

# 基于可见-近红外光谱技术的水稻穗颈瘟染病程度分级方法研究

吴迪<sup>1</sup>, 曹芳<sup>1</sup>, 张浩<sup>2</sup>, 孙光明<sup>1</sup>, 冯雷<sup>1\*</sup>, 何勇<sup>1\*</sup>

1. 浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029

2. 浙江省农业科学院数字农业研究中心, 浙江 杭州 310021

**摘要** 采用 Vis-NIR 技术对水稻穗颈瘟染病程度分级方法进行了研究。分别基于原始光谱, 变量标准化 (SNV) 预处理后和多元散射校正 (MSC) 预处理后的光谱, 应用无信息变量消除法 (UVE) 结合连续投影算法 (SPA) 对 Vis-NIR 光谱区进行有效波长的选择。选择后的波长作为输入变量建立最小二乘-支持向量机 (LS-SVM) 模型。结果表明 SNV-UVE-SPA 建立的 LS-SVM 模型预测效果最好。通过 SNV-UVE-SPA 从全波段 600 个波长中选择了 6 个最能够反应光谱信息的波长 (459, 546, 569, 590, 775 和 981 nm)。SNV-UVE-SPA-LS-SVM 组合模型对预测集样本预测得到的确定系数 ( $R_p^2$ ), 预测集的预测标准差 (RMSEP) 和剩余预测偏差 (RPD) 分别达到了 0.979, 0.507 和 6.580。结果表明, 采用 Vis-NIR 光谱技术对水稻穗颈瘟染病程度进行分级是可行的。通过 UVE-SPA 得到的有效波长能够很好地代替全波长。

**关键词** Vis-NIR 光谱; 水稻穗颈瘟; 无信息变量消除法; 连续投影算法; 变量选择

**中图分类号**: O657.3 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2009)12-3295-05

## 引言

稻瘟病是水稻生产上最主要的病害之一, 其中以穗颈瘟对产量的影响最大。目前穗颈瘟常用的染病程度鉴定方法主要是培养稻穗离体接种<sup>[1]</sup>, 然后通过植保专家进行目测调查, 参照国家稻瘟病穗瘟发病率分级标准, 对病情进行记载分级。因为穗颈瘟的病情划分标准要求比较精确, 普通人很难进行准确的病情区分。即使是专家分析评判, 也需要花费大量时间和精力。在现代农业生产中, 对稻瘟病预测预报、调查分级和防治措施评价等工作频率、快速、准确程度要求越来越高, 这为水稻病害诊断技术提出了新的研究课题。

Vis-NIR 分析技术, 是一种快速、无损、低成本、无污染的分析技术<sup>[2]</sup>。其中, 吴迪等应用 Vis-NIR 光谱技术分别对番茄和茄子叶片的灰霉病进行了检测研究<sup>[3]</sup>。当植物受到病虫害侵染时, 光谱反射率发生明显变化, 这与植物叶片的光合色素含量的变化和细胞结构的改变有着密切关系<sup>[4]</sup>。但是由于水稻品种不同, 叶绿素含量、水分含量、组织结构、叶层构造等生物特性也不同, 在发生病菌胁迫之后, 其光谱特征也表现各异。因此需要从水稻不同光谱中发现这些不同和差异, 以获取对病害进行分级的有用信息。

本文将无信息变量消除算法 (uninformative variable elimination, UVE) 结合连续投影算法 (successive projections algorithm, SPA) 提取的有效波长作为最小二乘支持向量机 (least square-support vector machine, LS-SVM) 的输入, 建立了水稻穗颈瘟的快速、无损分级模型。

## 1 材料与方法

### 1.1 样本制备

实验水稻选用不同抗性类型的原丰早等 17 个籼稻和粳稻品种, 进行离体接种。在抽穗初期, 每个株系用剪刀剪下 12 个稻穗, 要求穗颈节在叶鞘内; 然后, 剥去叶鞘, 将穗颈节剪成长约 6 cm 的节段 (穗颈节左右各 3 cm), 去除残留颖花, 在一用 70% 乙醇消毒过的铝盘 (35 cm × 25 cm × 2.5 cm) 内铺好灭菌滤纸, 并用浓度为 100  $\mu\text{g} \cdot \text{mL}^{-1}$  的无毒苯咪唑 (benzimidazole) 润湿, 将剪好的穗颈节置于湿滤纸上, 穗颈节两端用相同浓度的苯咪唑浸湿的脱脂棉条压住, 然后接种。选用分属于中国稻瘟病菌 ZA, ZB, ZC, ZD, ZE, ZF 和 ZG 等 7 群 10 个小种的单孢菌株作为测定用菌种, 孢子悬浮液浓度为 20~30 万  $\cdot \text{mL}^{-1}$ , 内含 2% 的羧甲基纤维素, 以便孢子吸附于穗颈节部。接种时用毛笔蘸适量孢子液涂抹

收稿日期: 2008-11-10, 修订日期: 2009-02-12

基金项目: 国家自然科学基金项目 (60605011, 30671213), 国家科技支撑项目 (2006BAD10A04) 和浙江省 2009 年度重大科技专项资助

作者简介: 吴迪, 1984 年生, 浙江大学生物系统工程与食品科学学院博士生 e-mail: ilkb8mvp@yahoo.com.cn

\* 通讯联系人 e-mail: yhe@zju.edu.cn

于穗颈节部,再用保鲜膜密封铝盘保湿,在 28 °C 下光照培养 10~12 d 后测定穗颈发病情况。穗颈瘟染病程度等级按照 GB/T 15790—1995 标准划分。

### 1.2 光谱采集及样本集划分

实验使用美国 ASD(analytical spectral device, Boulder, USA)公司的 Handheld Field Spec 光谱仪,探头视场角为 20°,光谱扫描次数设定为 30 次。光源采用 14.5 V 卤素灯。对不同染病程度等级采集 325~1 075 nm 的漫反射光谱一共 51 个样本。采用 Kennard-Stone 算法<sup>[5]</sup>选取 30 个样本作为建模集,剩余 21 个样本作为预测集。分析软件采用 Unscrambler V9.7 (CAMO AS, Oslo, Norway)以及 Matlab V7.6 (The Math Works, Natick, USA)。

### 1.3 无信息变量消除算法原理及算法实现

UVE 是基于分析 PLS 回归系数  $b$  建立的波长选择算法,用于消除不提供信息的变量,减少模型中包含的变量数,降低模型的复杂性<sup>[7]</sup>。在 PLS 模型中,UVE 把一定变量数目的随机变量矩阵加入光谱矩阵中,然后通过交叉验证建立 PLS 模型。分析每个变量的系数向量的平均值和标准偏差的商  $C$  的稳定性。根据  $C$  的绝对值大小确定是否把该变量用于最后 PLS 模型中。具体算法请见参考文献<sup>[6]</sup>。

### 1.4 连续投影算法

SPA 是一种新的变量提取方法<sup>[7,8]</sup>。它能够利用向量的投影分析,寻找含有最低限度的冗余信息的变量组,并使变量之间的共线性达到最小,同时能大大减少建模所用变量的个数,提高建模的速度和效率。SPA 的算法简要介绍如下<sup>[10]</sup>。

记  $x_{k(0)}$  为初始迭代向量,  $N$  为需要提取的变量个数。光

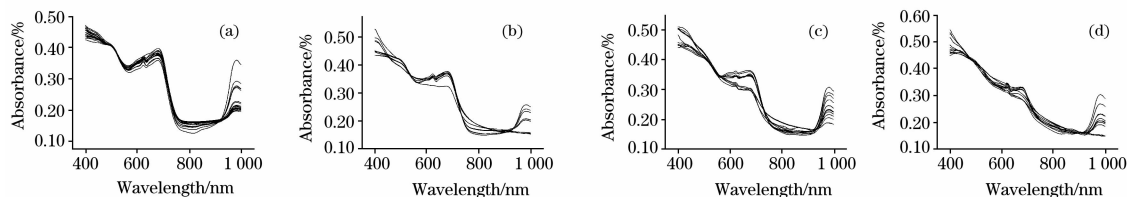


Fig. 1 Original absorbance spectra of different disease levels of rice panicle blast

Disease level: (a): 0; (b): 0.1~4.9; (c): 5.0~7.0; (d): 7.1~9.0

### 2.2 全波段光谱分析

分别采用原始光谱,变量标准化(standard normal variate, SNV)预处理后的光谱和多元散射校正(multiplicative scatter correction, MSC)预处理后的全波段光谱作为输入变量建立 LS-SVM 模型。径向基函数(radial basic function, RBF)作为 LS-SVM 的核函数,并用网格搜索法结合留一交叉验证确定模型的参数  $\gamma$  和  $\sigma^2$ 。具体的 LS-SVM 算法描述可参见参考文献<sup>[10]</sup>。建模集和预测集的染病程度分级的预测值与真实值之间的确定系数( $R_c^2$ ,  $R_p^2$ ),建模集的预测标准差(root mean square error for calibration, RMSEC),预测集的预测标准差(root mean square error for prediction, RMSEP)和剩余预测偏差(residual predictive deviation, RPD)被用于模型的评估。表 1 显示了基于不同预处理方法的预测结果。从表 1 可以看到采用原始光谱全波长作为输入变量时的建模

谱矩阵为  $j$  列。

(1)迭代开始前,任选光谱矩阵的 1 列  $j$ ,把建模集的第  $j$  列赋值给  $x_j$ ,记为  $x_{k(0)}$ ;

(2)把未选入的列向量位置的集合记为  $s$ ,  $s = \{j, 1 \leq j \leq J, j \notin \{k(0), \dots, k(n-1)\}\}$ ;

(3)分别计算  $x_j$  对剩下列向量的投影:  $Px_j = x_j - (x_j^T x_{k(n-1)})x_{k(n-1)} / (x_{k(n-1)}^T x_{k(n-1)})$ ,  $j \in s$ ;

(4)记  $k(n) = \arg(\max(\|Px_j\|))$ ,  $j \in s$ ;

(5)令  $x_j = Px_j$ ,  $j \in s$ ;

(6) $n = n + 1$ ,如果  $n < N$ ,回到第二部循环计算;

最后,提取出的变量为  $\{x_{k(n)} = 0, \dots, N-1\}$ 。对应于每一个  $k(0)$  和  $N$ ,循环一次后进行多元线性回归分析(MLR),得到验证集的预测标准偏差(RMSEV),最小的 RMSEV 值对应的  $k(0)$  和  $N$  就是最优值。

## 2 试验结果与分析

### 2.1 不同等级水稻穗颈瘟 Vis-NIR 光谱图

由于测量到的光谱两端存在噪声,因此选用 401~1 000 nm 波长范围内共 600 个变量进行分析。图 1 显示了不同等级水稻穗颈瘟的 Vis-NIR 光谱。当染病程度为 0 时为典型的植物光谱。但是随着染病程度的加重,光谱曲线发生了变化。绿光区和红光区的吸光度分别呈现上升和下降的趋势。当染病程度大于 7 时,绿光区和红光区的吸收峰已基本消失。然而光谱曲线和染病程度之间虽有一定程度的相关性,但却不明显。因此仅从光谱特征上,难以对染病程度进行分级。

集结果很好,但 RMSEC 和 RMSEP 相差过大,说明模型过拟合。而采用 MSC 和 SNV 的结果都差不多。两种预处理方法得到的模型的 RMSEC 和 RMSEP 很接近,  $R^2$  也很高。而 RPD 达到了 4.5 左右。以上结果说明应用 Vis-NIR 光谱技术能够对水稻穗颈瘟染病程度进行分级。

### 2.3 应用 UVE 选取有效波长

虽然采用全波长作为输入变量已经达到了较好的分类结果,但是 LS-SVM 模型的输入变量有 600 个,部分变量可能包含无用的或者不相关的信息,并且上百个输入变量造成模型冗余、计算复杂<sup>[11]</sup>。为此采用 UVE 分别对原始光谱(Original-UVE),MSC 预处理后(MSC-UVE)和 SNV 预处理后(SNV-UVE)的全波段 600 个光谱波长进行选择。UVE 中产生的随机变量个数同样设置为 600 个。UVE 中的最优潜在变量个数是经内部交叉验证得到的。当潜在变量分别为 6,

13, 12 时得到的 RMSECV 最小。基于上述几个潜在变量时的 Original-UVE, MSC-UVE 和 SNV-UVE 分别得到了 202, 42, 112 个波长。将得到的波长作为 LS-SVM 的输入变量, 预测结果见表 1。可以看到 Original-UVE, MSC-UVE 得到的波长建立的 LS-SVM 模型均明显过拟合。RMSEC 和 RMSEP 的值相差过大。相反, 虽然 SNV-UVE 得到的变量数大

于 MSC-UVE, 但 SNV-UVE 模型的 RMSEC 和 RMSEP 很接近,  $R^2$  也很高。而 RPD 达到了 6.8 左右, 高于其他 2 个模型。说明 SNV-UVE 的模型效果最好。模型输入变量也从 600 个减少到 112 个, 而且预测效果要好于全波段 SNV 建立的模型。

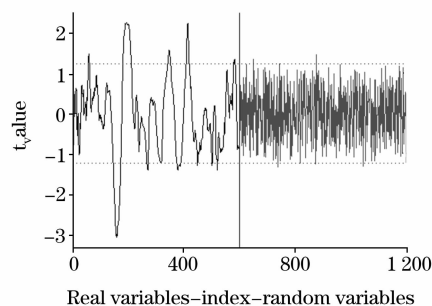
**Table 1 Prediction results of rice panicle blast's disease levels by LS-SVM models using different pretreatment and variable selection methods**

| Variable selection methods | Pretreatment | Variables | Calibration set( $n=30$ ) |       | Prediction set( $n=21$ ) |       |       |
|----------------------------|--------------|-----------|---------------------------|-------|--------------------------|-------|-------|
|                            |              |           | $R_c^2$                   | RMSEC | $R_p^2$                  | RMSEP | RPD   |
| None                       | None         | 600       | 0.993                     | 0.320 | 0.944                    | 0.842 | 3.915 |
|                            | MSC          | 600       | 0.955                     | 0.756 | 0.960                    | 0.697 | 4.548 |
|                            | SNV          | 600       | 0.955                     | 0.759 | 0.957                    | 0.717 | 4.423 |
| UVE                        | None         | 202       | 0.996                     | 0.250 | 0.961                    | 0.700 | 4.984 |
|                            | MSC          | 42        | 1.000                     | 0.025 | 0.969                    | 0.569 | 5.597 |
|                            | SNV          | 112       | 0.978                     | 0.523 | 0.982                    | 0.480 | 6.869 |
| UVE-SPA                    | None         | 13        | 0.992                     | 0.317 | 0.895                    | 1.283 | 2.704 |
|                            | MSC          | 10        | 0.973                     | 0.582 | 0.936                    | 0.817 | 3.884 |
|                            | SNV          | 6         | 0.976                     | 0.551 | 0.979                    | 0.507 | 6.580 |

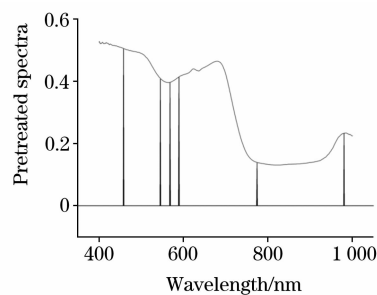
SNV-UVE 变量选择过程见图 2。蓝色竖线的左侧为 600 个变量的稳定性 C 分布曲线, 右侧为 UVE 中产生的 600 个随机变量的稳定性 C 分布曲线。两条蓝色水平的虚线表示变量选择的阈值上下限, 在虚线外的数值对应的变量被用于建模。阈值的选择标准为随机变量稳定性最大值的 99%。SNV-UVE 选择的波长在 546~573 和 581~610 nm 的绿光部分和 742~752 nm 的红边部分, 以及 775~921 nm 的多处近红外部分有较多的波长被选择。此外在 458 nm 左右的紫光部分, 668~671 nm 的红光部分以及 979~984 nm 的水分吸收峰部分也有波长被选择。说明无论是可见光还是近红外光谱都包含对于水稻穗颈瘟染病程度分级有用的信息。

#### 2.4 应用 UVE-SPA 选取有效波长

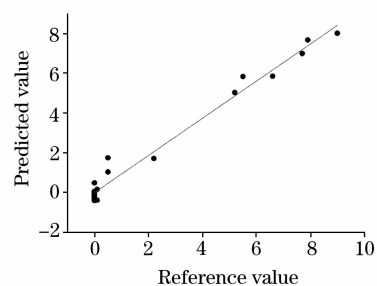
由于经过 UVE 选择后的变量个数仍过于庞大, 因此采用 SPA 对经过 UVE 选择后的变量进一步进行选择。分别提取出 13, 10 和 6 个有效波长时 Original-UVE-SPA, MSC-UVE-SPA 和 SNV-UVE-SPA 的 RMSEV 趋于平坦。将得到的有效波长分别作为输入变量建立 LS-SVM 模型, 预测结果见表 1。可以看到当不经过预处理时, Original-UVE-SPA 建立的模型再次过拟合。而 MSC-UVE-SPA 建立的模型也有一定程度的过拟合。并且和 MSC-UVE 建立模型相比, 虽然 MSC-UVE-SPA 仅仅采用了 10 个变量, 少于 MSC-UVE 的 42 个。但其 RPD 要远低于 MSC-UVE 的 RPD 值。而 SNV-UVE-SPA 建立的模型, 其 RMSEC 和 RMSEP 很接近。甚至 SNV-UVE-SPA 模型的预测集效果比建模集效果还要好。并且 SNV-UVE-SPA 将模型输入变量从 SNV-UVE 的 112 个成功地减少到了 6 个。虽然 SNV-UVE-SPA 的预测效果要略逊于 SNV-UVE 模型, 但考虑到其仅有 6 个输入变量, 并且其 RPD 值也达到了 6.580, 其模型的预测效果仍然可以接受。表明采用 SNV-UVE-SPA 所提取的有效波长能充分代表原始光谱的有效信息, 可以作为波长提取的有效手段。



**Fig. 2 Stability distribution of SNV-UVE. Two horizontal dot lines indicated the lower and upper boundaries**



**Fig. 3 Plot of 6 selected wavelengths by SNV-UVE-SPA**



**Fig. 4 Predicted vs reference values of rice panicle blast by SNV-UVE-SPA-LS-SVM model**

由 SNV-UVE-SPA 得到的这 6 个波长分别为 459, 546, 569, 590, 775 和 981 nm(图 3), 分处于蓝光、绿光、红边以及近红外光谱, 其中 981 nm 反映了水的吸收峰。说明可以通过上述几个波长来对水稻穗颈瘟染病程度进行分级。图 4 为由 SNV-UVE-SPA 得到的 6 个波长建立的 LS-SVM 对预测集样本染病程度的预测结果散点分布图。图中样本点基本都分布在回归线附近。表明采用 SNV-UVE-SPA 选择的少数几个 Vis-NIR 光谱范围内的有效波长建立的 LS-SVM 模型对水稻穗颈瘟染病程度进行分级是可行的, 并且获得了满意的分级精度。并且经过 SNV 预处理后光谱建立的模型能够避免如原始光谱和经 MSC 预处理后建立的模型造成的过拟合现象, 预测结果也优于原始光谱和经 MSC 预处理后建立

的模型。说明 SNV 能够较好地避免模型过拟合, 并且能够提高模型预测精度。

### 3 结 论

本文研究结果表明应用 Vis-NIR 光谱技术对水稻穗颈瘟染病程度进行分级是可行的, 为染病程度分级提供了新的方法。采用 UVE-SPA 组合的变量选择方法能够选取最优的建模变量。在进一步的研究中可以通过采集这 6 个波段的光谱信息或者相应波段的光谱图像信息来对水稻穗颈瘟染病程度进行快速精确分级。

### 参 考 文 献

- [1] SUN Guo-chang, SHEN Zong-tan(孙国昌, 申宗坦). Chinese Journal of Rice Science(中国水稻科学), 1992, 6(1): 39.
- [2] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄, 赵龙莲, 韩东海, 等). The Foundation and Application of Near Infrared Spectroscopy Analysis(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005.
- [3] WU Di, FENG Lei, ZHANG Chuan-qing, et al(吴迪, 冯雷, 张传清, 等). J. Infrared Millim. Waves(红外与毫米波学报), 2007, 26(4): 269.
- [4] Dawson T P, Curran P J, North P. Remote Sensing of Environment, 1999, 67(2): 147.
- [5] Macho S, Iusa R, Callao M P, et al. Anal. Chim. Acta, 2001, 445(2): 213.
- [6] Centner V, Massart D L. Anal. Chem., 1996, 36: 8851.
- [7] Araújo M C U, Saldanha T C B, Galvão R K H, et al. Chemom. Intell. Lab. Syst., 2001, 57: 65.
- [8] Galvão R K H, Araújo M C U, Fragoso W D, et al. Chemom. Intell. Lab. Syst., 2008, 92: 83.
- [9] CHEN Bin, MENG Xiang-long, WANG Hao(陈斌, 孟祥龙, 王豪). Journal of Instrumental Analysis(分析测试学报), 2007, 26(1): 66.
- [10] Wu D, He Y, Feng S, et al. Journal of Food Engineering, 2008, 84: 124.
- [11] Chauchard F, Cogdill R, Roussel S, et al. Chemom. Intell. Lab. Syst., 2004, 71: 141.

## Study on Disease Level Classification of Rice Panicle Blast Based on Visible and Near Infrared Spectroscopy

WU Di<sup>1</sup>, CAO Fang<sup>1</sup>, ZHANG Hao<sup>2</sup>, SUN Guang-ming<sup>1</sup>, FENG Lei<sup>1\*</sup>, HE Yong<sup>1\*</sup>

1. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China

2. Digital Agricultural Research Centre, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

**Abstract** Visible and near infrared (Vis-NIR) spectroscopy was used to fast and non-destructively classify the disease levels of rice panicle blast. Reflectance spectra between 325 and 1 075 nm were measured. Kennard-Stone algorithm was operated to separate samples into calibration and prediction sets. Different spectral pretreatment methods, including standard normal variate (SNV) and multiplicative scatter correction (MSC), were used for the spectral pretreatment before further spectral analysis. A hybrid wavelength variable selection method which is combined with uninformative variable elimination (UVE) and successive projections algorithm (SPA) was operated to select effective wavelength variables from original spectra, SNV pretreated spectra and MSC pretreated spectra, respectively. UVE was firstly operated to remove uninformative wavelength variables from the full-spectrum. Then SPA selected the effective wavelength variables with less collinearity after UVE. Least square-support vector machine (LS-SVM) was used as the calibration method for the spectral analysis in this study. The selected effective wavelengths were set as input variables of LS-SVM model. The LS-SVM model established based on SNV-UVE-SPA obtained the best results. Only six effective wavelengths (459, 546, 569, 590, 775 and 981 nm) were selected from the full-spectrum which has 600 wavelength variables by UVE-SPA, and their LS-SVM model's performance was further improved. For SNV-UVE-SPA-LS-SVM model, coefficient of determination for prediction set ( $R_p^2$ ), root mean square error for prediction (RMSEP) and residual

predictive deviation (RPD) were 0.979, 0.507 and 6.580, respectively. The overall results indicate that Vis-NIR spectroscopy is a feasible way to classify disease levels of rice panicle blast fast and non-destructively. UVE-SPA is an efficient variable selection method for the spectral analysis, and their selected effective wavelengths can represent the useful information of the full-spectrum and have higher signal/noise ratio and less colinearity.

**Keywords** Visible and near infrared (Vis-NIR) spectroscopy; Rice panicle blast; Uninformative variable elimination (UVE); Successive projections algorithm (SPA); Variable selection

\* Corresponding author

(Received Nov. 10, 2008; accepted Feb. 12, 2009)

(上接 3287 页)

2010 Winter Conference on Plasma Spectrochemistry Monday-Saturday, January 4-9, 2010

2010 Workshop on New Plasma Instrumentation Tuesday-Thursday, January 5-7, 2010

2010 Instrument Exhibition Tuesday-Thursday, January 5-7, 2010

2010 Conference Dinner Cruise Friday, January 8, 2010

Conference Manuscripts Submission Deadline February 26, 2010

**University Research Institute for Analytical Chemistry**  
**P. O. Box 666, Hadley, MA 01035-0666**

The University Research Institute for Analytical Chemistry (URIAC), the research and laboratory division of ICP Information Newsletter, Inc., located in Amherst near the University of Massachusetts, provides consulting, chemical research, analysis, testing, and training and is the home of analytical experts performing non-routine, specialty chemical analyses for academic laboratories, industry, and government agencies. For information contact Director, ICP Information Newsletter, Inc., P. O. Box 666, Hadley, MA 01035-0666 USA. Telephone (239) 674-9430, fax (239) 674-9431, e-mail icpnews@chem.umass.edu.

**Research for**

- Biological, environmental and physical modeling and simulations
- Chemical speciation
- Chemical statistics (chemometrics)
- Environmental forensics
- Instrument evaluation and development
- Method development and testing for agriculture, biology, biomedicine, advanced materials, environmental chemistry, and semiconductor industries
- Problem solving
- Project management
- Research and development grants
- Source and country of origin identification
- Ultra-trace elemental analysis

**Ultra-Clean and Secure Laboratories for**

- Dedicated space for industrial and commercial clients
- Equipment demonstrations and advanced training
- New equipment evaluation and testing
- New technology development
- On-site experienced chemists to assist clients
- Proprietary instrumentation development

**High-Quality Sample Analysis for**

- Industry
- Environment
- University and Research Institutions
- Semiconductors
- Commerce
- Government
- Agriculture, Medicine, and Nutrition