



# [12] 发明专利申请公开说明书

[21] 申请号 200310112331.6

[43] 公开日 2004年11月10日

[11] 公开号 CN 1544921A

[22] 申请日 2003.11.26

[21] 申请号 200310112331.6

[71] 申请人 广东省农业科学院作物研究所  
地址 510640 广东省广州市天河区五山路

[72] 发明人 曹干

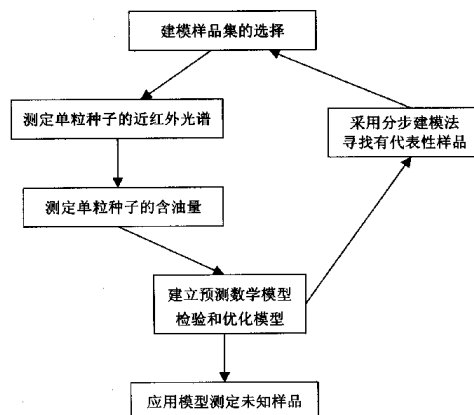
[74] 专利代理机构 广州市华学知识产权代理有限公司  
代理人 李卫东

权利要求书2页 说明书7页 附图2页

[54] 发明名称 一种非破坏性测定花生单粒种子含油量的方法

[57] 摘要

本发明是一种非破坏性测定花生单粒种子含油量的方法，该方法以傅立叶变换近红外光谱技术为基础，结合化学计量学方法，采用漫反射测定方式，以多种基因型的花生单粒种子为样品背景建立多元回归数学模型，再通过模型预测未知样品的含油量。本发明方法是非破坏性的，不需要对样品进行任何预处理，不伤害种子活力和组织结构。本发明方法操作简单，测定速度快，测定准确性较高。本发明适用于花生高油品质育种、种质资源评价和遗传规律研究。



1.一种非破坏性测定花生单粒种子含油量的方法，其特征是，它包括如下步骤：

第一步 选择有代表性的花生单粒种子作为建立多元回归预测模型的标准样品集；

第二步 使用傅立叶变换近红外光谱仪和光导纤维固体探头测定标准样品集中每个花生单粒种子样品的漫反射光谱；

第三步 采用索氏残存法测定标准样品集中每个花生单粒种子样品的含油量，作为与样品集近红外光谱一一对应的化学值；

第四步 采用多元校正方法建立和优化多元回归预测模型，通过比较各种可能组合下预测模型的决定系数  $R^2$  和均方根差 RMSECV，选取  $R^2$  尽可能大而 RMSECV 尽可能小的组合，RMSECV 和  $R^2$  按照以下方式确定：

$$RMSECV = \sqrt{\frac{1}{M} \sum (Differ_i)^2} \quad R^2 = 1 - \frac{\sum (Differ_i)^2}{\sum (y_i - y_m)^2}$$

其中， $Differ_i$  为第  $i$  个样品的化学值和预测值之差， $M$  为样品数， $y_i$  为第  $i$  个样品的化学值， $y_m$  为  $M$  个样品预测值的平均值；

第五步 按照第二步的方法采集未知样品的近红外光谱，将光谱输入预测模型，从而确定未知样品的含油量。

2.根据权利要求 1 所述的一种非破坏性测定花生单粒种子含油量的方法，其特征是，所述有代表性的花生单粒种子的选择采用分步建模法，先随机选择 50~60 份样品材料，建立一个基本模型；应用基本模型从大量的材料中寻找和筛选有代表性的样品 50~60 份，测定其近红外光谱和含油量化学值，将其作为建模样品加入到基本模型中，按第四步的方法进行建模和优化，获得第一个优化模型；再重复上述步骤，获得第二个、第三个以及更多的优化模型，直到获得满意的预测模型。

3.根据权利要求 1 所述的一种非破坏性测定花生单粒种子含油量的方法，其特征是，所述漫反射光谱的测定采用如下装置：一个垂直固定有固定支架的底座，在固定支架上依次安装有固定夹和可上下水平移动的平台，平台由旋钮控制，连接光导纤维的固体探头的头部垂直向下由固定夹安装在固定支架上，在平台上正对固体探头安装有白陶瓷，白陶瓷上固定有样品垫圈，

其内放置单粒种子样品。

4.根据权利要求 3 所述的一种非破坏性测定花生单粒种子含油量的方法，其特征是，所述固体探头头部直接接触种子子叶背面的中部。

5.根据权利要求 1 所述的一种非破坏性测定花生单粒种子含油量的方法，其特征是，测定漫反射光谱时，所述花生单粒种子的装样方法是种子的子叶背面向上。

6.根据权利要求 1 所述的一种非破坏性测定花生单粒种子含油量的方法，其特征是，建立多元回归预测模型的最佳谱区范围是  $10000\sim 6098\text{cm}^2$  和  $4601\sim 4247\text{cm}^2$ 。

7.根据权利要求 1 所述的一种非破坏性测定花生单粒种子含油量的方法，其特征是，建立多元回归预测模型的最佳光谱预处理方法是先进行一阶导数处理，再进行直线递减处理。

8.根据权利要求 1 所述的一种非破坏性测定花生单粒种子含油量的方法，其特征是，所述未知样品是完整的花生单粒种子。

9.根据权利要求 1 所述的一种非破坏性测定花生单粒种子含油量的方法，其特征是，所述多元校正方法包括偏最小二乘算法、主成分回归法、逆最小二乘法、多元线性回归法。

## 一种非破坏性测定花生单粒种子含油量的方法

### 技术领域

本发明涉及花生种子品质成分的定量分析技术领域，具体是指一种应用傅立叶变换近红外光谱分析技术、结合现代化学计量学方法的非破坏性测定花生单粒种子含油量的方法。

### 背景技术

测定花生种子含油量的常规化学方法一般采用索氏法。该测定方法的取样方式和样品预处理方法均为破坏性的，也存在耗时费力、测定程序烦琐、测定成本较高等缺点。在花生品质育种和遗传研究中极为需要快速实时、操作简单、不破坏样品的分析方法。在以提高含油量为目的的花生品质育种中，迫切需要能在早期分离世代非破坏性地检测出花生单粒种子含油量的分析技术，使育种者能在变异世代定向选择高油性状，从而使高油基因得以累加并能与高产、抗病等优良基因整合。毋庸置疑，非破坏性分析技术已成为花生高油育种研究的瓶颈，严重阻碍育种效率的提高。

近十年来，由于近红外光谱技术、化学计量学和计算机软件技术的有机结合，使近红外光谱技术在天然复杂物质的品质分析上发展迅速，应用范围越来越广泛，并以其测定速度快、操作简便和对样品的非破坏性等特点而日益受到重视。虽然应用近红外光谱技术分析农产品品质成分的报道很多，但迄今为止未见有应用近红外光谱技术非破坏性测定花生单粒种子含油量的报道。

### 发明内容

本发明的目的就是为了解决上述现有技术中存在的不足之处，提供一种快速实时、操作简单、非破坏性测定花生单粒种子含油量的方法。

本发明所述一种非破坏性测定花生单粒种子含油量的方法，其特征是，它包括如下步骤：

第一步 选择有代表性的花生单粒种子作为建立多元回归预测模型的标准样品集；

第二步 使用傅立叶变换近红外光谱仪和光导纤维固体探头测定标准样

品集中每个花生单粒种子样品的漫反射光谱；

第三步 采用索氏残存法测定标准样品集中每个花生单粒种子样品的含油量，作为与样品集近红外光谱一一对应的化学值；

第四步 采用多元校正方法建立和优化多元回归预测模型，通过比较各种可能组合下预测模型的决定系数（ $R^2$ ）和均方根差（RMSECV），选取  $R^2$  尽可能大而 RMSECV 尽可能小的组合，RMSECV 和  $R^2$  按照以下方式确定：

$$RMSECV = \sqrt{\frac{1}{M} \sum (Differ_i)^2} \quad R^2 = 1 - \frac{\sum (Differ_i)^2}{\sum (y_i - y_m)^2}$$

其中， $Differ_i$  为第  $i$  个样品的化学值和预测值之差， $M$  为样品数， $y_i$  为第  $i$  个样品的化学值， $y_m$  为  $M$  个样品预测值的平均值；

第五步 按照第二步的方法采集未知样品的近红外光谱，将光谱输入预测模型，从而确定未知样品的含油量。

为了更好地实现本发明，本发明的另一发明点是用分步建模法来解决寻找有代表性样品难的问题，可节省建模的时间和费用。花生种子是化学成分复杂的天然有机物质，无法象纯化学物质那样随意人工配制各种浓度梯度的样品，只能从大量的材料中筛选有代表性的样品，而采用常规化学方法测定花生单粒种子含油量耗时费力，工作量很大且成效低。分步建模法的具体步骤是：先随机选择 50~60 份样品材料，按上述 1~4 步骤先建立一个基本模型；应用基本模型从大量的材料中寻找和筛选有代表性的样品 50~60 份，测定其近红外光谱和含油量化学值，将其作为建模样品加入到基本模型中，按步骤 4 的方法进行建模和优化，获得第一个优化模型；再重复上述步骤，获得第二个、第三个以及更多的优化模型，直到获得满意的预测模型。

本发明的原理是，以傅立叶变换近红外漫反射技术为基础，结合现代化学计量学方法，采用漫反射测定方式，以多种基因型的花生单粒种子作为样品背景，通过化学计量学方法建立多元回归预测数学模型，再由模型测定未知样品的含油量。

本发明适用于以提高花生含油量为目的的品质育种研究，可用于分离世代单粒种子含油量的非破坏性检测，也适用于花生油分性状遗传规律研究和种质资源的评价利用。

本发明与现有技术相比，具有如下优点和有益效果：

1.本发明是一种非破坏性的分析方法，以完整的花生单粒种子为测定对象，单粒种子不需要任何预处理就可快速检测出其含油量，对种子的活力和组织结构无任何损伤。

2.本发明方法操作简单、测定成本低，测定速度快，一个样品仅耗时一分钟，因此可在短期内处理大量样品。

3.本发明方法的测定准确性较高，可满足花生高油育种的要求。

#### 附图说明

图 1 是本发明的流程图；

图 2 是本发明实施例的漫反射光谱测定装置的结构示意图。

#### 具体实施方式

下面结合附图和实施例，对本发明做进一步地详细说明。

如图 1 所示，本发明实施例的流程如下：

##### 第一步 建模样品集的选择

采用分步建模的方法来寻找有代表性样品，选择一定数量有代表性的不同基因型花生单粒种子作为建立模型的标准样品集，样品的数量以不少于 150 份为好，本实施例共选择和收集了 230 个样品作为标准样品集。花生单粒种子材料来源于 200 个不同品种或品系，在分类上属珍珠豆类型；材料来源还包括不同年份和季节以及不同产地。样品集含油量的变化幅度为 30.80%~66.89%，浓度梯度空间分布合理。在自然条件下晒干花生种子，种子的含水量保持在 8%以下。本发明所说的样品代表性是指样品遗传背景的多样化、时间（年份和季节）和空间（产地来源）分布的多样化。样品集代表性的好坏对预测模型的稳定性、适应性有很大影响。样品集的材料类型越多越好，含油量变化幅度越大越好，浓度梯度的空间分布越均匀越好。

##### 第二步 花生单粒种子近红外光谱的测定

本实施例使用德国布鲁克公司生产的 VECTOR 22 / N 型傅立叶变换近红外光谱仪，仪器参数设置为：扫描谱区为  $4000\sim 10000\text{cm}^{-1}$ ，扫描次数为 64 次，分辨率为  $8\text{cm}^{-1}$ ，以白陶瓷为背景。样品近红外光谱的测定是在花生单粒种子漫反射光谱测定装置上完成。

花生单粒种子近红外光谱的采集是利用光导纤维固体探头、以漫反射方式进行的，可以在漫反射光谱测定装置上完成。本实施例提供的一个专门设

计的漫反射光谱测定装置的结构如图 2 所示,它的底座 10 上垂直固定有固定支架 1,在固定支架 1 上依次安装有 2 个固定夹 3 和可上下水平移动的平台 8,平台 8 由旋钮 9 手动控制,连接光导纤维 2 的固体探头 4 的头部垂直向下由固定夹 3 安装在固定支架 1 上,在平台 8 上正对固体探头 4 安装有白陶瓷 7,白陶瓷 7 上固定有样品垫圈 6,其内放置单粒种子样品 5。

**装样方法:** 在可移动平台上安放表面平滑的白陶瓷背景,将种子放置在其上,在种子下面可放置大小合适的垫圈,防止种子滚动;种子的子叶背面向上。

**测定方法:** 利用平台上下移动使种子子叶背面中部与垂直的固体探头头部接触,指令光谱仪开始扫描,1 分钟内即完成扫描,获得样品的漫反射近红外光谱,扫描期间不要移动或振动样品和固体探头。

所有样品近红外光谱的测定均严格按照上述装样方法和测定方法进行。

### 第三步 花生单粒种子含油量的测定

采用常规化学方法测定标准样品集中每一样品的含油量,作为与样品集近红外光谱一一对应的化学值。采用索氏残存法测定花生单粒种子的含油量,在 YG—2 型脂肪抽提器中以乙醚抽提,具体操作步骤参考国家标准 GB 2909—82。严格控制测定误差,测定结果的标准差控制在 1%左右为好。

### 第四步 预测数学模型的建立和优化

建立预测数学模型采用的多元校正方法可以是偏最小二乘法(PLS)、主成分回归法(PCR)、逆最小二乘法(ILS)和多元线性回归法(MLR)等。目前以偏最小二乘法(PLS)最为成熟,应用最广。采用偏最小二乘法(PLS)建立多元回归预测数学模型。本实施例应用德国布鲁克公司的 OPUS 4.2 版本定量分析软件进行多元回归模型的校正和优化,同样可采用其他商业化的同类型定量分析软件来完成。

将标准样品集中每个样品的近红外光谱和含油量化学值一一对应输入 OPUS 定量分析软件。利用在 OPUS 软件中设置的自动寻找和优化功能寻找建立模型的最佳条件,通过比较各种可能组合下预测模型的决定系数( $R^2$ )和均方根差(RMSECV),选取  $R^2$  尽可能大而 RMSECV 尽可能小的组合。采用内部交叉证实对预测数学模型进行验证。内部交叉证实是指依次剔除建模样品集中一个(或多个)样品,用剩余样品来建模预测被剔除样品的含量,

比较被剔除样品预测值与化学值的差异，由此判断所建模型的预测准确性。

RMSECV 和  $R^2$  由以下方式确定：

$$RMSECV = \sqrt{\frac{1}{M} \sum (Differ_i)^2} \quad R^2 = 1 - \frac{\sum (Differ_i)^2}{\sum (y_i - y_m)^2}$$

其中： $Differ_i$  表示第  $i$  个样品的化学值和预测值之差， $M$  为样品数， $y_i$  为第  $i$  个样品的化学值， $y_m$  为  $M$  个样品预测值的平均值。

由于在常规化学分析和近红外光谱测定中不可避免出现误差，在建模时需要将样品近红外光谱进行统计学上的显著性检验，商业化软件中均有此功能。在 F 检验时，设置检验异常阈值（F 概率值）为 0.99，将每一个建模样品的化学值与预测值之间的误差与平均误差进行比较，得到相应的 F 值和 F 概率值。经 F 值检验，剔除误差显著的异常样品。另外，样品集的个数并非越多越好。根据“少而精”的原则，对建模样品进行选择和优化。在保证样品集含油量变幅大、浓度梯度分布均匀的前提下，剔除在空间分布过度密集

的样品，同时尽量避免空间分布出现空洞。

经反复测试比较，本实施例获得建立花生单粒种子含油量多元回归预测模型的最佳条件为：最佳主成分维数为 15，最佳谱区范围为  $10000 \sim 6098 \text{cm}^2$  和  $4601 \sim 4247 \text{cm}^2$ ，最佳光谱预处理方法为一阶导数 + 直线递减（先进行一阶导数处理，再进行直线递减处理）。由此而建立的预测模型有很高的质量，模型的决定系数（ $R^2$ ）为 0.9785，测定误差（RMSECV）为 1.16%，与索氏残存法的测定误差 1.15% 相接近。

#### 第五步 应用模型测定未知样品

建立好预测数学模型之后，就可以测定未知样品（完整的花生单粒种子）的含油量。重复第二步采集未知样品的近红外光谱，将光谱输入预测模型，计算机立即给出未知样品的含油量。

用本实施例建立的模型测定 30 个花生单粒种子样品的含油量，同时采用索氏残存法测定其含油量作为对比，测定结果列于表 1。由结果可见，两种方法的测定结果很相近，绝对误差和相对误差都较小，表明本发明方法有较高的测定准确性。作为一种非破坏性和快速的测定方法，实际测定结果无疑是可以接受的。

表 1. 本实例测定花生单粒种子含油量的效果



样品编号	本发明方法%	索氏残存法%	绝对误差%	相对误差%
1#	46.20	45.61	-0.59	-1.29
2#	46.44	47.65	1.21	2.54
3#	46.96	47.18	0.22	0.46
4#	47.33	47.96	0.62	1.30
5#	47.50	45.55	-1.95	-4.29
6#	48.80	49.32	0.52	1.05
7#	49.25	49.92	0.67	1.34
8#	49.53	50.86	1.33	2.62
9#	49.56	51.24	1.68	3.28
10#	49.66	51.02	1.36	2.66
11#	49.88	50.51	0.63	1.24
12#	50.12	51.97	1.85	3.56
13#	50.28	50.58	0.30	0.59
14#	51.28	52.00	0.71	1.37
15#	51.41	54.01	2.60	4.81
16#	52.21	51.62	-0.59	-1.13
17#	52.80	55.61	2.82	5.07
18#	52.80	53.67	0.87	1.63
19#	53.19	55.63	2.43	4.38
20#	53.57	55.03	1.46	2.65
21#	53.66	53.73	0.07	0.14
22#	53.70	56.33	2.64	4.68
23#	54.68	56.76	2.07	3.65
24#	55.79	56.85	1.07	1.88
25#	44.07	42.82	-1.25	-2.92
26#	43.56	46.04	2.48	5.38
27#	41.86	44.19	2.33	5.28

---

28#	39.60	42.00	2.40	5.71
29#	35.49	38.43	2.94	7.65
30#	38.11	37.52	-0.58	-1.56

如上所述，即可较好地实现本发明。

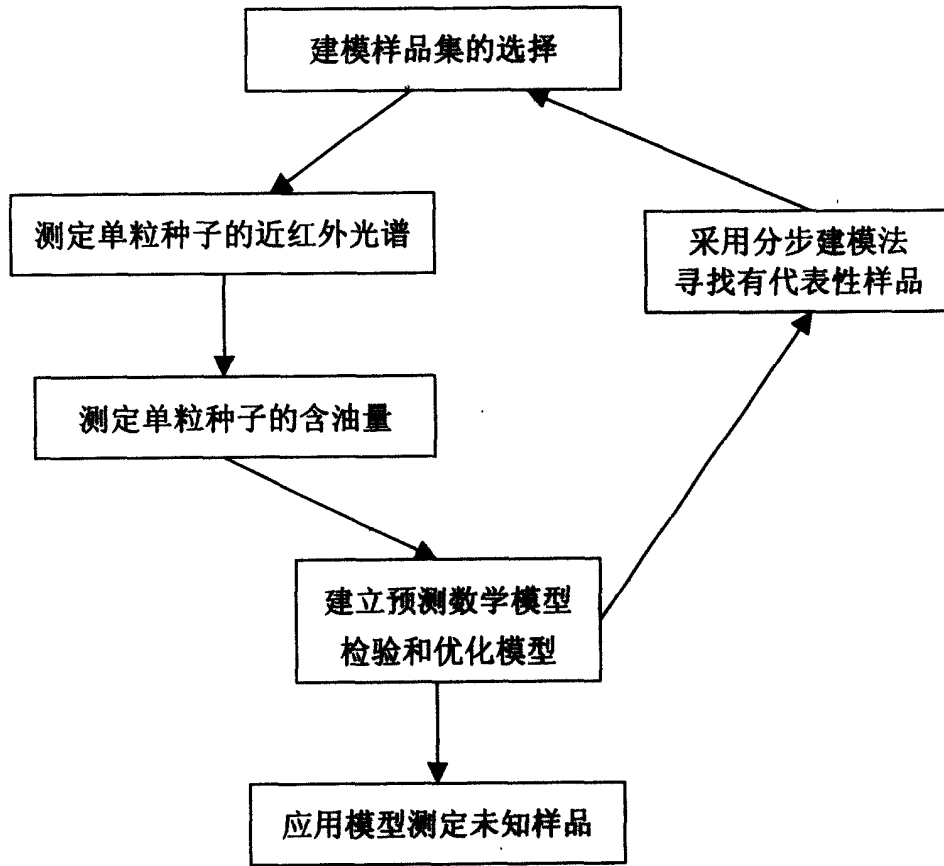


图1

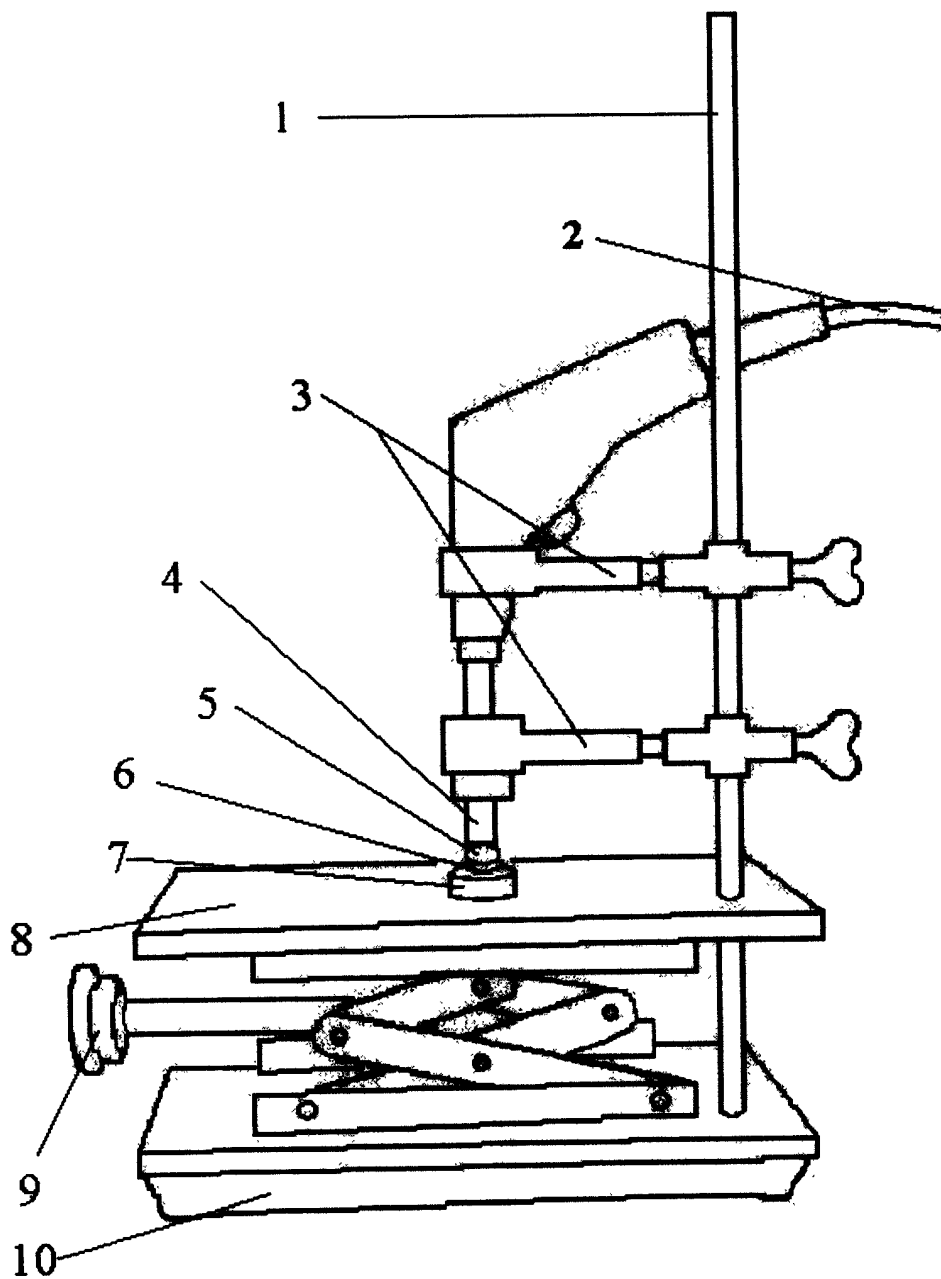


图 2