\text{REF}}$ 是子模型的预测结果, m 是分解的尺度, C 为模型融合后的预测结果,即最终的模型传递结果。

[0065] 下面结合具体实例和附图对本发明的基于多尺度建模的模型传递新方法做出详细说明。

[0066] (1) 使用化学方法测得玉米样品的湿度数据,采集样品的新旧仪器的原始光谱数据,结合小波基的数学特性和光谱的性质选择最佳小波基。

[0067] 采集光谱数据,是在同一测量条件下采集同一样品的光谱数据。新旧仪器是指同一型号,不同的两台仪器。

[0068] 图 1 (a)和图 1 (b)是两台不同近红外仪器 m5 和 mp5 对 80 组玉米样品在相同测量条件下测得的近红外光谱,设定 m5 光谱为主机光谱,mp5 光谱为从机光谱。图 2 是 m5 仪器光谱和 mp5 仪器光谱二者之间的差异。

[0069] 通过计算信号重构的均方根误差来衡量小波基重构信号的能力,最终选择 db 小波作为分析样品原始光谱的最佳小波基。

[0070] 其中,均方根误差公式如下:

$$[0071] \quad E = \sqrt{\frac{\sum_1^N |S_r(i) - S(i)|}{N}}$$

[0072] 式中: N 为采样点数, $S(i)$ 样品的原始光谱信号, $S_r(i)$ 为重构的光谱信号。

[0073] (2) 使用最佳小波基对原始光谱进行多尺度分解,并对每层小波系数进行重构。

[0074] 在使用最佳小波基对原始光谱进行多尺度分解,其分解层数与光谱自身长度有关,本发明实施例采用的是二进制离散小波,因此分解层数与 2 的幂数有关,如果光谱长度为 2^n ,那么分解尺度不应超过 $n-1$,同时分解尺度也不应过小,会导致信号中的有用消息难以与干扰因素分离,因此设定分解尺度范围为 $\frac{n-1}{2} \sim n-1$ 。确定分解尺度范围后,使用 RMSEP 值和相关系数来确定最佳分解尺度并对分解后的小波系数进行重构。

[0075] 其中,图 3 (a) 是不同小波基在不同分解尺度下的 RMSEP,图 3 (b) 是不同小波基在不同分解尺度下的相关系数,综合上述两个指标,选择了最佳分解尺度为 8。

[0076] (3) 对每一层重构的光谱分别进行多变量校正,并得到交叉验证的均方根误差。

[0077] 多变量校正是常规的模型传递方法,其中最常用的是直接标准化算法(DS),分段直接标准化算法(PDS)等,这里采用直接标准化算法,使用转换矩阵 F 将标准样品在主、从机上测定的训练集光谱矩阵 S_m 和 S_s 进行关联,关系如下:

$$[0078] \quad S_m = S_s \cdot F$$

$$[0079] \quad F = S_s^{-1} \cdot S_m$$

[0080] 这样,对于待测样品在从机上测得的光谱 X_s 就可以通过转换矩阵转化后就实现了模型传递:

$$[0081] \quad X_{s, ds} = X_s \cdot F$$

[0082] 其中, S_m 是训练集的 m5 光谱, S_s 是训练集的 mp5 光谱, $X_{s, ds}$ 是校正后的光谱, X_s 是测试集的 mp5 光谱。

[0083] (4) 在每一个尺度上建立 m5 光谱和化学法测得的玉米样品湿度浓度矩阵的预测模型,然后使用校正后的 mp5 光谱通过建立好的预测模型去预测样品的湿度。

[0084] 图 4 (a) (b) (c) 是预测模型预测的结果,鉴于篇幅所限,只列举 3 个尺度的结果。

[0085] (5) 使用有效的融合策略将子模型融合到一起。

[0086] 计算出每层尺度上预测模型的交叉验证均方根误差,图 5 是就是所有预测模型的交叉验证均方根误差,使用交叉验证均方根误差计算每个预测模型的权重,图 6 是每个预测模型的权重。使用计算出的权重将预测模型进行决策级融合,图 7 是本发明整体步骤的流程图。图 8 (a) 是常规 DS 法模型传递的结果,图 8 (b) 是本发明模型传递的结果。

[0087] 由图可见,本发明通过小波变换将新旧仪器光谱进行多尺度分解,将每一层的小波系数分别进行重构,然后分别在每一层上对新旧光谱进行多变量校正,对校正后的光谱进行 PLS 建模,最后采取有效的融合手段将预测模型融合。本方法较常规的模型传递方法精度更高、性能更高,对于不同仪器间模型转移起到了重要作用,可广泛应用于多个光谱领域。

[0088] 本领域技术人员可以理解附图只是一个优选实施例的示意图,上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0089] 以上所述仅为本发明的较佳实施例,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

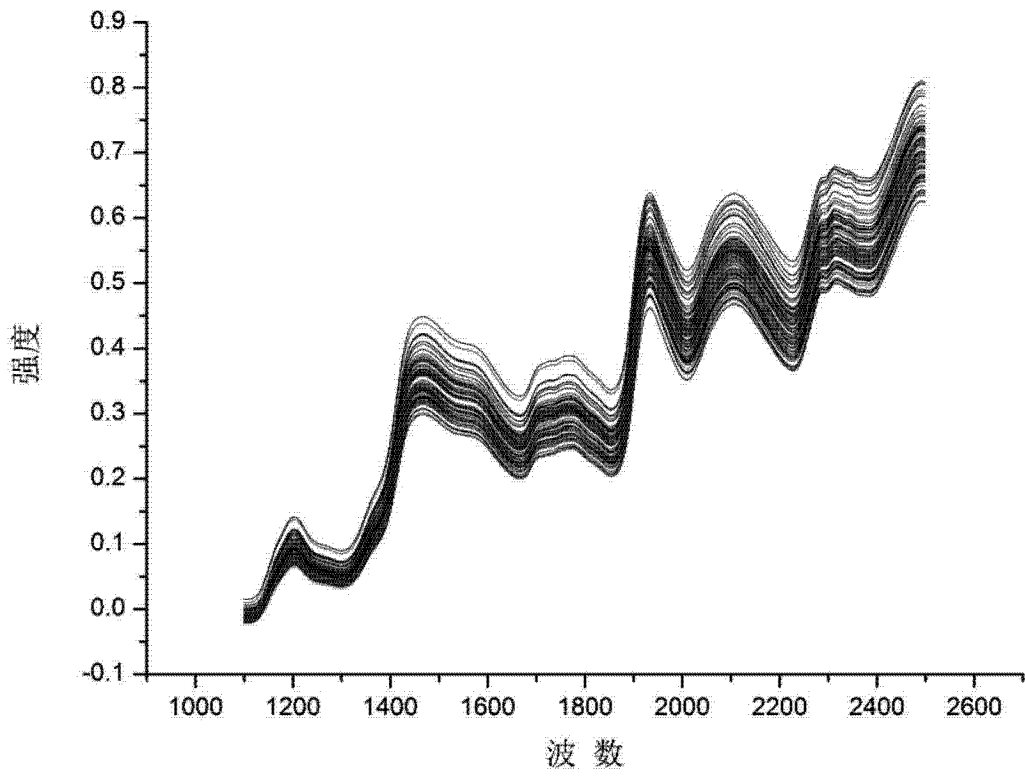


图 1 (a)

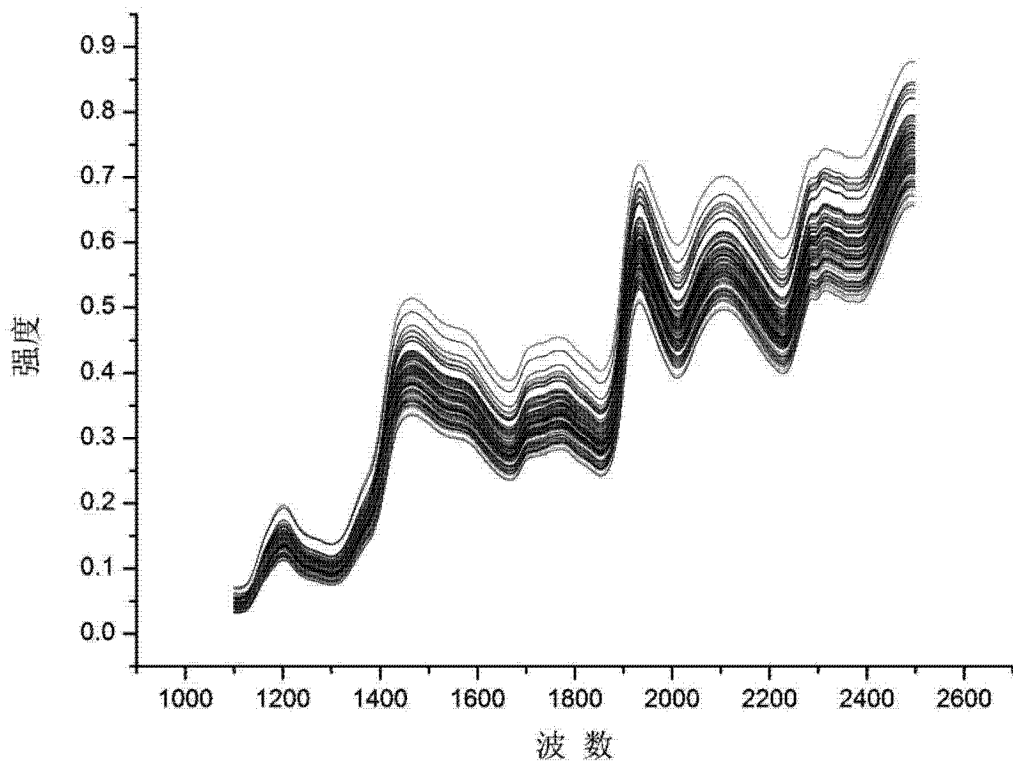


图 1 (b)

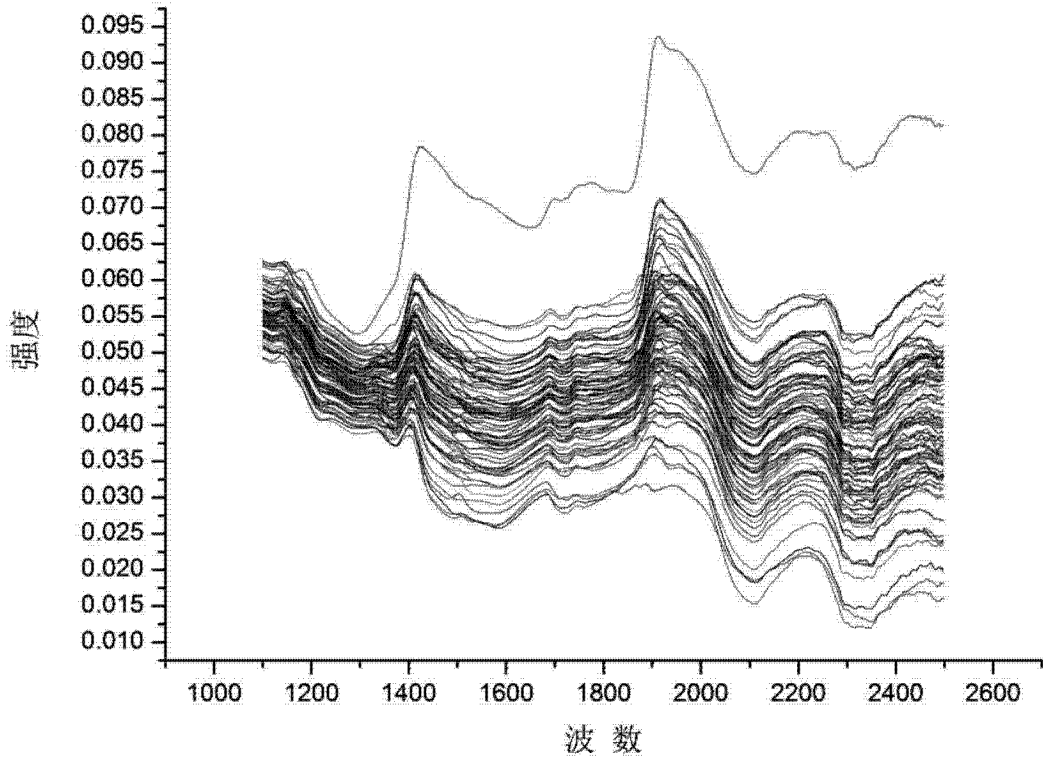


图 2

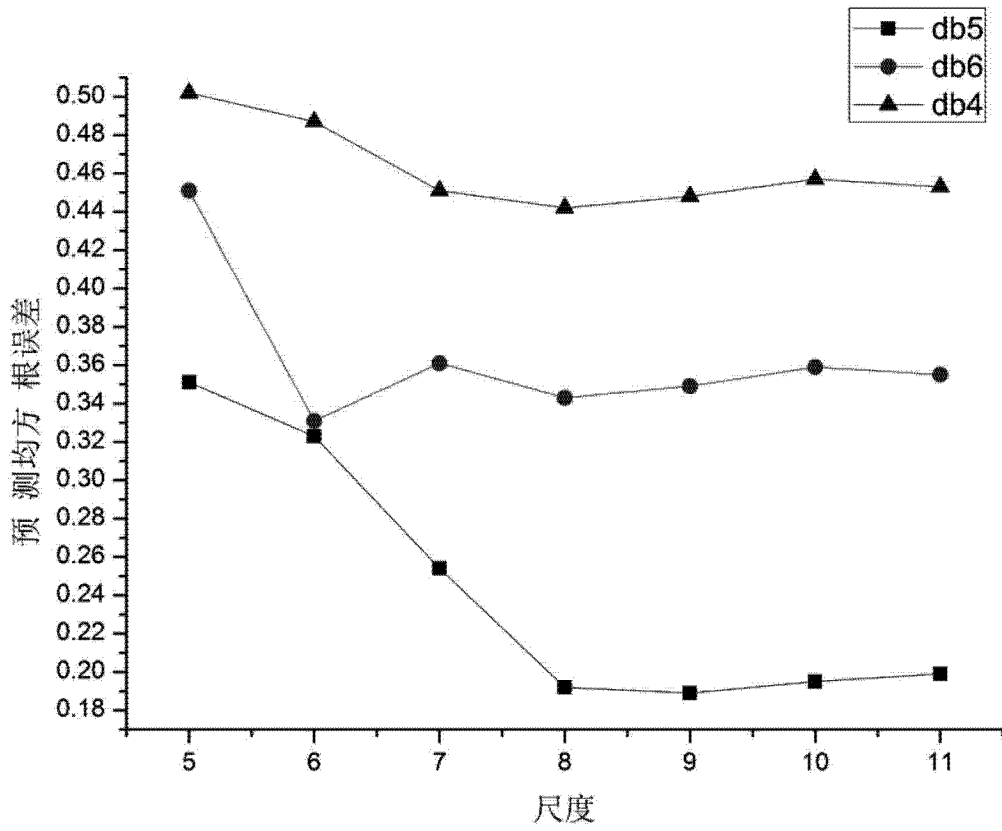


图 3 (a)

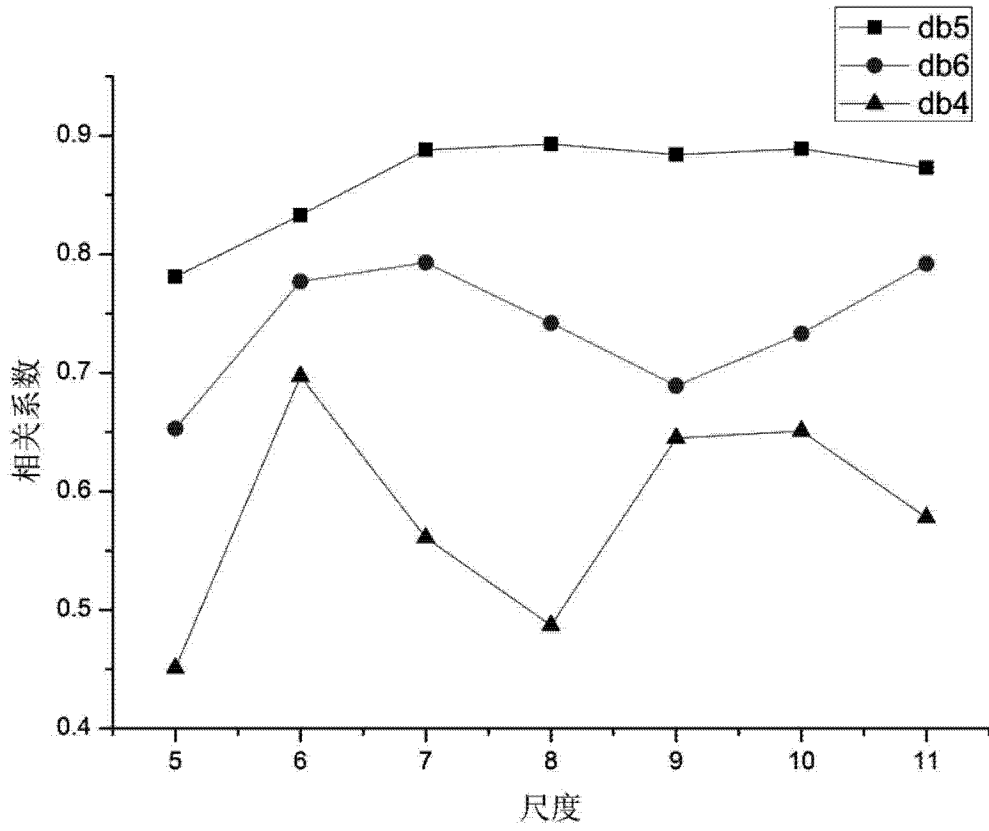


图 3 (b)

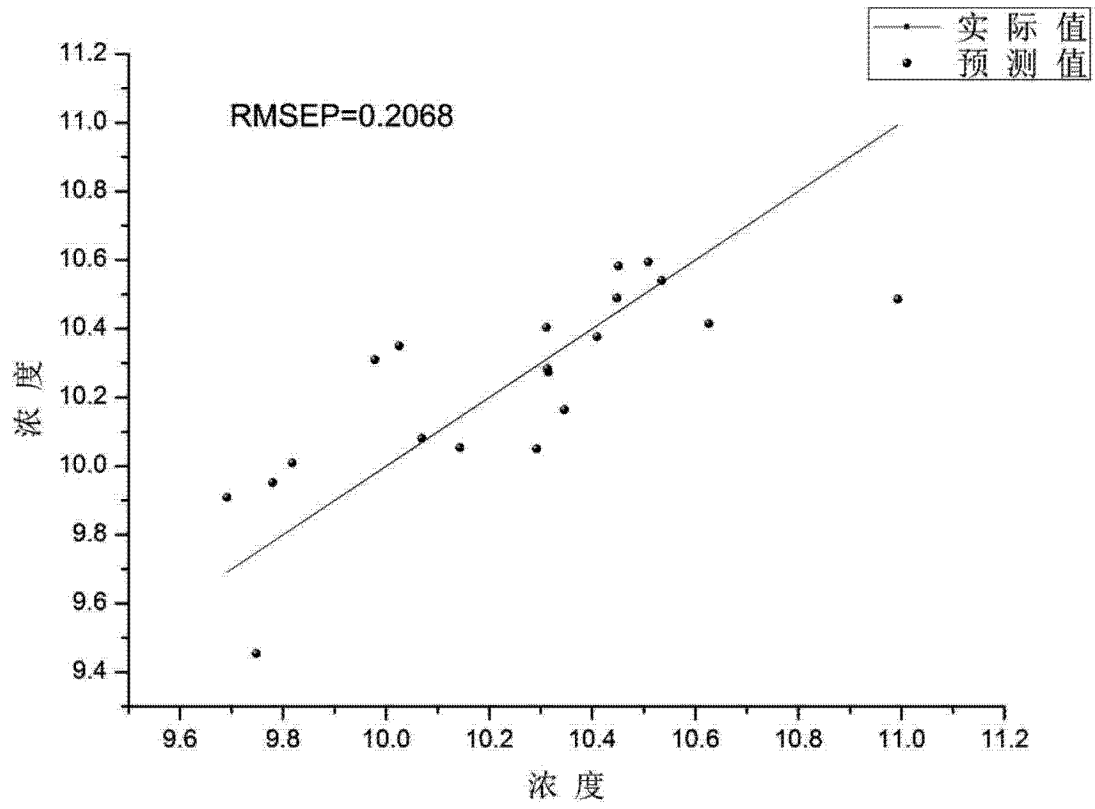


图 4 (a)

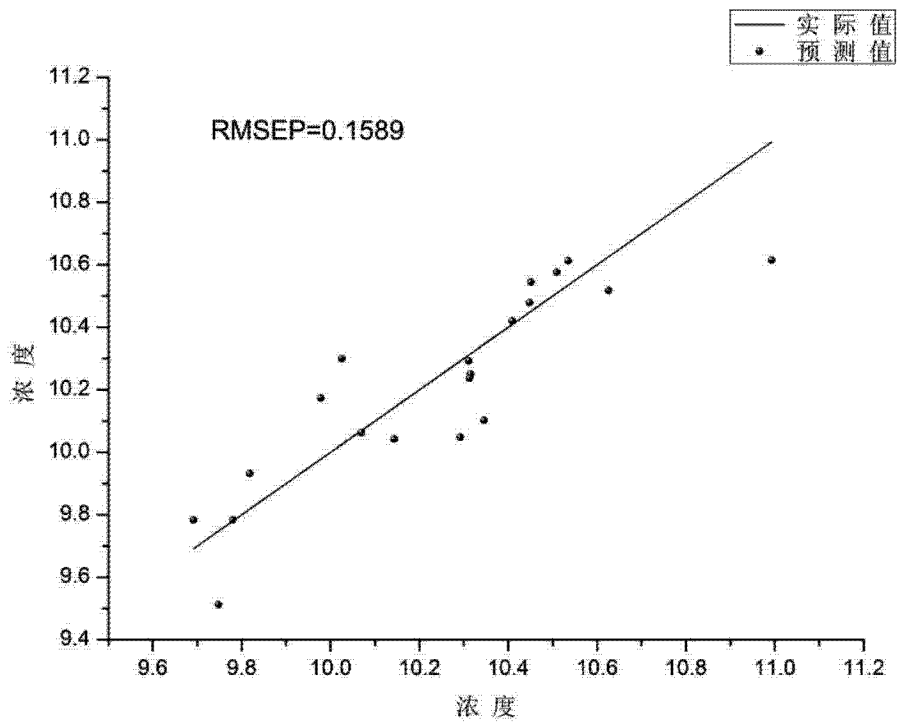


图 4 (b)

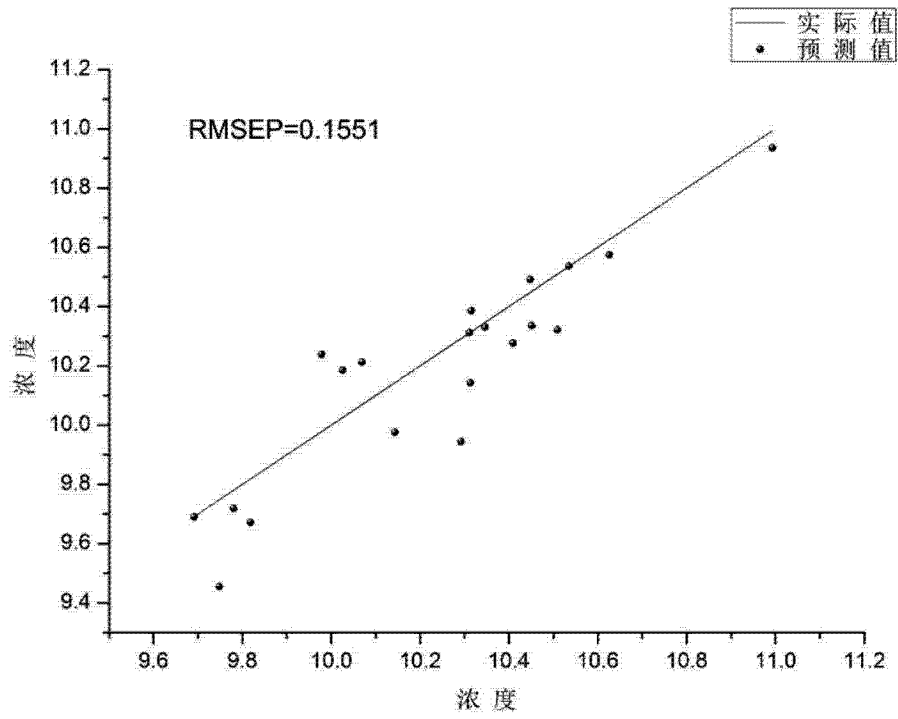


图 4 (c)

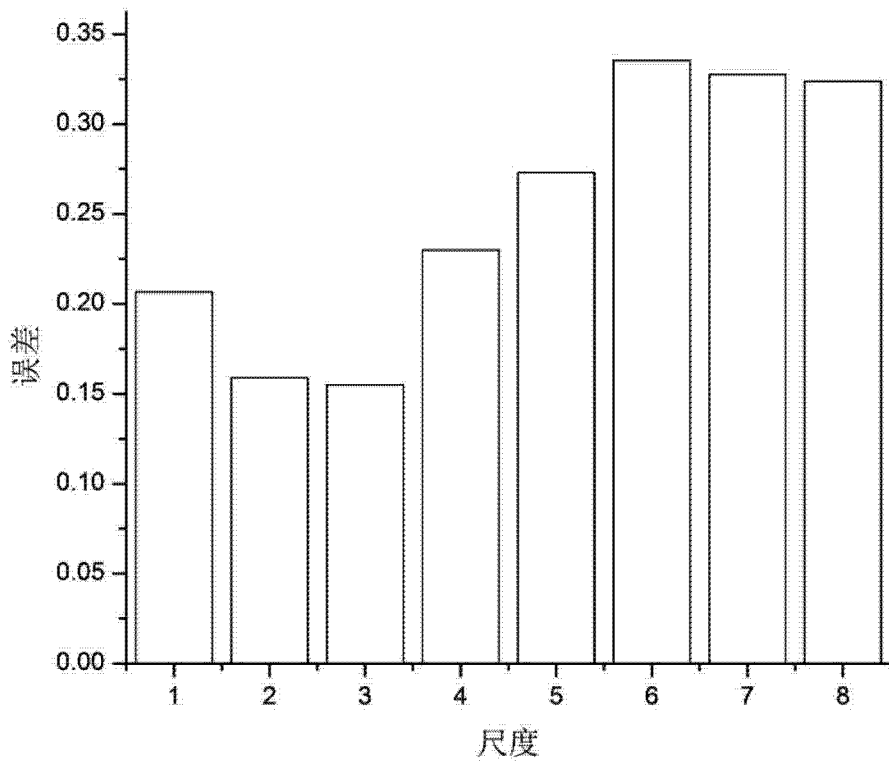


图 5

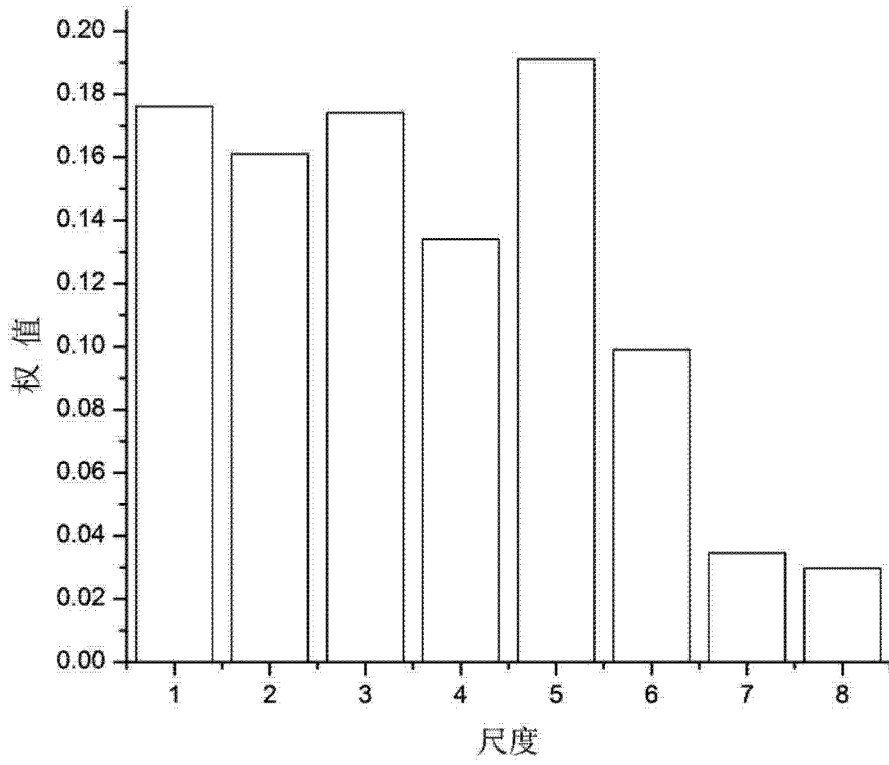


图 6

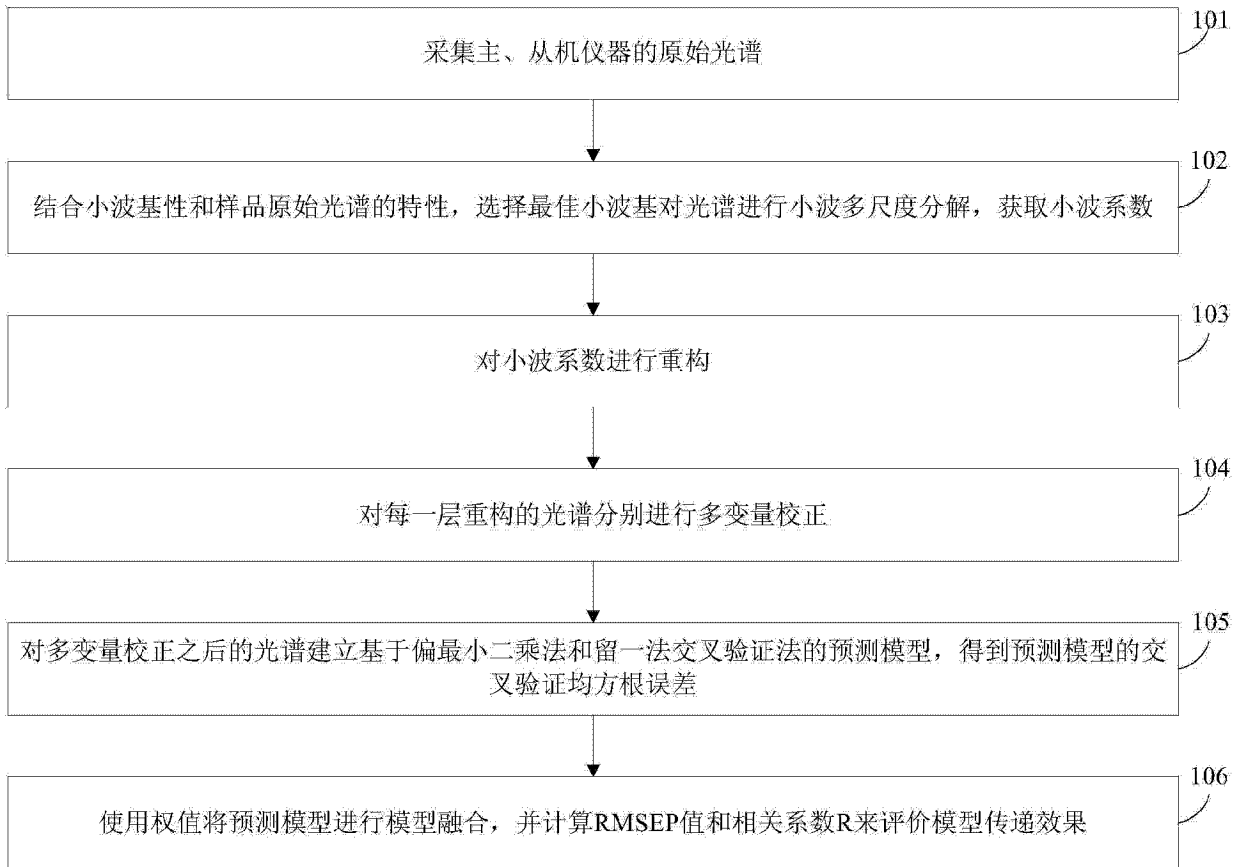


图 7

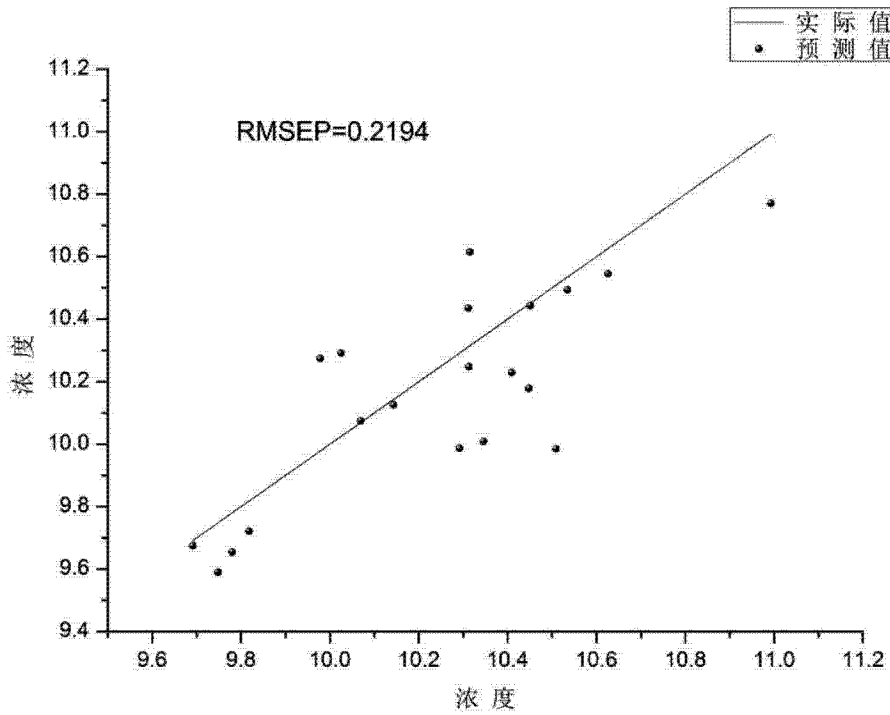


图 8 (a)

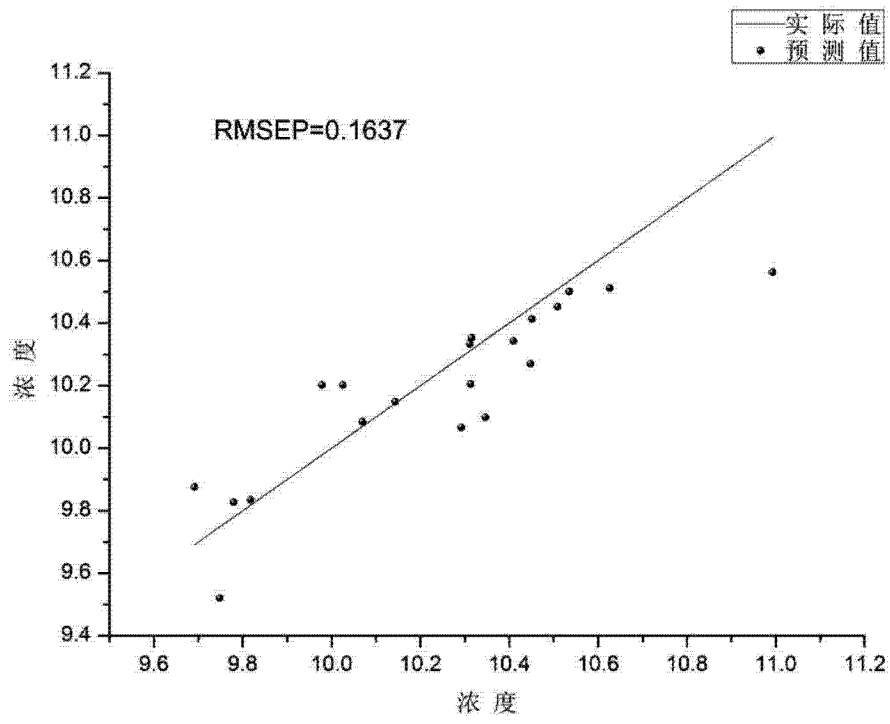


图 8 (b)