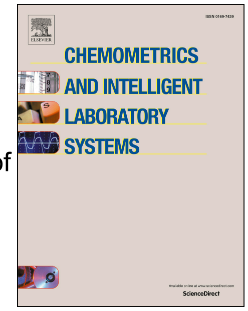
# Accepted Manuscript




Gaussian process regression coupled with MPT-AES for quantitative determination of multiple elements in ginseng

Yangwei Ying, Wei Jin, Yuwei Yan, Ying Mu, Qinhan Jin

PII: S0169-7439(17)30147-8

DOI: 10.1016/j.chemolab.2018.03.002

Reference: CHEMOM 3596

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received Date: 21 March 2017

Revised Date: 28 February 2018

Accepted Date: 6 March 2018

Please cite this article as: Y. Ying, W. Jin, Y. Yan, Y. Mu, Q. Jin, Gaussian process regression coupled with MPT-AES for quantitative determination of multiple elements in ginseng, *Chemometrics and Intelligent Laboratory Systems* (2018), doi: 10.1016/j.chemolab.2018.03.002.

# Gaussian process regression coupled with MPT-AES for quantitative determination of multiple elements in ginseng

Yangwei Ying[a], Wei Jin[a*], Yuwei Yan[a], Ying Mu[a], and Qinhan Jin[a]

**Abstract:** A comprehensive understanding of spectral information is the most reliable approach to identify ginseng characteristics and quality. More data is produced due to the use of a high resolution spectrometer, which in turn requires high-efficiency computational capability. To well address such problems, chemometrics plays a significant part and shows satisfactory results. In this paper, a Gaussian process regression (GPR) is proposed to couple microwave plasma torch-atomic emission spectrometry (MPT-AES) for quantitative determination of multiple elements in ginseng. GPR as a probabilistic method is capable of dealing with massive data and illustrating the outputs with probabilistic meanings. With high-dimension input variables involved, the principal component analysis (PCA) is proposed to reduce computational amount and improve variable representativeness. The results show that GPR is superior to the conventional approach compared with support vector regression (SVR) by evaluation indexes. All the work is performed based on real ginseng spectral data. Taking into account the computational amount caused by large datasets of scales, three approximation methods (subset of datasets (SD), subset of regressors (SR) and projected process (PP)) are applied to address this problem, which in a way shrink the training set with a certain similar result. It is concluded that the Gaussian process performs well in spectral analysis and has the potential for further practical applications.

**Key Words:** Gaussian process (GP), microwave plasma torch (MPT), quantitative analysis

## 1. Introduction

Ginseng is one of the most valuable herbal plants in China, and it is a well-known strong tonic with high medicinal value. It can be divided into a great number of types in terms of different origins. The better quality implies the higher nutritional value, but rarer it is. As a result, it will be more expensive. Thus, in order to make exorbitant profits, plenty of adulterate ginsengs lead to a serious and enormous challenge in the quality control of ginseng. Ginseng grown in different origins tends to possess different qualities and characteristics. The purpose of identification of ginseng can be achieved by identifying the origins. Therefore, establishing an efficient identification method of recognizing the quality of ginseng is significant. Also, it will improve the quality identification system and identify the ginseng species rapidly and stably. In recent years, conventional chemical analytical methods have been well used for qualitative and quantitative analysis of elements in a wide variety of samples, including plants. For example, Mao et al [1] identified the Chinese traditional medicine rapidly and reliably using molecular spectroscopy. Woo et al [2] classified the cultivation area of ginseng using near infrared spectroscopy and ICP-AES. Lu et al [3] applied two-dimensional near-infrared (NIR) correlation spectroscopy to the discrimination of Fructus Lycii of four different geographic regions successfully. A linear relationship exists between the spectral signal intensity and the elemental concentration of the specimen at a given wavelength, with a quantitative description achieved by the standard calibration curve. Building multivariate models associated with the spectral information is potential to predict the elemental concentrations accurately. Furthermore, it can match the real origin reliably and conveniently. As a consequence, to control the quality of the ginseng concerned.

Microwave plasma torch (MPT) is a newly developed excitation source for atomic emission spectroscopy, which has been widely applied in elemental analysis fields like heavy metals in the water, nonmetallic elements in the oil and rare earth elements detection. Moreover, HeMPT-AES has the potential to identify almost all elements in the Periodic Table. Fei et al [4] used MPT-AES coupled with online preconcentration-vapor generation to determine the content of iodine. Kong et al [5] determined the cobalt and nickel by coupling online preconcentration by ion-exchange with MPT-AES. Obviously, by combining the abundant spectral information thus obtained by MPT-AES with chemometric analysis methods, one can establish the multivariate calibration models of various samples accurately.

For multivariate calibration models, chemometrics is an essential and powerful tool for processing a vast amount of data rapidly. Currently, various multivariate analytical techniques have been applied in spectroscopic fields. Chen et al [6] determined the content of caffeine and total polyphenols in green tea rapidly using NIR spectroscopy with a partial least squares (PLS) algorithm to perform the calibration. He et al [7] predicted a variety of 40 unknown tea samples using near infrared spectroscopy coupled principal component analysis and BP model, and the recognition rate reached 100%. Bu et al [8] determined the Panax ginseng by near-infrared spectroscopy, and compared several multivariate regression methods and showed that the partial least squares regression (PLSR) was the most suitable one. It is well-known that the natural ginseng consists of complex components such as organic compounds and various chemical elements. Therefore, both linear and

[a] *Research Center for Analytical Instrumentation, Institute of Cyber-Systems and Control, State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, Zhejiang, P. R. China. E-mail: jinweimy@zju.edu.cn*, TEL: +86-571-88208383, FAX: +86-571-88208383