

## 干涉光谱结合偏最小二乘法反演热液 CH<sub>4</sub> 的研究

刘青松<sup>1,2</sup>, 胡炳樑<sup>1\*</sup>, 唐远河<sup>3</sup>, 于涛<sup>1</sup>, 王雪霁<sup>1,2</sup>, 刘永征<sup>1</sup>, 杨鹏<sup>4</sup>, 王浩轩<sup>3</sup>

1. 中国科学院西安光学精密机械研究所光谱成像技术重点实验室, 陕西 西安 710119
2. 中国科学院大学, 北京 100049
3. 西安理工大学理学院, 陕西 西安 710048
4. 青岛海洋科学与技术国家实验室海洋观测与探测联合实验室, 山东 青岛 266200

**摘要** 热液释放的高温甲烷气体经扩散作用先后进入海洋和大气, 并对地球物理、化学和生物方面产生深刻影响。由于海洋溶解甲烷数据的缺乏, 导致人们对深海热液释放甲烷的活动机制和环境效应还缺乏足够的认识。我们前期提出一种光学被动成像干涉系统 OPIIS 用于热液甲烷浓度、温度和压强的实时探测和长期观测。为了从 OPIIS 的干涉光谱中精确、稳定、快速的获取热液甲烷信息, 采用将干涉光谱与偏最小二乘法相结合的方法处理 OPIIS 数据。首先分别建立三个甲烷浓度、温度和压强的单因变量预测模型, 再利用干涉条纹与辐射光谱的关系, 间接建立干涉光谱与甲烷浓度、温度和压强的 PLS 预测模型, 提高了预测模型在实际应用中的抗干扰能力和稳定性。基于洛仑兹线型建立了不同于大气环境的深海气体辐射模型, 并利用 HITRAN2016 分子光谱数据库的光谱参数, 建立了深海甲烷在任意浓度、任意温度和任意压强下的辐射光谱数据库。挑选热液其他气体对甲烷探测干扰较小的甲烷泛频带 1.64~1.66 μm 内的六条谱线建立甲烷辐射光谱与浓度、温度和压强的偏最小二乘回归模型。另外, 分析了训练集取样个数、取样间隔和主成分个数对提高预测模型综合性能的作用。利用不同训练集样本数, 不同训练集取样间隔和不同的主成分数, 分别建立 96 个浓度、温度和压强预测模型, 并分别利用 25 组预测集样本对预测模型进行交叉验证。不同模型预测均方根误差和决定系数的对比表明, 训练集取样个数、取样间隔和主成分个数等单一因素的改变并不能同时提高预测模型的预测精度、稳定性、适用范围和运算量等综合性能。经过平衡选取各项指标确定的最优回归模型的参数为: 浓度、温度和压强的适用范围分别为 5~375 mmol·L<sup>-1</sup>, 580~678 K, 10~34.5 MPa, 浓度、温度和压强的训练集取样个数分别为 50 组, 25 组, 25 组, 采样间隔分别为 5 mmol·L<sup>-1</sup>, 2 K, 0.5 MPa, 浓度、温度和压强预测模型的主成分数分别为 2, 2, 5。浓度、温度和压强预测模型的预测均方根误差分别为 3.082×10<sup>-6</sup>, 0.977 0, 5.052×10<sup>-3</sup>, 决定系数分别为 0.999 9, 0.998 9, 0.999 9。浓度、温度和压强的预测误差分别为 ±1.21×10<sup>-7</sup>, ±3.63×10<sup>-3</sup>, ±9.49×10<sup>-4</sup>, 对应的预测精度分别为 ±45.4 nmol·L<sup>-1</sup>, ±2.5 K, ±3.3×10<sup>-2</sup> MPa。结果表明, 干涉光谱结合偏最小二乘法的反演算法可以精确、稳定、快速的获取热液甲烷气体的浓度、温度和压强信息。

**关键词** 热液甲烷; 偏最小二乘法; 成像干涉; 反演

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)08-2415-06

## 引言

海水沿洋壳裂缝渗透至地壳岩浆房的过程中, 形成热液流体, 并喷发至海底, 在海底热液喷口上方将会形成热液羽

状流, 并伴随大量甲烷气泡<sup>[1]</sup>。热液释放的高温甲烷气体经扩散作用先后进入海洋和大气, 并对地球物理、化学和生物方面产生深刻影响<sup>[2-3]</sup>。由于海洋甲烷探测手段的限制和溶解甲烷数据的缺乏, 人们对热液释放甲烷的活动机制和环境效应还缺乏足够的认识<sup>[4]</sup>。相比于采样实验室测量的方法,

收稿日期: 2018-06-05, 修订日期: 2018-10-12

基金项目: 国家重点研发计划项目(2017YFC1403700), 国家自然科学基金面上项目(61675165), 中国科学院光谱成像技术重点实验室开放基金项目(LSIT201714D)资助

作者简介: 刘青松, 1991 年生, 中国科学院西安光学精密机械研究所光谱成像技术重点实验室博士研究生 e-mail: s-liuqingsong@opt.cn

\* 通讯联系人 e-mail: hbl@opt.ac.cn

原位探测方法不需要考虑样品保真的问题, 可以获取更为准确、连续、实时的数据<sup>[5]</sup>。我们前期提出一种光学被动成像干涉系统(Optical passive imaging interference system, OPIIS)用于热液甲烷的原位探测。光学成像干涉技术是国际上被动探测大气的主流方法, 由于成像干涉技术的物理机制较为复杂<sup>[6]</sup>, 且部分深海物理机制还有待于进一步研究和验证, 给 OPIIS 的数据反演工作提出更高的要求。

偏最小二乘法(partial least squares, PLS)是构造回归方程的一种数学方法, 可以进行单(多)变量的回归分析, 且其建立的预测模型具有较高的预测稳定性<sup>[7]</sup>。目前研究较多的是采用改进型偏最小二乘算法和光谱数据变换方法提高模型的预测性能<sup>[8]</sup>。文中将干涉光谱与偏最小二乘法相结合, 分别建立三个甲烷浓度、温度和压强单因变量预测模型, 并着重分析建模样本个数、样本间隔和主成分个数的选取对预测模型精度、稳定性、运算量的影响, 平衡取舍模型各项性能指标, 使模型的预测精度、稳定性、适用范围和运算量等综合性能满足应用场景需要。

## 1 OPIIS 工作原理

OPIIS 的工作原理如图 1 所示, 菲涅尔透镜收集的甲烷谱线以不同角度入射干涉滤光片, 根据干涉滤光片的透过率与谱线入射角和中心波长的关系<sup>[9]</sup>, 不同入射角不同中心波长的谱线得以分离, 并于探测器上形成干涉条纹。本文主要研究通过算法处理从干涉条纹中提取出甲烷浓度、温度和压强信息。

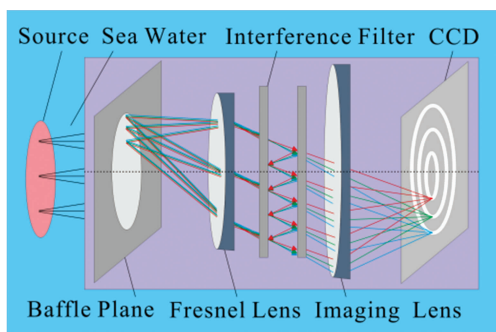


图 1 OPIIS 工作原理图

Fig. 1 Schematic diagram of OPIIS

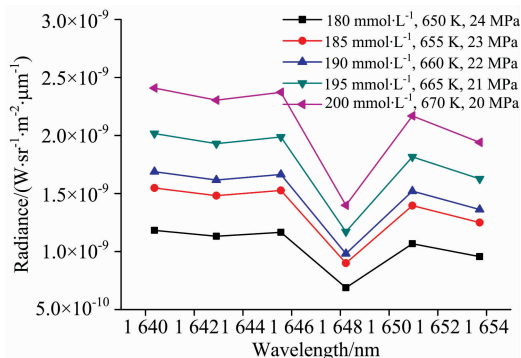


图 2 部分建模样本辐射光谱

Fig. 2 Radiance spectra of partial modeling samples

OPIIS 干涉条纹强度与入射光强的对应关系可以表示为

$$I = I_0 \tau_I(\lambda, \theta_I) = I_0 \tau_{\lambda_0} \left\{ 1 + \left[ \frac{2(\lambda - \lambda_0)}{D} + \frac{\lambda}{D} \frac{\theta_I^2}{n_c^2} \right]^2 \right\}^{-1} \quad (1)$$

其中,  $I$  为干涉条纹强度,  $I_0$  为洛伦兹型入射光强,  $\tau_I$  为干涉滤光片的透过率,  $\tau_{\lambda_0}$  为滤光片峰值透过率,  $D$  为滤光片半带宽,  $\theta_I$  为滤光片入射光的入射角,  $n_c$  为滤光片的有效折射率。为了从干涉条纹中精确、稳定、快速的提取出甲烷气体的浓度、温度和压强信息, 需要建立干涉条纹强度与甲烷分子浓度、温度和压强的回归模型。但在实际测量过程中, OPIIS 干涉条纹强度值易受环境和仪器自身因素的干扰, 极大影响回归模型的稳定性。而甲烷辐射光谱主要由深海气体辐射模型、甲烷分子光谱参数和热液环境参数决定, 其数值受其他因素影响较小。因此本文主要研究建立甲烷气体辐射光谱与分子浓度、温度和压强的回归关系, 再根据甲烷气体辐射强度与干涉条纹强度的辐射定标关系<sup>[10]</sup>, 间接确定干涉条纹与甲烷分子浓度、温度和压强的回归模型。

## 2 偏最小二乘回归建模

本文分别以甲烷辐射光谱为自变量, 以甲烷浓度、温度和压强为因变量建立单因变量预测模型, 其中甲烷光谱取自我们建立的合成光谱数据库, 图 2 为部分建模样本甲烷辐射光谱。热液区域通常位于海底 2 000 m 以下, 压强高达 20 MPa 以上, 热液喷口温度甚至高达 400 °C, 热液区域甲烷浓度高达 190 mmol·L<sup>-1</sup> 以上。根据热液环境参数, 分别选取不同浓度的样本 50 组, 75 组, 100 组, 125 组建立甲烷浓度预测模型, 其中预测集样本数都为 25 组, 训练集样本数分别为 25 组, 50 组, 75 组, 100 组。浓度取样间隔分别为 2, 3, 5 和 10 mmol·L<sup>-1</sup>, 主成分个数分别取 1, 2, 3, 4, 5, 6。通过对以上 96 组浓度预测模型的对比, 挑选出同时满足预测精度, 稳定性, 适用范围, 运算量等要求的浓度预测模型。同理, 设置温度取样间隔分别为 1, 2, 5 和 10 K, 压强取样间隔分别为 0.05, 0.1, 0.25 和 0.5 MPa, 通过对比分析, 挑选出综合性能较优的温度预测模型和压强预测模型。本文采用决定系数(coefficient of determination,  $R^2$ )和预测均方根误差(root mean square error of prediction, RMSEP)作为模型的评价指标。决定系数  $R^2$  表征模型的稳定性, 其值越大, 则模型的预测稳定性越高。预测均方根误差表征模型的预测精度, 其值越小, 则模型的预测精度越高。

## 3 结果与讨论

### 3.1 不同训练集样本间隔的模型预测结果

本文分析了训练集取样个数、训练集取样间隔和主成分个数对模型预测的影响。以压强预测模型为例, 设置训练集样本数为 100 组, 压强取样间隔分别为 0.5, 0.25, 0.1 和 0.05 MPa, 对应的模型适用范围分别为 0.5~62.5, 0.25~31.25, 20.1~32.5 和 20.05~26.25 MPa。从本文光谱数据库中提取 400 条原始光谱曲线建立压强预测模型。图 3 为不

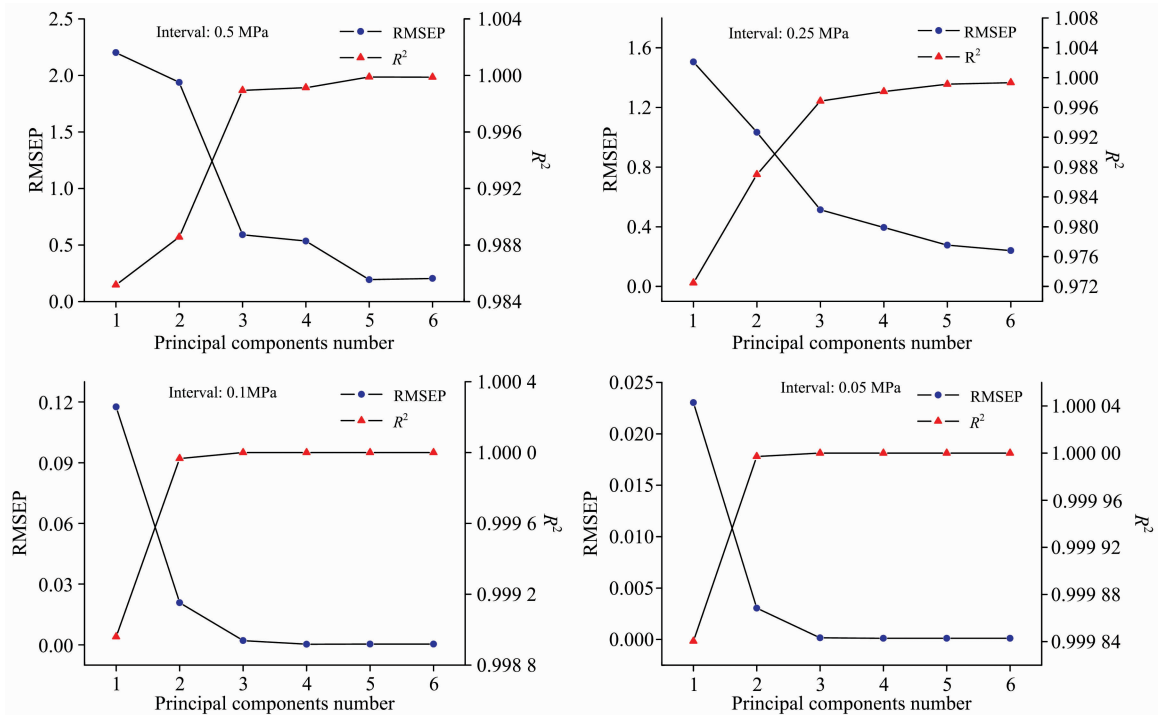


图 3 不同训练集取样间隔的压强预测模型的 RMSEP 和  $R^2$

Fig. 3 RMSEP and  $R^2$  of pressure prediction model for various intervals of training samples

同取样间隔的压强模型预测均方根误差 RMSEP 和决定系数  $R^2$ 。

图 3 的每幅图中展示了主成分对压强预测模型的预测均方根误差 RMSEP 和决定系数  $R^2$  的影响。当主成分数增加时, 压强模型预测均方根误差 RMSEP 呈现减小趋势, 决定系数  $R^2$  呈现增加趋势直至接近于 1。在每个模型中, 存在一个最优主成分值使得模型的预测精度和稳定性最佳, 当主成分数较小时, 模型性能较低, 当主成分较大时, 运算量也较大, 同时未必能够提高模型性能, 甚至可能会使模型性能降低。另外, 对比图 3 中四幅图可以看出不同取样间隔对压强预测模型的预测均方根误差 RMSEP 和决定系数  $R^2$  的影响。保持样本数一定, 当取最优主成分时, 压强取样间隔越小, 模型的预测均方根误差 RMSEP 值越小, 模型预测精度越高, 但此时对应压强模型的适用范围却越小。

### 3.2 不同训练集样本数的模型预测结果

图 4 为不同样本数的压强模型预测均方根误差 RMSEP 和决定系数  $R^2$ 。设置压强取样间隔为 0.05 MPa, 训练集样本数分别为 100 组, 75 组, 50 组, 25 组, 对应的模型适用范围 20.05~26.25, 20.05~25, 20.05~23.75 和 20.05~22.5 MPa。从本文光谱数据库中提取 250 条原始光谱曲线建立压强预测模型。

从图 4 的每幅图中同样可以看出, 主成分对压强预测模型的预测均方根误差 RMSEP 和决定系数  $R^2$  的影响与图 3 类似。在每个模型中, 存在一个最优主成分值使得模型的预测精度和稳定性最佳。另外, 对比图 4 中四幅图可以看出, 保持样本间隔一定, 当取最优主成分时, 压强取样数

越少, 模型的预测均方根误差 RMSEP 值越小, 模型预测精度越高, 但此时对应压强模型的适用范围却越小。在浓度预测模型和温度预测模型中, 训练集取样个数、训练集取样间隔和主成分个数对模型的影响与压强预测模型大致相同。

### 3.3 最佳预测模型及应用结果

本文根据预测精度、预测稳定性、模型适用范围、模型运算量等综合指标确定最优模型。由 3.1 节, 3.2 节可知, 模型的预测精度、预测稳定性、适用范围和运算量取决于训练集取样个数、训练集取样间隔和主成分个数。上述单一影响因素的改变并不能同时提高模型的所有性能, 需要对模型各项指标平衡选取。

在每个模型中, 当主成分数较小时, 模型性能较低, 当主成分数较大时, 运算量也较大, 但未必能够提高模型性能, 甚至可能会使模型性能降低。因此, 当精度满足要求时, 可以适当选择较小主成分数, 以减小运算量。当取最优主成分时, 保持样本数一定, 样本间隔越小, 模型的预测精度越高, 稳定性越好, 但模型适用范围较小; 保持样本间隔一定时, 样本数越少, 模型的预测精度越高, 稳定性越好, 但模型适用范围同样较小。当精度满足要求时, 可以适当选择间隔较大、样本数较多的训练样本, 以扩大模型的适用范围。综合考虑以上因素, 本文最终确定的三个预测模型的参数如表 1 所示。

本文将三个单自变量模型用于甲烷的温度, 浓度和压强的实际预测。选择 25 组样本的干涉光谱数据, 分别代入浓度、温度和压强模型, 图 5 为三个预测模型的甲烷浓度、温度和压强预测误差。

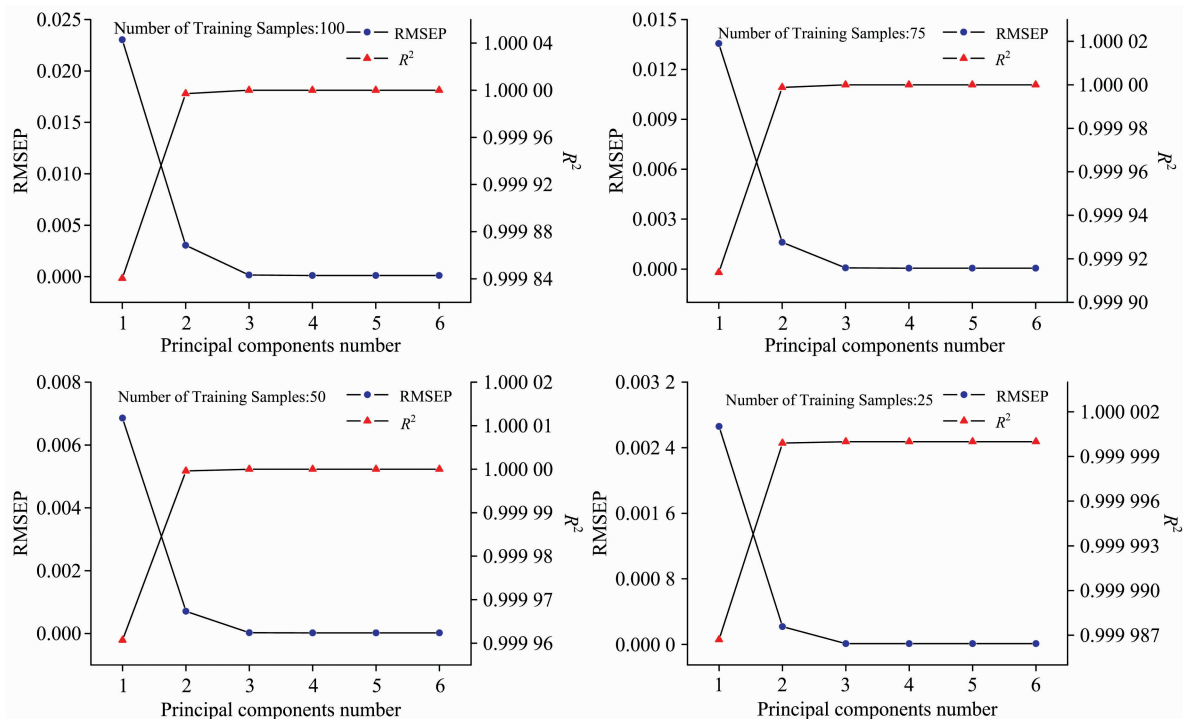


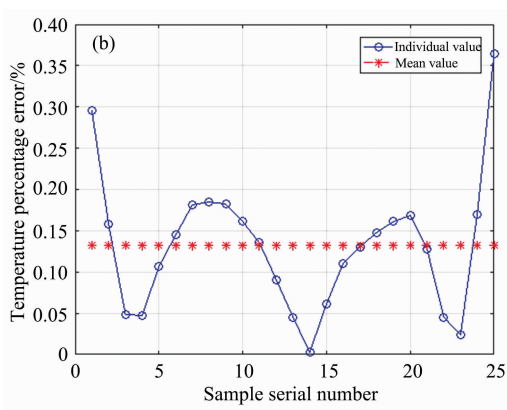
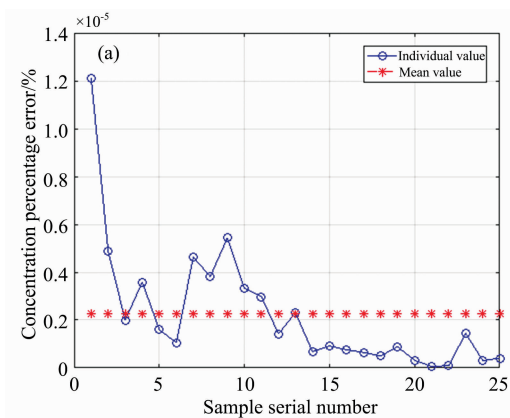
图 4 不同训练集样本数的压强预测模型的 RMSEP 和  $R^2$

Fig. 4 RMSEP and  $R^2$  of pressure prediction model for various numbers of training samples

表 1 浓度、温度和压强预测模型的参数

Table 1 Parameters of concentration, temperature and pressure prediction models

	Concentration model	Temperature model	Pressure model
Application ranges	5~375 mmol · L <sup>-1</sup> 670 K & 24 MPa	580~678 K 200 mmol · L <sup>-1</sup> & 20 MPa	10~34.5 MPa 200 mmol · L <sup>-1</sup> & 670 K
Training samples number	50	25	25
Prediction samples number	25	25	25
Training samples interval	5 mmol · L <sup>-1</sup>	2 K	0.5 MPa
Principle component number	2	2	5
RMSEP	3.082 × 10 <sup>-6</sup>	0.977 0	5.052 × 10 <sup>-3</sup>
$R^2$	0.999 9	0.998 9	0.999 9
Error	± 1.21 × 10 <sup>-7</sup>	± 3.63 × 10 <sup>-3</sup>	± 9.49 × 10 <sup>-4</sup>
Precision	± 45.4 nmol · L <sup>-1</sup>	± 2.5 K	± 3.3 × 10 <sup>-2</sup> MPa



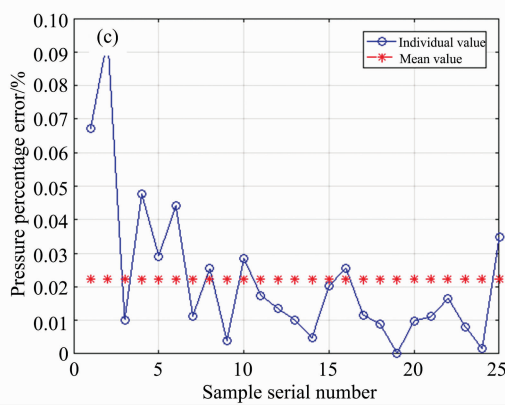


图 5 三个模型的预测误差

(a): 浓度误差; (b): 温度误差; (c): 压强误差

Fig. 5 Prediction errors of three models

(a): Concentration error; (b): Temperature error;  
(c): Pressure error

## 4 结 论

针对深海热液甲烷光学被动成像干涉系统 OPIIS 的数据

反演, 提出利用将成像干涉技术与偏最小二乘法相结合的方法精确、稳定、快速的提取甲烷浓度、温度和压强信息。首先分别建立甲烷辐射光谱与甲烷浓度、温度和压强的单因变量预测模型, 再利用干涉条纹与辐射光谱的关系, 间接建立干涉光谱与甲烷浓度、温度和压强的 PLS 预测模型, 可以提高预测模型在实际应用中的抗干扰能力和稳定性。另外, 分析了训练集取样个数、训练集取样间隔和主成分个数对模型预测性能的影响。在每个模型中, 存在一个最优主成分数值使得模型的预测精度和稳定性最佳。当取最优主成分数值, 保持样本数一定, 样本间隔越小, 模型的预测精度越高, 稳定性越好, 但模型适用范围较小; 保持样本间隔一定时, 样本数越少, 模型的预测精度越高, 稳定性越好, 但模型适用范围同样较小。平衡取舍模型预测精度、稳定性、适用范围和运算量等性能指标, 分别确定浓度、温度和压强三个最优单因变量 PLS 预测模型。将 25 组样本的干涉光谱应用于三个最优预测模型, 其浓度、温度和压强的预测误差分别为  $\pm 1.21 \times 10^{-7}$ ,  $\pm 3.63 \times 10^{-3}$ ,  $\pm 9.49 \times 10^{-4}$ , 对应的预测精度分别为  $\pm 45.4 \text{ nmol} \cdot \text{L}^{-1}$ ,  $\pm 2.5 \text{ K}$ ,  $\pm 3.3 \times 10^{-2} \text{ MPa}$ 。结果表明, 干涉光谱结合偏最小二乘的反演算法可以精确、稳定、快速的获取热液区甲烷气体的浓度、温度和压强信息。

## References

- [ 1 ] QIN Wei-hua, TAO Chen-chao, CAI Zhen, et al(秦伟华, 陶辰超, 蔡 真, 等). Marine Science(海洋科学), 2016, 40(6): 119.
- [ 2 ] Farias L, Sanzana K, Sanhueza-Guevara S, et al. Estuaries and Coasts, 2017, 40: 1592.
- [ 3 ] Li Yuhong, Zhan Liyang, Zhang Jiexia, et al. Acta Oceanologica Sinica, 2015, 34(6): 34.
- [ 4 ] SHEN Zheng-wei, SUN Chun-yan, HE Hui-ce, et al(申正伟, 孙春岩, 贺会策, 等). Journal of Ocean Technology (海洋技术学报), 2015, 34(5): 19.
- [ 5 ] Di Pengfei, Feng Dong, Chen Duofu. Continental Shelf Research, 2014, 81: 80.
- [ 6 ] Tang Yuanhe, Duan Xiaodong, Gao Haiyang, et al. Applied Optics, 2014, 53(11): 2273.
- [ 7 ] OUYANG Ai-guo, TANG Tian-yi, WANG Hai-yang, et al(欧阳爱国, 唐天义, 王海洋, 等). Chinese Optics(中国光学), 2017, 10(3): 363.
- [ 8 ] LAO Wan-li, HE Yu-chan, LI Gai-yun, et al(劳万里, 何玉婵, 李改云, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(1): 55.
- [ 9 ] TANG Yuan-he, WANG Shu-hua, CUN Jin, et al(唐远河, 王淑华, 崔 进, 等). Acta Physica Sinica(物理学报), 2016, 65(18): 184201.
- [ 10 ] Liu Qingsong, Tang Yuanhe, Qin Lin, et al. International Journal of Remote Sensing, 2019, 40(1): 1.

# Retrieval of Hydrothermal CH<sub>4</sub> Based on Interference Spectroscopy and PLS Methods

LIU Qing-song<sup>1,2</sup>, HU Bing-liang<sup>1\*</sup>, TANG Yuan-he<sup>3</sup>, YU Tao<sup>1</sup>, WANG Xue-ji<sup>1,2</sup>, LIU Yong-zheng<sup>1</sup>, YANG Peng<sup>4</sup>, WANG Hao-xuan<sup>3</sup>

1. Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China
2. University of Chinese Academy of Sciences, Beijing 100049, China
3. School of Science, Xi'an University of Technology, Xi'an 710048, China
4. Joint Laboratory for Ocean Observation and Detection, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266200, China

**Abstract** The methane (CH<sub>4</sub>) gas released by hydrothermal enters into the ocean and atmosphere successively by diffusing and causes inestimable effect on earth in physics, chemistry and biology. The principle and environment effect of abyssal hydrothermal still require further study because limited information is available about dissolved methane. In our previous work, we propose an optical passive imaging interference system (OPIIS) for the real-time detection and long-term observation of hydrothermal methane's concentration, temperature, and pressure. To accurately, stably, and rapidly obtain the information of hydrothermal methane from OPIIS's interferogram, this paper processes OPIIS's data by combining interference spectra and partial least squares (PLS) algorithm. We built three single-dependent variable models between methane radiance spectra and gas concentration, temperature and pressure, respectively. Then we can establish the PLS prediction model between interference fringes indirectly on the basis of relationship between interference fringes and radiance spectra, which can improve the capacity of resisting disturbance and stability of prediction models in practical application. On the basis of Lorentz profile, we build the deep ocean gas emission model different from atmosphere emission and obtain the synthetic methane radiance spectrum database at any concentration, temperature and pressure by using the methane spectral parameters from HITRAN2016 molecular spectroscopy database. The six spectral lines of methane in the range of 1.64~1.66 μm are selected for the PLS regression model between methane radiance spectra and gas concentration, temperature and pressure. Furthermore, this paper analyzes the contribution of number of training samples, interval of training samples and number of principal components to the improvement of the comprehensive performance of regression model. The 96 groups of concentration, temperature and pressure regression model are built by using different groups, intervals and principal components, and those regression models are cross-validated using 25 groups of prediction samples. The comparison results of those regression models' root mean square error of prediction (RMSEP) and coefficient of determination ( $R^2$ ) indicate that the change of single factors such as the number of training samples, the interval of training samples and the number of principal components can not improve the prediction model's comprehensive performance about prediction accuracy, stability, application scope and computation. Finally, the optimized model with balanced performance is determined with concentration, temperature and pressure application ranges at 5~375 mmol · L<sup>-1</sup>, 580~678 K, 10~34.5 MPa, training samples of concentration, temperature and pressure are 50 groups, 25 groups, 25 groups, intervals at 5 mmol · L<sup>-1</sup>, 2 K, 0.5 MPa, principal components are 2, 2, 5. The RMSEPs of concentration, temperature and pressure are  $3.082 \times 10^{-6}$ , 0.977 0,  $5.052 \times 10^{-3}$ , and  $R^2$ s are 0.999 9, 0.998 9, 0.999 9, respectively. The prediction errors of concentration, temperature and pressure are  $\pm 1.21 \times 10^{-7}$ ,  $\pm 3.63 \times 10^{-3}$ ,  $\pm 9.49 \times 10^{-4}$ , and the corresponding precisions are  $\pm 45.4$  nmol · L<sup>-1</sup>,  $\pm 2.5$  K,  $\pm 3.3 \times 10^{-2}$  MPa. The results indicate that this retrieval algorithm can accurately, stably, and rapidly obtain concentration, temperature and pressure of hydrothermal methane.

**Keywords** Hydrothermal methane; Partial least squares (PLS); Imaging interference; Retrieval

(Received Jun. 5, 2018; accepted Oct. 12, 2018)

\* Corresponding author