HYNUTRI: ESTIMATING THE NUTRITIONAL COMPOSITION OF WHEAT FROM MULTI-TEMPORAL PRISMA DATA

Mariana Belgiu[1], Michael Marshall[1], Mirco Boschetti[2], Monica Pepe[2], Alfred Stein[1], Caroline Lievens[1]

[1]Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands, m.belgiu@utwente.nl, m.t. marshall@utwente.nl, a.stein@utwente.nl, c.lievens@utwente.nl

[2] Institute for Electromagnetic Sensing of the Environment, Italian National Research Council (CNR-IREA), Via Diocleziano, 328, 80124 Napoli, Italy, boschetti.m@irea.cnr.it, pepe.m@irea.cnr.it

ABSTRACT

The goal of this work is to investigate the potential of PRecursore IperSpettrale della Missione Applicativa (PRISMA) hyperspectral data to predict the concentration of four macronutrients (K, P, N, S) and four micronutrients (Ca, Fe, Mg, Zn) in final wheat production. All investigated nutrients are essential to improving human nutrition. The initial findings indicate accurate predictions for Zn, P, Mg, S, K, Ca and Fe (R^2 ranging from 0.57 to 0.74). N was less accurately estimated (R^2 of 0.49). We conclude that the foliar chemical properties and temporal dynamics as detected by hyperspectral data translate successfully to the target micro- and macronutrients composition of the wheat production.

***Index Terms*—**malnutrition, food security, agriculture, hyperspectral imaging, remote sensing, machine learning

1. INTRODUCTION

The "environmental nutrition" domain addresses the sustainability of food systems in terms of nutritional outcomes, environmental impacts and feedbacks [1]. Nutritional outcomes are defined by the nutritional value of food intake, which is affected by the concentration of essential nutrients available in major crops produced locally. In addition, the nutritional composition of the consumed crops varies depending upon crop variety or species (i.e. genetics), environmental conditions (i.e. soil, total rainfall, relative humidity, temperature etc.) and farm management practices [2]. Previous studies have recommended the analysis of context-specific nutrient composition of crops in order to assess nutrient status and make appropriate dietary recommendations at the local level [3, 4]

The conventional method of measuring the nutritional composition of crops is to collect crop seed samples and perform chemical analysis of macro- and micro- nutrient composition [5]. This method is time-consuming and expensive, which invariably reduces the sample size and therefore robustness of the results. In this context, remote sensing technologies and data might provide a complementary or alternative solution for estimating the nutrient composition of crops. Previous remote sensing studies have focused on nutrients that increase food production and therefore abundance through N, P, K and S application [6], while steps are made at present to also study micro-nutrients. To the best of our knowledge, no comprehensive study of macro and micro-nutrients composition of the wheat seeds has been performed with spaceborne hyperspectral remote sensing.

In this study we aim to evaluate the potential of the PRISMA Hyperspectral Narrow Bands (HNB) data to predict the concentration of nutrients in the wheat seeds critical to evaluating the quality of crop yield. More specifically, we focus on predicting the concentration of four macro-nutrients (K, P, N and S) and four micronutrients (Ca, Fe, Mg, Zn) in final wheat production.

2. STUDY AREA AND DATASETS

A total number of 40 wheat samples were collected for the study area situated in Jolanda di Savoia, in the northern part of Italy. The distribution of the samples is presented in Fig. 1.

PRISMA is a hyperspectral remote sensing platform that is funded and managed by the Italian Space Agency. PRISMA was launched on 22 March 2019 with a 5-year mission. The sensor consists of 250 HNBs from 400-2500nm at a spatial resolution of 30m. Sixty-six bands are in the visible/very near-infrared (400-1010nm) and 171 bands are in the shortwave infrared (920-2505nm) at ~12nm intervals. Images available for our study are depicted in Table 1.

Table 1: PRISMA image acquisition dates

Date	Status
2020/04/07	Clean
2020/05/05	Haze/cloud contamination
2020/05/11	Cloud/cloud shadow
2020/05/17	Cloud/cloud shadow
2020/05/23	Clean
2020/26/06	Clean

IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9553614

We targeted at collecting PRISMA images during the sprouting, flowering/tasseling, and scenecense phases. Yet, PRISMA images during collected during the wheat sprouting phase were not available. In addition, three out of four images available in May (Table 1) were discarded because of the high cloud coverage. In the end, three images were used as input for developing the prediction models: April 04, May 23, and June 26. Smoothing Splines algorithm was used for noise removal purposes.

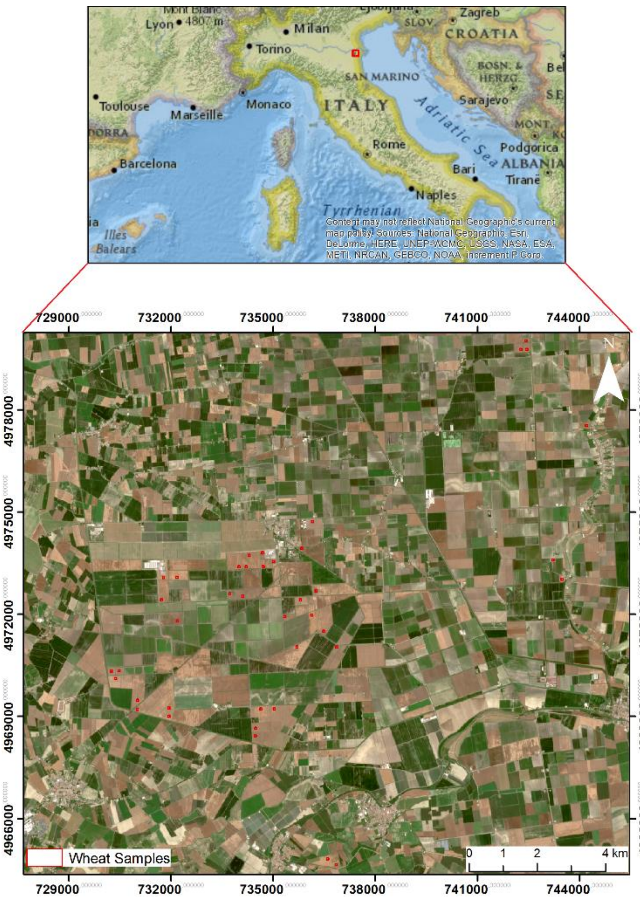

Fig. 1: Study area and spatial distribution of wheat samples

3. METHODOLOGY

The methodology adopted in our work involves the following steps:

(1) A field campaign to collect wheat samples (Elementary Sampling Unit -ESUs). Each ESU was selected in a way that was internally homogeneous but had spatial heterogeneity with other ESUs in terms of wheat variety, soil properties, management practices, etc. These factors increase the variability in nutrients required to test the robustness of our predictions. Each ESU consisted of a 60×60 m^2 area that was characterized by several 0.25×0.25 m^2 sub-samples randomly selected within the ESU. The seeds were separated from the clear cuts, dried and weighed.

(2) The chemical analysis of the wheat seeds using carbon hydrogen and nitrogen analyzer (CHN) for N and the inductively coupled plasma-optical emission *spectroscopy* (ICP/OES) for the remaining nutrients.

(3) Development of Random Forests (RF) regression models [7] to predict the concentrations of nutrients in the crops (measured as % for N and mg kg^{-1} for the remaining nutrients). RF is an ensemble of decision trees (*ntree* parameter) built by randomly selecting a user-defined number of input variables (*mtry* parameter) and a sample subset.

We developed RF prediction models for all target nutrients and PRISMA images. Following [8], the *ntree* parameter was set equal to 1000 and *mtry* as the square root of the total number of input spectral bands and calculated indices. To increase the robustness of the reported results, we ran 100 iterations for each RF-prediction model. The number of spectral bands was reduced using the backward feature elimination library implemented in the *caret* package available in R.

The prediction results are assessed using the coefficient of determination (R^2) and Root Mean Square Error (RMSE).

4. RESULTS AND DISCUSSION

Developed RF prediction models obtained accurate results for the target nutrients (Table 2 and Table 3). The image collected in April, which represents the peak of productivity of the wheat crop, proved to be the best predictor for K and Fe (0.63 R^2 and 0.57 R^2). The remaining nutrients were predicted using the image collected during the scenecense phase, namely May 23 and June 26 (Fig. 2).

Table 2: Nutrients prediction results (R^2) obtained using PRISMA input images collected on April 07, May 23, and June 26

Date	K	P	Zn	Mg	Ca	Fe	S	N
07.04	0,63	0,66	0,53	0,56	0,44	0,57	0,5	0,4
23.05	0,54	0,62	0,74	0,55	0,61	0,46	0,47	0,41
26.06	0,48	0,68	0,7	0,66	0,56	0,5	0,66	0,49

The highest R^2 value was obtained for Zn (0.74), followed by P (0.68), Mg (0.66), S (0.66), K (0.63), Ca (0.61) and Fe (0.57) (Fig. 2). N concentration in wheat seeds was predicted with an R^2 value equal to 0.49. N is associated with biomass accumulation and photosynthesis which does not automatically translate to grain yield. Therefore, it is not surprising that previous studies focused on foliar N concentration predictions and reported accurate results [9].

The spectral regions explaining the highest variability of target nutrients concentrations in the final wheat production are presented in Table 3. The input variables that explain most variation, i.e. the best predictors, for the RF models varied between different nutrients. Almost all nutrients were estimated using five predictors, expect for S that could be estimated with four predictors. Shortwave-infrared regions of the spectrum had an important contribution for predicting target nutrients (Table 3).

Table 2: Spectral bands used for predicting the target nutrients

Nutrient	Spectral bands				
K	B1626	B1636	B1656	B728	B522
P	B2483	B822	B2245	B1595	B2169
Zn	B2214	B2428	B538	B1784	B2252
Mg	B485	B2190	B2245	B1152	B1163
Ca	B343	B1273	B1784	B2252	B2061
Fe	B1491	B2335	B2061	B2086	B1078
S	B1636	B1646	B1656	B1328	
N	B1316	B1533	B515	B1328	B739

Preliminary results reported in this paper showed that the foliar chemical properties and temporal dynamics as detected by hyperspectral data translated successfully to the target micro- and macronutrients composition of the investigated crop.

Our future work will investigate additional data dimensionality reduction approaches (e.g. Principal Component Analysis) including those implemented in RF such as the backward feature reduction method proposed by Díaz-Uriarte and De Andres [10] or the conditional feature importance (CS) method proposed in [11]. In addition, we will perform a similar analysis for maize, soybean, and rice crops in the same study area. Different data smoothing algorithms will also be implemented to increase the quality of the PRISMA data.

5. CONCLUSIONS

The analysis performed in our study showed prediction accuracy (measured by R^2) ranging from 0.49 to 0.74.

The results of this work are relevant to food security programs dedicated to “integrating nutrition interventions into the agricultural investment plans developed in different countries” [12]. Specifically, the proposed methodology could be used to operationalize monitoring and forecast of the nutrient composition of croplands at the national, regional, and global levels, which are vital to emergency response and other short-term interventions.

The nutrient composition of crops is projected to decline under elevated CO2 concentrations (Smith and Myers 2018). This methodology could be used to address long-term planning via scenario-building. Ultimately, the coordinated use of such activities would assist in achieving Sustainable Development Goal (SDG) 2 (“Zero Hunger”) while indirectly assisting other SDGs: SDG 1 (“No Poverty”), SDG 12 (“Responsible Production and Consumption”), SDG 13 (“Climate Action”) and SDG 15 (“Life on Land”) [13].

ACKNOWLEDGEMENT

This work is funded by the European Space Agency (ESA) – EO Science for society permanently open call for proposals EOEO-5 Block 4, 4000130277/20/I-DT. More information about the ‘HyNutri: Sensing “Hidden Hunger” with Sentinel-2 and Hyperspectral’ can be found here: www.hynutri.nl.

6. REFERENCES

[1] J. Sabaté, H. Harwatt, and S. Soret, "Environmental nutrition: A new frontier for public health," *American journal of public health,* vol. 106, no. 5, pp. 815-821, 2016.

[2] M. Reguera *et al.*, "The impact of different agroecological conditions on the nutritional composition of quinoa seeds," *PeerJ,* vol. 6, p. e4442, 2018.

[3] S. A. Wood, "Nutritional functional trait diversity of crops in south-eastern Senegal," *Journal of Applied Ecology,* vol. 55, no. 1, pp. 81-91, 2018.

[4] G. Shively and C. Sununtnasuk, "Agricultural diversity and child stunting in Nepal," *The journal of development studies,* vol. 51, no. 8, pp. 1078-1096, 2015.

[5] E. M. Abdel-Rahman, O. Mutanga, J. Odindi, E. Adam, A. Odindo, and R. Ismail, "Estimating Swiss chard foliar macro-and micronutrient concentrations under different irrigation water sources using ground-based hyperspectral data and four partial least squares (PLS)-based (PLS1, PLS2, SPLS1 and SPLS2) regression algorithms," *Computers and Electronics in Agriculture,* vol. 132, pp. 21-33, 2017.

[6] D. J. Mulla, "Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps," *Biosystems engineering,* vol. 114, no. 4, pp. 358-371, 2013.

[7] L. Breiman, "Random Forest," *Machine Learning,* vol. 45, 2001.

[8] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 114, pp. 24-31, 4// 2016.

[9] G. R. Mahajan *et al.*, "Monitoring the Foliar Nutrients Status of Mango Using Spectroscopy-Based Spectral Indices and PLSR-Combined Machine Learning Models," *Remote Sensing,* vol. 13, no. 4, p. 641, 2021.

[10] R. Díaz-Uriarte and S. A. De Andres, "Gene selection and classification of microarray data using random forest," *BMC bioinformatics,* vol. 7, no. 1, p. 3, 2006.

[11] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," *BMC bioinformatics,* vol. 9, no. 1, p. 307, 2008.

[12] M. T. Ruel, A. R. Quisumbing, and M. Balagamwala, "Nutrition-sensitive agriculture: what have we learned so far?," *Global Food Security,* vol. 17, pp. 128-153, 2018.
[13] M. Herrero *et al.*, "Farming and the geography of nutrient production for human use: a transdisciplinary analysis," *The Lancet Planetary Health,* vol. 1, no. 1, pp. e33-e42, 2017.