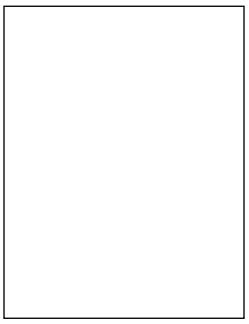

Title: Estimation of Sunflower Seed Yield Using Partial Least Squares Regression and Artificial Neural Network Models

Author: ZENG Wenzhi, XU Chi, ZHAO Gang, WU Jingwei, HUANG Jiesheng

PII: S1002-0160(17)60336-9
DOI: 10.1016/S1002-0160(17)60336-9
Reference: NA

To appear in:

Received date: NA
Revised date: NA
Accepted date: NA

Please cite this article as: ZENG Wenzhi, XU Chi, ZHAO Gang, WU Jingwei, HUANG Jiesheng, Estimation of Sunflower Seed Yield Using Partial Least Squares Regression and Artificial Neural Network Models, *Pedosphere* (2017), 10.1016/S1002-0160(17)60336-9.

PEDOSPHERE

Pedosphere ISSN 1002-0160/CN 32-1315/P doi:10.1016/S1002-0160(17)60336-9

Estimation of Sunflower Seed Yield Using Partial Least Squares Regression and Artificial Neural Network Models

ZENG Wenzhi[1], XU Chi[1], ZHAO Gang[2], WU Jingwei[1], HUANG Jiesheng[1]★

[1]*State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, China*

[2]*Crop Science Group, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Katzenburgweg 5, 53115 Bonn, DE*

★ Corresponding author. Email: sdjshuang@whu.edu.cn

ABSTRACT

Statistical models can efficiently establish the relationships between crop growth and environmental conditions, and explicitly quantify the uncertainties. This study aimed to test the efficiency of statistical models, established by partial least squares regression (*PLSR*) and artificial neural networks (*ANN*), in predicting the seed yields of sunflower. Two years field trial data about the sunflower growth under different salinity levels and nitrogen application rates in the Yichang Experimental Station in Hetao Irrigation District, Inner Mongolia, China were used to calibrate and validate the statistical models. The calculation of variable importance in the projection (*VIP*) was used to select the sensitive crop indices for seed yield prediction. We found that, by using the most sensitive indices as inputs for seed yield estimation, *PLSR* model could attain a comparable accuracy (*RMSE* = 0.93 t ha^{-1}, R^2 = 0.69) comparing with using all measured indices (*RMSE*=0.81 t ha^{-1} and R^2=0.77). The *ANN* outperformed the *PLSR* for yields predicting in different combination of inputs of both microplots and field data. The results indicated that sunflower seed yield could be reasonably estimated by using a small number of crop characteristic indices under complex environmental conditions and management options (e.g., saline soils, nitrogen application). Since leaf area index (*LAI*) and plant height (*H*) were found to be the most sensitive crop indices for sunflower seed prediction, it implicates that remotely sensed data and *ANN* may be joint for regional crop yield simulation.

Key Words: Artificial neural network; Partial least square regression; Salinity; Sunflower; Statistical crop model

INTRODUCTION

As a widely planted crop for the production of food and oil, more than 25 million hectares of Sunflower (*Helianthus annuus*) is distributed across the world croplands (National Sunflower Association, 2013). Since it is well adapted to moderately saline soil conditions (Francois, 1996), sunflower is one of the best crops in arid or semiarid regions undergoing soil salinization (Zeng *et al.*, 2014). The Hetao Irrigation District, Inner Mongolia, China, is one of the regions suffering from salinity due to low precipitation and high evapotranspiration, approximately 70% of cultivated lands are suffering from salinization (Shi *et al.*, 2012; Li *et al.*, 2014). The profit from selling the sunflower yield is becoming a major source of income for local farmers. In 2009, according to the local government, sunflower has grown on approximately 0.18 million hectares in Hetao Irrigation District, which