معاونت پژوهش و فناوری

گزارش نهایی طرح تحقیقاتی
بررسی امکان پیش بینی برخی ویژگی های گیاهی گندم با استفاده از داده های طیف سنجی و محاسبات نرم

مجری (مجریان) طرح:
ولی رسولی شربیانی
یوسف عباسپور گیلانده
سینا فیض اله زاده اردبیلی
سما عمید

گروه مهندسی بیوسیستم، دانشکده کشاورزی و منابع طبیعی اردبیلی اجرا گردیده است.

تیرماه 1396
Prediction of Crop Growth Monitoring by Using Spectral Data and Soft Computing on Wheat

Researchers:
Vali Rasooli Sharabiani
Yousef Abbaspour Gilandeh
Sina Faiz-o-llah zade Ardabili
Sama Amid

Department of Biosystem Engineering
Faculty of Agriculture and Natural Resources

This Research Project Has Been Financially Supported by the Office of Vice Chancellor for Research

Date:
July 2017
Contents

Abstract ................................................................................................................................. 1

2 Introduction ..................................................................................................................... 2

Precision agriculture ........................................................................................................ 5
Need for an agriculture monitoring.................................................................................. 7
Crop monitoring ................................................................................................................ 9

Crop investigation .......................................................................................................... 10
Crop growth survey ......................................................................................................... 10
Prediction of crop yield .................................................................................................... 12

The role of remote sensing in crop monitoring............................................................... 13

Crop mapping .................................................................................................................... 13
The role of remote sensing in combination with crop growth models for crop yield prediction ....................................................................................................................... 14

Yield ................................................................................................................................. 15

Artificial neural network .................................................................................................. 16

Adaptive neuro fuzzy inference system (ANFIS) ............................................................. 19

Organization of study ...................................................................................................... 23

3 Material and methods ................................................................................................... 23

Experimental field .......................................................................................................... 23

Spad Value Measurements ............................................................................................. 25

Data processing ................................................................................................................. 27

RBF Modelling ................................................................................................................ 29

ANFIS Modelling ............................................................................................................. 31

4 Results and discussion ................................................................................................. 33

Primary results of Data .................................................................................................... 33

Evaluation of RBF model ................................................................................................. 34

Evaluating of ANFIS model ............................................................................................. 36

Comparing ANFIS and RBF models ................................................................................ 41

5 Conclusion ..................................................................................................................... 44
Table Contents

Table 2.1. Forward and backward pass for ANFIS .................................................. 21
Table 4.1. Selected property of the investigated crop variables. .......................... 33
Table 4.2. Performance indices (R and RMSE) for RBF models....................... 36
Table 4.3. Performance indices (R and RMSE) for ANFIS models..................... 40
Table 4.4. Average amount of performance indices (R and RMSE) for predictor models. .......................................................... 41
Figure contents

Fig 2.1. Anfis structure ................................................................. 20

Fig 3.1 Map of experimental field including blocks (different fertilizer), tractor paths for using remote sensing and two different places in two years ........................................ 24

Fig. 3.2 photo of experimental field on 2011 that difference of growth condition after additional fertilizer can be recognized................................................................. 25

Fig. 3.4. Average of reflectance spectrum of the different experimental treatments for years (n= 56 + 40).................................................................................. 28

Fig. 3.5. RBF neural network structure ........................................................................ 30

Fig. 4.1. Plots for predicted and actual values of growth status. Left plots indicates 2011 datasets and right plots indicates 2012 datasets....................................................... 35

Fig. 4.4. Plots for predicted and actual values of growth status for 2011 datasets. Left plots indicates results of Gaussian MFs and right plots indicates results of Trap MFs. 38

Fig. 4.5. Plots for predicted and actual values of growth status for 2012 datasets. Left plots indicates results of Gaussian MFs and right plots indicates results of Trap MFs. 39
Abstract

Recent advances in precision agriculture technology have led to the development of ground-based active remote sensors (or crop canopy sensors) that calculate normalized difference vegetation index (NDVI) readings. Vegetation indices obtained from remote sensing data can help to summarize climate conditions. Artificial neural networks (ANNs), another Artificial Intelligence approach, are one of the most efficient computational methods rather than other analytical and statistical techniques for spectral data. This study was employed experimental radial basis function (RBF) of ANN models and adaptive neural-fuzzy inference system to design a network in order to predict the SPAD, protein content and grain yield of wheat plant based on spectral reflectance value and to compare two models. Results indicated that the obtained results of RBF method with high average correlation coefficient (0.997, 0.997 and 0.996 in 2011 for SPAD, yield and protein, respectively and 0.994, 0.995 and 0.997 in 2012) and low average RMSE (0.271, 103.315 and 0.111 in 2011 for SPAD, yield and protein, respectively and 0.407, 105.482 and 0.096 in 2012) has the high accuracy and high performance compared to ANFIS models.

Keywords: Remote sensing; Spectral data analysis; ANFIS; Artificial neural networks; Vegetation index;
2 Introduction

In today's world, there is increasing concern with respect to the agriculture sector and the estimated longevity of a sufficient food production system. Environmental issues can hinder food production systems (e.g. soil erosion, water quality, climatic change), while socioeconomic issues can be equally as damaging. Concerns from both the consumer and the farm community stems from issues surrounding the volatility of the international agricultural marketplace, and the requirement of farmers to meet the food demands of an increasing population while maintaining quality. These factors coupled with increasing production costs have resulted in a substantial decrease in the farms and public concern surrounds the subsequent increase in industrial and corporate farms[1]. The adoption of sustainable agriculture practices by farmers involves daily management strategies that strive to protect the land resources required to grow food. A sustainable food production system has been defined as an agri-food sector that over the long term can simultaneously maintain environmental quality, provide economic and social rewards for all individuals involved in the system, and produce an adequate and accessible food supply. Essentially, if the food production system cannot meet these criteria then the system is deemed unsustainable. Site-specific agriculture is one approach to farm management that can promote sustainable agriculture. Site-specific agriculture, also known as precision agriculture, can be defined as the application of technologies and principles to manage spatial and temporal variability associated with all aspects of agricultural production to improve crop performance and environmental quality[2]. This style of agriculture practice refers to a "knowledge-based system" that allows farmers to manage variability at scales that are within a defined farm unit (e.g. section, quarter section) and to specific spatial regions of the farm unit where required[3]. Spatially variable crop yield can exist due to many factors such as soil nutrient and moisture content, topography, as well as insect and weed infestations that change over time. Site specific agriculture requires both spatial and temporal management, which in the case of
farming can require highly time-sensitive information over large agricultural fields. In the past this type of real-time information has not been easily accessible and farmers have treated fields as homogenous units applying average rates of crop inputs over the entire field. As a result of this practice, farmers tend to over or under-apply crop inputs (e.g. fertilizer) which can result in both economic loss and environmental contamination. More recently, the increased availability of remote sensing imagery accompanied by comprehensive site-specific crop management plans has offered farmers a more definitive means of implementing sustainable agricultural practices in large agricultural areas. Remote sensing can play a unique role in agriculture because it is a non-invasive, time-specific method of acquiring information about seasonally variable crop and soil conditions. Remote sensing is a geospatial tool often incorporated into a management strategy for the whole farm operation that together with the benefits of global navigation satellite system (GNSS) and Geographic Information Systems (GIS) can be used to develop Variable Rate Application (VRA) maps for crop inputs[4]. Launched commercial Earth Observation (EO) satellites can provide the spatial resolution, timeliness, and high quality imagery that site-specific agriculture requires[5]. Remote sensing as part of a site-specific agriculture management strategy can provide the farm enterprise with the ability to satisfy increasing environmental, economic, and market demands[6]. The monitoring of crops can be done by means of ground survey at the local scale. However, at a regional scale, remote sensing appears appropriate both in terms of spatial and temporal coverage [7]. Recent advances in precision agriculture technology have led to the development of ground-based active remote sensors (or crop canopy sensors) that calculate normalized difference vegetation index (NDVI) readings. Because of the amount of applications, the classification of crops using remote sensing images is an important topic in remote sensing research. The advantages of using remote sensing techniques, instead of field survey, are the lower cost and the possibility of covering large areas. Another important reason is that it is easier to update the classifications, due to the possibility of repeated time frequency of the data [7].
The leaf surface properties, internal structure, plant stress, and the concentration and distribution of biochemical components effects on plant reflectance; therefore, To assess plant biomass and the physiological status of a plant, the analysis of reflected light may be used [8]. Wavelengths in the red (R) and near-infrared (NIR) wavebands are frequently used for indirect measurements of plant characteristics [9].

One of the main goals of this study is to evaluate the prediction of vegetation indices for subsequent studies of optimal vegetation indices for prediction of crop growth monitoring. Vegetation indices obtained from remote sensing data can help to summarize climate conditions [10-12]. These methods are based on regression models between the final crop yields, the climate data and vegetation indices. Although these methods are widely used, they have the problem that predictions are site specific from local measurements and sometimes the spatial extrapolation is difficult. Various methods of mathematical and statistical analyses have been used for setting up linear and non-linear calibration models. Card et al [13] found that N in dried and ground tree leaves could be determined accurately from reflectance with a laboratory spectrometer. Stepwise multiple linear regression (SMLR) was used to select 580 nm and 480 nm for total nitrogen prediction ($R^2=0.90$). Hansen et al. [14] used multi-way partial least squares regression (N-PLS) to predict grain yield and protein content, and they showed that the relation between reflectance measurements and protein content was slightly better in wheat, where especially N-PLS improved the prediction of grain protein content. Lee et al. [15] found that SPAD (Soil and Plant Analyzer Development, Minolta, Inc.) readings, based on transmittance at 659 and 940 nm, were well correlated with N content in corn ear leaves ($R^2=0.962$). They developed prediction models by partial least squares (PLS) regression, principal component regression (PCR), and multiple linear regression (MLR). Their results showed that models built by PLS and PCR were better than models established from MLR. Rasooli Sharabian et al. [16] used the multivariate analysis including partial least squares regression (PLSR) and stepwise multiple linear regression (SMLR)
procedures to determine important wavelengths related to winter wheat growth characteristics in three consecutive years. The results showed strong relationships between predicted and actual crop variables. The best prediction model built on wavelengths selected by SMLR so that $R^2$ for the validation dataset were respectively 0.85, 0.89 and 0.84 for SPAD, grain yield and protein content. Tumbo et al. [17] used a back-propagation neural network model for corn nitrogen prediction in field conditions. The model used 201 spectral bands as input, covering a range of 407-940 nm, and results proved that the neural network model could considerably reduce interfering effects of cloud cover and solar angle. The model showed good correlation between predicted and actual chlorophyll meter readings of the training set ($R^2 = 0.91$). A good relationship was also found in the validation dataset ($R^2 = 0.74$).

But, there is a need for developing an alternative technique to determine crop growth status in the field. In the past, mathematical models were used to find the relationships between inputs and outputs of a production process. But this classic logic approach requires an exact definition of the mathematical model equations to describe the phenomenon. Artificial neural networks (ANNs), another Artificial Intelligence approach, are one of the most efficient computational methods rather than other analytical and statistical techniques [18].

**Precision agriculture**

The use of innovative technologies collectively named “Precision Agriculture” is a promising approach to optimize agricultural production of crops. In field crop production precision agriculture methodologies are applied to site-specific application of fertilizer or pesticides, automatic guidance of agricultural vehicles, product traceability, on-farm research or management of production systems[19]. Recently precision agriculture also enhances management decisions in livestock production, pasture management, viticulture, and horticulture[20]. Precision crop production aims to match agricultural input and practices to the spatial and
temporal variability within a field, instead of managing an entire field based on a hypothetical average. Small-scale site specific differences can lead to great differences in yield and quality, thus a better use of resources to preserve the quality and quantity of agricultural products with respect on environmental resources is essential[19]. The philosophy behind precision agriculture is not only including a direct economical optimization of agricultural production, it also stands for a reduction of harmful outputs into environment and non-target organisms. In particular a contamination of water, soil, and food resources with pesticides has to be minimized in crop production[21]. Against the background of food security and sustainable production, adequate technologies are fundamental for this agricultural practice[22]. The implementation of information-based management systems into crop production since the mid-1980s implies a huge potential to modernize the agricultural practice. Since then different techniques for the characterization of soils and crops have been engineered and included into decision making systems. For the future an information-driven crop production as a combination of geospatial and agricultural data management will encourage the actual utilization of precision agriculture applications[23]. Current research on precision agriculture for crop production focuses on the development of sensors for remote detection of crops and soil in real time. Relevant field parameters like soil properties, topography, water status, crop micro-climate, nutritional status, weeds, and pests and diseases as well as yield can be monitored and estimated. Integration of different remote sensing techniques and image analysis in combination with a global positioning system will be an essential step towards online application. Still one limiting factor of a successful use of precision agriculture is the interpretation of properties derived from sensor data, rather than the collection of relevant data[20]. The interpretation of information and its implementation into robust decision support systems will improve the acceptance and implementation of precision agriculture techniques.
**Need for an agriculture monitoring**

The economic and social importance of the agricultural sector in many regions of the world, together with the concern about world population increase, economic development and the uncertainty in the changes of production caused by climate change, made necessary the development of procedures and techniques to monitor the conditions of crops, to improve the crop field management and also to be able to make early prediction of crop production. This need for an efficient crop monitoring and management, as well as, the prediction of crop production is thus enhanced by climate change issues and by the changes in agriculture related to human activities.

Regarding to human activities, the Food and Agricultural Organization of the United Nations (FAO) states that the world population will increase at a rate of 43 million per year in the period 2045-2050[24]. This rise of human beings in the world will be a consequence of the population growth in developing countries (45 million), and prognosis is that half of this accruement will occur in the sub-Saharan Africa (23 millions). In those developing countries, especially in Africa, the increase of population will aggravate even more the current world undernourished state. It is expected that industrial countries will have some reactions for increasing food production in concordance to this population growth. Thus, there is a matter of fact that it exists a general concern about increasing agricultural production.

Furthermore in the frame of what is also known as food security strategies, there is an interest in predicting problems like pest infections and drought periods than can damage the crop production. In the large arid and semiarid regions of the world droughts are frequent and they commonly cause a decrease or a total failure of crop production, important economic losses in developed countries, and famine in undeveloped countries. An increase of some of these problems is expected with climate change, mainly in the Mediterranean region, which might be one of the most vulnerable regions to global change in Europe. Climate change projections for the Mediterranean region show a reduction of agricultural areas and losses of...
agricultural potential during the twentieth century[25] due to the pronounced decrease in precipitation that is predicted[26]. The current change of alimentary habits in some important emerging countries, like India and China is increasing the demand for agricultural products. In these countries, a growing sector of the population is becoming wealthy enough to change from a mostly vegetarian diet, based on rice and other cereals, to a diet that includes more meat. Livestock needs to be fed with cereals, which increases the demand and, therefore, market prices. Furthermore, some developed countries are increasing their production of biofuels, which diminishes the quantity of crops used for human consumption. This is also helping to increase the prices of cereals. Last, but not least, the upward trend of prices is attracting speculators to the markets. This increase of prices is causing political tensions, as happened during the spring of 2008 in countries like Haiti, where the prime minister had to resign, Cameroon, Senegal, but also Egypt or Thailand. A means for limiting price increases would be to increase production, which would need an increase of agricultural productivity. In fact, during the FAO summit of June 2008, it was stated that more investments should be done to increase agricultural productivity. The need for an increase of production has induced important changes in the agricultural practices during the last decades. For example the use of fertilizers has been extended worldwide and genetically modified crops are used as a solution for a “sustainable” production increase[27]. There is also a concern about the impacts associated with these new agricultural practices. The expansion of agriculture needs to be done in a sustainable way as the generalized use of fertilizers or over exploitation of water resources represents environmental risks and can even have consequences for human health. The increase intensive processes like irrigation and/or the abuse of fertilizers also might produce some negative consequences on water quality and the degradation of irrigated lands for instance as a consequence of salinization. Finally, another noticeable change in modern agriculture is that more and more frequently crop rotations are decided by market fluctuations and policy regulations. This introduces an additional dynamics in crop distribution, which make necessary to
update crop maps with a high temporal frequency. The other important issue that requires a system for crop monitoring is the impacts of climate change in agriculture. Studies conducted over the last decades have provided evidence about the modifications of several climatic parameters[28]. For instance, noticeable trends in surface temperatures have been recorded during the twentieth century at the global scale[29]. The feedback effect of climate change on agriculture is complex. The increase in temperature and the increase in the concentration of atmospheric CO$_2$ could affect the plant biological processes (photosynthesis, respiration, growth, etc)[30]. The fertilizing effect of the atmospheric carbon could produce a general increase of the vegetation activity and production. Nevertheless, the positive response of vegetation activity and production to climate change is only expected in areas with an adequate availability of water, on the contrary, the areas affected by an increase of temperatures and evapotranspiration together with a decrease of precipitation will suffer from a higher water stress in vegetation, which, in turn, would cause a decrease of the production[31]. Finally, extreme events (hot waves, droughts, extreme rainfalls) have a negative effect in crop production[32]. The undesirable impacts that climatic change can have in crop production show the strong requirement for the monitoring of crops at present and to be maintained in the future.

**Crop monitoring**

In the previous section it was highlighted that the relationship between agriculture and climate and the important changes in agriculture practices during the last part of the 20$^{th}$ century shows that agricultural monitoring systems are necessary. To be efficient, such systems should satisfy at least the three requirements listed hereafter: they should be able to provide a map of crops timely, to survey the growth of crops and if possible to predict the yields. Below, each of these requirements is discussed in more detail.
Crop investigation

The substantial increase of intensive agriculture together with the influence of the policy regulations and market demands leads to frequent changes in the surface meant to agriculture and in the distribution of crops within the land devoted to agriculture. Therefore, the timely identification, inventory and cartography of crops becomes necessary for estimations of crop yield. In addition to the crop production assessment, crop mapping is also useful for the management of water resources or the estimations of sequestration of carbon by the soil, among others.

Crop growth survey

Crop growth survey consists in the monitoring during the growth period of several crop and soil parameters, which are indicators of the plant condition, together with the actual plant phonological stage. Those parameters are for example plant height, LAI (Leaf Area Index), biomass or nitrogen content. Typically, the survey of crop growth is focused in the following issues, which are in-turn interconnected:

a) Phenology development: That is the succession of biological events during the plant life. The survey of phenology implies, for example, the observation of the exact moment in which certain crop organs appear (ex. wheat ears). Phenology is often simulated in terms of the sum of degree-days and crop specific characteristics, for instance vernalisation factors.

b) Canopy development: it can be quantified by the measurement of the LAI, the plant biomass of the plant height. In terms of biological processes, canopy development is the result of photosynthesis, respiration and biomass allocation. The amount of energy received and the capacity of the plant to use this energy will determine the biomass production. The amount of intercepted radiation is a function of the LAI. Only a part of the intercepted radiation, denoted as fPAR, is efficiently used by the crop and will be used for biomass accumulation. The way in which biomass partitioning is performed is specific to each cultivar type. In the modelling of canopy
development the high vegetative structural diversity is controlled by genetic variables that intervene in this partitioning.

c) Roots growth and uptake ability: the function of plant roots is to uptake water and nutrients from the soil. This is closely related to the soil chemical and physical properties as well as the soil moisture conditions. Any lack of nutrients, especially nitrogen, or any water deficiencies would negatively impact the plant development. The shortages in mineral content or basic nutrients in the soil can be detected with periodical analysis of soil samplings and compensate with fertilization. The monitoring of moisture conditions is also necessary. Regarding biological aspects, there is a big difference between the root system of annual crops (ex. wheat, corn, potatoes…) and perennial crops (ex. vineyards).

d) Water balance among the plant, the soil and the atmosphere. The water requirements of a crop in a particular moment depend on the environmental variables (ex. air temperature), the soil conditions and the crop phenology. The processes involved in the water balance include evaporation and transpiration, both in the soil and in the plant. The list of variables that take part in the water balance, mainly describing the soil status and soil water behavior, can be very extensive (soil albedo, drainage coefficients, etc…) but the most important is soil moisture.

e) Nitrogen balance in the soil and in the plant. The content of nitrogen in the soil can change as a result of organic decompositions, fertilization, etc. Crops absorb nitrogen through the roots system and fix it in their elements. The nitrogen content in the leaves is related to the chlorophyll content, which is easier to measure than nitrogen content.

The information obtained from the survey of the previous points through the quantification of several parameters is of great valuable for the management of fields and are the basis of the human interventions like the use of fertilizers or a particular irrigation schedule. However, the monitoring of the parameters of crops
along a growing season is expensive and time consuming, and therefore, there is a need for developing remote sensing techniques that will be useful in this context.

**Prediction of crop yield**

Several techniques have been used to obtain an early prediction of crop production, most of them based on previous climate conditions summarized by means of drought indices, vegetation indices obtained from remote sensing data[10, 12] and both of them[31]. These methods are based on regression models between the final crop yields, the climate data and vegetation indices. Although these methods are widely used, they have the problem that predictions are site specific from local measurements and sometimes the spatial extrapolation is difficult, as a consequence of the geographic and topographic diversity and the different crop types. To solve these problems, more complex models, based on biophysical processes, can also be used. These are likely to be more general than the statistical methods based on local regressions. A model of crop growth describes how a plant grows, that is, how the carbon is allocated in the plant. These models require daily meteorological data: incoming solar radiation, temperature and precipitation. Many models have been developed or adapted to a unique cultivar, a reduced number of them or to particular crop conditions like water stress, nitrogen stress, salinity conditions, etc. and make use of many parameters. Thus, the benefits of using a monitoring system that provides crop parameters describing canopy development, for instance LAI would be very important for model calibration, forcing, etc. A huge diversity of crop growth models exists in the scientific literature. Some well-known models and their related ‘families’ are SUCROS (Simple and Universal Crop Growth Simulator)[33] CERES (Crop Environment Resource Synthesis)[34] that was developed for cereals, CROPGRO is a family of grain legumes models and STICS (Simulator multi disciplinarian pour les Cultures Standard) developed at the INRA, France. There are also software “packages” like DDSAT (Decision Support System for Agro technology Transfer) and APSIM
(Agricultural Production Systems simulator) that integrate several of the previously cited models. Nevertheless, despite the great usefulness of these models, there are noticeable limitations concerning its calibration. Crop parameters describing canopy development and dynamic are commonly needed for the calibration of the models. This involves time and cost consuming field samplings and very often there is a lack of spatial representation, mainly in areas in which spatial diversity of crops, soil characteristics and climates are important. Therefore, due to these limitations, there is a need to develop methods based on remote sensing data, which allow the monitoring of crop parameters over large areas, to improve the yield prediction.

The role of remote sensing in crop monitoring
The monitoring of crops can be done by means of ground survey at the local scale. However, at a regional scale, remote sensing appears appropriate both in terms of spatial and temporal coverage.

**Crop mapping**
As it was said before, crop mapping is necessary in land change studies, climate change, hydrological studies and other applications like yield prediction and the efficient management of water resources, the later usually based in the estimates of evapotranspiration[35]. Crop maps are usually used in combination with crop growth models for yield prediction or to model for example soil carbon sequestration[36]. Because of the amount of applications, the classification of crops using remote sensing images is an important topic in remote sensing research. The advantages of using remote sensing techniques, instead of field survey, are the lower cost and the possibility of covering large areas. Another important reason is that it is easier to update the classifications, due to the possibility of repeated time frequency of the data. The use of optical remote sensing data is well established for crop mapping and the methodologies have been proved to be quasi operational. Crop
classification using optical data is often performed with data with a spatial resolution compatible with the field size: in general Landsat-TM or SPOT-HRV data at regional scale are used\cite{37, 38}. A well-known limitation of optical data is the presence of the cloud cover that prevents the acquisition of images at the desiderate time. Radar data, in contrast, has the advantage of being independent from cloud cover and thus show a high potential for crop classification. It may also happen that vegetation needs to be monitored at a specific phenology stage. This is the case, for example, when two crops have similar behavior during the growing season except for a specific development stage. However, satellite radar data have not often been used for this purpose\cite{39, 40}. Mainly because, until very recently, satellites were only able to measure single linear polarizations at a single frequency: ERS-1 and ERS-2 operate at C Band at VV polarization, RADARSAT operated at C band and HH polarization, JERS operated at L Band, HH polarization. Future missions will measure the complete scattering matrix at a single frequency and there is a need for developing adequate classification methods.

Several algorithms use radar data for the classification of crops. In a general way, they can be classified into knowledge-based approaches, classification by scattering mechanism and statistical data-driven methods\cite{41}. Knowledge based approaches are based on the analysis of the physics that determines the measured backscattering for each crop type. Those classifiers have the advantage of being more robust and easier to adapt to the specific conditions of the area to classify.

*The role of remote sensing in combination with crop growth models for crop yield prediction*

It is difficult for the models to account for the spatial heterogeneity in vegetation and soil conditions as well as the inherent difficulties of phenology modelling. Crop growth depends on many factors (weather, species, soil status, soil characteristics and management strategies) and, as a result, models need many parameters. For instance STICS v3.0 depends on 132 parameters\cite{42}. It is frequent
that some of these parameters, like the sowing date, are unknown, or need to be adjusted for each crop type or geographical location. One solution consists in calibrating the models using measurements of biophysical parameters\cite{33, 43}. LAI, which accounts for the leaf surface intercepting in-coming radiation, and biomass are key variables to calibrate crop growth models.

The calibration can be done with in-situ measurement of biophysical parameters. However, in-situ measurements are expensive and time consuming and generally can only be done at a limited number of fields. Thus, calibration has the risk of becoming site and cultivar-specific. In this context satellite remote sensing is useful when integrated in the models of crop growth as it provides spatial information on actual vegetation status. Remote sensing can be used to estimate key variables in the models: LAI, aboveground biomass and other crop characteristics like chlorophyll or nitrogen content. This information can be integrated in the calibration process using for example forcing methodologies\cite{44}.

\textit{Yield}

In remote sensing agricultural applications, the most spatially accurate yield data available today is obtained using a yield monitor coupled with a Differential Global Positioning System (DGPS). A DGPS yield monitor is placed on the combine at the time of harvest and captures position as well as crop volume and moisture readings on a per second basis. The DGPS receiver allows the yield data to be "stamped" with a geographic coordinate and enables the yield across the field to be mapped. Most DGPS receivers used in agriculture today are 12 channel and use phase smoothed pseudo-range positioning to permit sub-meter accuracy\cite{6}. A typical example is the Trimble AgGPS 106 differential GPS antenna and receiver. The yield monitor units used to represent yield data can vary by both the yield monitor and the manufacturers software used for data post-processing. In the North American marketplace yield is represented as bushels per acre (bu/ac), kilograms per hectare (kg/ha), or tons per hectare (t/ha). Generally, yield monitors provide an accurate and reliable source of information for farmers over time\cite{45}. Yield maps
can be visualized in a raw format represented by a set of yield points, or points can be interpolated into a continuous map surface. The goal of yield map interpretation is enhanced profitability through better control of natural and management induced sources of yield variation. Successful yield mapping is heavily dependent on the auxiliary information from the farmer such as field history (e.g. soil type, perennial weed regions, and crop rotation), the analysts' geo statistical knowledge (e.g. appropriate data interpolation methods) and the available GIS tools. Sources of yield variation are not always easily identifiable and can be a result of weather, soil-water relationships, soil physical and chemical properties, slope and aspect of a region, pest infestation, crop inputs, field history, and cultural practices and errors. The yield map can be used as a seasonal "report card" whereby farmers can evaluate how well the crop performed due to the implementation of new site-specific management strategies.

**Artificial neural network**

Stuart[46] described that NIR is an analytical tool technique based on the vibrations of the atoms of a molecule. This atom’s vibration would show the infrared spectrum which commonly appears by passing the spectrum radiation through a sample and determining what fraction of the incident radiation is absorbed at a particular energy. The infrared spectrum then will give sufficient information about compound structure of the samples in the form of spectral data. In order to take the benefit and potential used of NIR for classification problem hence the use of an appropriate statistical method is needed to process these NIR spectral data. One such statistical technique that has found many applications and has proved to be useful in assigning an object to an appropriate group or classification is Neural Network (NN). NN is a tool that is useful in classification problem. NN is adopted from a biological model system with computational method which consists of some processing units in the system. NN is also known as modern computational statistical method in terms of soft computing technique.
with some adaptive nature in system. Though NN is a powerful method, several studies [47-50] have cited problems in the NN process, especially in network architecture process. Ozturk [48] noted that actually NN has its own effective algorithm to solve the neural architecture problems, but the technique is based only on the iterative procedure and takes a long time to learn the whole process. Hence, these often lead to inefficient and complex network architecture in NN. To address these problems, Ozturk[48] and Yalkin and korkmaz [49] proposed an optimization technique by combining NN with genetic algorithm (GA) to improve the convergence speed and reduce the computational complexity of network architecture in NN. This technique thus is known as Genetic Algorithm Neural Network (GANN).

Dimitris et al[51] explored the applicability of artificial neural networks (ANN) for predicting the spread of structural response under the presence of uncertain parameters described as random fields. The use of ANN is carried out in combination with Monte Carlo simulation (MCS) for calculating response statistics in stochastic analysis of structural systems using finite elements. To this extent, the ANNs are trained with a few samples, following a conventional MCS procedure and used henceforth to predict the stochastic response for the rest of samples. The basic idea is to achieve a dimensionality reduction of the input ANN training space by using as input vector the random phase angles of the spectral representation method instead of the random variables describing the uncertain input parameters. A further improvement of the efficiency of the proposed approach is achieved by exploiting the uniform distribution of the random phase angles, in order to span efficiently the training space using a Latin hypercube sampling (LHS) technique. Spectral input selection and structure optimization is an important step for the design of optimized retrieval algorithms based on NN for satellite data inversion. This also allows the identification of the most informative input wavelengths for the specific inverse problem under consideration.
Selito tested two methods for the input wavenumber selection/reduction and input spectral sensitivity analysis: the extended pruning (EP) and the auto associative neural network (AANN) techniques.

An ANN is a computational model inspired by networks of biological neurons, wherein the neurons compute output values from inputs. It learns from its past experience and errors in a non-linear parallel processing manner. The neuron is the basic calculating entities which computes from a number of inputs and deliver one output comparing with threshold value and turned on (fired). The computational processing is done by internal structural arrangement consists of hidden layers which utilizes the back propagation and feed forward mechanism to deliver output close to accuracy[52]. Neural network is a predicting method that has learning ability to approximate any nonlinear function [53]. The main advantage of ANNs over statistical methods is that they require no assumptions about the form of a fitting function. Instead, the network is trained with experimental data to find the relationship; so they have become very popular as an estimating tool and have known to be efficient and less time consuming in modeling of complex systems compared to other mathematical models such as regression[54]. Since agricultural systems and technologies are quite complicated and uncertain, ANNs can be widely applied for modeling of different components in this sector. There are many types of artificial neural networks (ANN). Feed forward neural network, Radial basis function (RBF) network, Kohonen self-organizing network, learning vector quantization, Recurrent neural network, Modular neural networks, Physical neural network and others, are different types of neural networks[55, 56]. Radial basis function (RBF), one of the neural network types, was selected as one of the prediction networks in this study because it is a powerful technique for interpolation in multidimensional space. A RBF is a function which has built into it a distance criterion with respect to a center. Radial basis functions have been applied in the area of neural networks where they may be used as a replacement for the sigmoidal hidden layer transfer characteristic in multi-layer perceptrons. RBF networks have two layers of processing: In the first, input is mapped onto
each RBF in the 'hidden' layer. The RBF chosen is usually a Gaussian. In regression problems the output layer is then a linear combination of hidden layer values representing mean predicted output. The interpretation of this output layer value is the same as a regression model in statistics. In classification problems the output layer is typically a sigmoid function of a linear combination of hidden layer values, representing a posterior probability. Performance in both cases is often improved by shrinkage techniques, known as ridge regression in classical statistics and known to correspond to a prior belief in small parameter values (and therefore smooth output functions) in a Bayesian framework. RBF networks have the advantage of not suffering from local minima in the same way as Multi-Layer Perceptron. This is because the only parameters that are adjusted in the learning process are the linear mapping from hidden layer to output layer. Linearity ensures that the error surface is quadratic and therefore has a single easily found minimum. In regression problems this can be found in one matrix operation. In classification problems the fixed non-linearity introduced by the sigmoid output function is most efficiently dealt with using iteratively re-weighted least squares. Although the implementation is very different, RBF neural networks are conceptually similar to K-Nearest Neighbor (k-NN) models. The basic idea is that a predicted target value of an item is likely to be about the same as other items that have close values of the predictor variables.

Adaptive neuro fuzzy inference system (ANFIS)

ANFIS is a kind of adaptive neuro-fuzzy inference system which connects fuzzy logic system with neural network and constructs a hybrid intelligent system and benefits from the advantages of both fuzzy logic and neural networks, and its efficiency in very accurate models has been proved[57]. Using a given input/output data set, the toolbox function ANFIS constructs a fuzzy inference system (FIS) whose membership function parameters are tuned (adjusted) using either a back-propagation algorithm alone or in combination with a least squares type of method.
(hybrid). The neuro-adaptive learning method works similarly to that of neural networks, and provides a learning method for the fuzzy modeling[58]. In ANFIS, both of the learning capabilities of a neural network and reasoning capabilities of fuzzy logic were combined in order to give enhanced prediction capabilities, as compared to using a single methodology alone. The goal of ANFIS is to find a model or mapping that will correctly associate the input values with the target values. The fuzzy inference system (FIS) is a knowledge representation where each fuzzy rule describes a local behavior of the system. The network structure that implements FIS and employs hybrid-learning rules to train is called ANFIS. As seen from Fig. 2.1, different layers of ANFIS have different nodes[59].

![Anfis structure](image)

**Fig 2.1. Anfis structure**

Different layers with their associated nodes are described below:

Layer 1. Every node I in this layer is an adaptive node. Parameters in this layer are called premise parameters.

Layer 2. Every node in this layer is a fixed node labeled P, whose output is the product of all the incoming signals. Each node output represents the firing strength of a rule.

Layer 3. Every node in this layer is a fixed node labeled N. The $i^{th}$ node calculates the ratio of the $i^{th}$ rules’ firing strength. Thus the outputs of this layer are called normalized firing strengths.
Layer 4. Every node $i$ in this layer is an adaptive node. Parameters in this layer are referred to as consequent parameters.

Layer 5. The single node in this layer is a fixed node labeled $R$, which computes the overall output as the summation of all incoming signals. Each node in a layer is either fixed or adaptive.

The learning algorithm for ANFIS is a hybrid algorithm, which is a combination of gradient descent and the least-squares method. More specifically, in the forward pass of the hybrid learning algorithm, node outputs go forward until layer 4 and the consequent parameters are identified by the least-squares method[59]. In the backward pass, the error signals propagate backwards and the premise parameters are updated by gradient descent. Table 2.1 summarizes the activities in each pass.

<table>
<thead>
<tr>
<th></th>
<th>Forward Pass</th>
<th>Backward Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Premise parameters</td>
<td>Fixed</td>
<td>Gradient descent</td>
</tr>
<tr>
<td>Consequent parameters</td>
<td>Least-squares estimator</td>
<td>Fixed</td>
</tr>
<tr>
<td>Signals</td>
<td>Node outputs</td>
<td>Error signals</td>
</tr>
</tbody>
</table>

Method used in neural networks. The overall output can be expressed as a linear combination of the consequent parameters. The error measure to train the above-mentioned ANFIS is defined as[60]:

$$E = \sum_{k=1}^{n} (f^k - f'^k)^2$$

Where $f^k$ and $f'^k$ are the $k^{th}$ desired and estimated output, respectively, and $n$ is the total number of pairs (inputs–outputs) of data in the training set.

The main aim of this study was employed experimental radial basis function (RBF) of ANN models and adaptive neural-fuzzy inference system to design a network in order to predict the SPAD, protein content and grain yield of wheat plant based on spectral reflectance value and to compare two models. This work contained 3 steps. The first step is to process the requirement data. Second step develops
prediction network method and the last step offers and compares the obtained results.
3 Material and methods

Experimental field

An experimental winter wheat field (Fig. 2.1) was conducted in the farming area. A conventional variety “Omid” of winter wheat in the studied area was cultivated. According to the map (Fig 3.1) the field was in south side but in 2011 it shifted to the north part as a rotation. Dimensions of this field were 40 m x 120 m. The field was divided into 8 areas (blocks) and 0, 30, 60, and 90 kg ha⁻¹ fertilizer (ammonium nitrate) was applied with two repetitions at the regrown stage, so that the difference of the growth condition can be appeared (Fig. 3.2). After the bloom stage, growth investigation was done. Reflectance data by using a Spectroradiometer (Field Spec 3), SPAD value, and height of crop, number of stem, nitrogen content and protein content of wheat ear was taken from target points as reference area which was randomly set in the field.
Fig 3.1 Map of experimental field including blocks (different fertilizer), tractor paths for using remote sensing and two different places in two years
Fig. 3.2 photo of experimental field on 2011 that difference of growth condition after additional fertilizer can be recognized

**SPAD Value Measurements**

Chlorophyll content of the leaf is an important parameter for plant physiologists and agriculturists; it is used as an indicator of leaf senescence and nitrogen status of plants and can be altered in response to environmental stresses. There are methods now that are able to determine approximate chlorophyll content in the leaf
non-destructively, using a measurement of the leaf transmittance (T) at certain suitable wavelength(s).

Several instruments based on this principle are now available including a SPAD-502 chlorophyll meter (Konica Minolta Sensing, Japan). This chlorophyll meter measures intensity of light transmitted through the leaf sample at two wavelengths (650 and 940 nm) using light emitting diodes with approximate half width of the emission spectrum of 15 and 50 nm, respectively. The display shows a value M in relative SPAD units[61].

The quantity M is defined as E.q 5:

\[
M = k \left( \log(I'(940)/I(940))-\log(I'(650)/I(650)) \right) + C = k \left( \log T(940) - \log T(650) \right) + C
\]  

(5)

Where I (650) and I (940) are signals without the sample and I’ (650) and I’ (940) signals with the sample and log is a common logarithm. The quantity k (a confidential proportionality coefficient) defines the relative SPAD units, and C is the compensation value adjustable in the instrument software. For practical usage it is supposed that the negative common logarithm of the transmittance T at 650 nm related to that at 940 nm is proportional to the chlorophyll content. In each individual experiment, the reading of the SPAD-502 chlorophyll meter should be calibrated for the real chlorophyll content in the leaf as their relationship can differ among species or cultivars as well as among plants grown under different conditions. Also the producer declares that SPAD reading can vary in dependence on the leaf type.
Data processing

The required reflectance spectral data were obtained on three different dates (2011, 2012) with four growth stages in each year using a portable spectral diameter, FieldSpec_3 (FS3) (Analytical Spectral Devices, Inc., USA). These growth stages were GS36, GS37, GS45 and GS60 [62]. Field in-season measurements including measurements of SPAD value (soil plant analysis development) and canopy reflectance in 2011 at the flag leaf stage (GS 37) and anthesis stage (GS 60) were done in 56 target points as well as in the 2012 (40 target points) after the stem elongation (GS 36) and anthesis stage (GS 60) growth investigations were performed. The protein content and grain yield were measured after harvesting and threshing a 1 m × 3 m area in each target point in each of the two years. The SPAD value was determined the relative amount of chlorophyll concentration in plant leaves by measuring the absorbance of the leaf in two wavelength regions of red
(650 nm) and near infrared (940 nm) by using a SPAD meter (MINOLTA Co. LTD.). According to the catalogue of SPAD 502 [63] there is strong relationship ($R^2 > 0.9$) between SPAD value and leaf nitrogen concentration, and SPAD value has therefore been widely used for estimation crop chlorophyll and nitrogen contents and for guidance of plant health status and topdressing [64]. In this study, SPAD value was used as an index of actual nitrogen content in crop leaves. Wheat canopy reflectance measurements in the 350-2500 nm wavebands (1 nm in width) were made under cloudless conditions and as close to solar noon as possible [16]. The first 50 readings (from 350 nm to 400 nm) at the lower visible wavelengths and last the 1150 readings (from 1350 nm to 2500 nm) at the shortwave infrared (SWIR) were deleted due to their low signal-to-noise ratio; thus, the revised spectra began at 400 nm (Fig. 3.4). The reflectance on wavelength range of 400-1350 was considered as independent variable and variables of SPAD, Yield and Protein were considered as dependent variables.

![Reflectance Spectrum](image)

Fig. 3.4. Average of reflectance spectrum of the different experimental treatments for years ($n=56+40$).

Based on Fig. 3.4, Generally, the reflectance of 2010 in green and red visible (VIS) and middle infrared waveband (MIR) was higher and about infrared (NIR) region was lower than reflectance of 2011 and 2012 on average.
When simulation of complex systems is required and limited amount of experimental data is available, ANN is used to simulate system performance. A neural network consists of a large number of neurons or processing elements connected by synaptic weights. The used network in this study is Radial Basis Function network (RBF). RBF is a feed forward network and local type of learning which responds only to a limited section of input space. In one hidden layer of a RBF network, hidden node maps measure the distances between input vectors and center vectors to outputs by means of a nonlinear kernel or radial function [65].

The basis of a RBF neural network is supervised learning. RBF networks were independently proposed by several researchers to increase training procedures, and to produce precise approximations with simpler network architecture and an alternative population of RBF networks is then compared to the other neural networks [66]. RBF neural networks are used for nonlinear function approximation, data classification, systems modeling, and control [67]. They have a feed-forward architecture composed of one input layer, one hidden layer with a non-linear RBF activation function, and one linear output layer. The most important characteristic of RBF networks is the hidden layer neurons in the middle of the basis function that produce only local reactions for the input function. This is the reason that the basic function can produce a significant nonzero response although the input space falls only in a local area. In other cases, the output of the basic functions may be small [68]. The structure of a RBF network is presented in Fig.3.5.
The output of network is:

\[ y_k = \sum_{j=1}^{M} w_{kj} \phi_j(x) + w_{k0} \]  

(6)

Where \( M \) is the number of basic functions, \( x \) is the input data vector, \( w_{kj} \) represents a weighted connection between the basic function and output layer, and \( \phi_j \) is the nonlinear function of the \( j \)th unit, which is typically a Gaussian function:

\[ \phi_j(x) = \exp \left( -\frac{\| x - \mu_j \|}{2\sigma_j^2} \right) \]  

(7)

where \( x \) and \( \mu \) are the input and the center of RBF unit, respectively, and \( \sigma_j \) is the spread of the Gaussian basis function [69].

A least mean square (LMS) algorithm is used to optimize the weights once the centers of the RBF units have been determined. There are two ways to choose the centers: randomly or using clustering algorithms. In this study, centers were selected randomly from a data set.

Percentages of reflectance of each wavelength were the inputs of network and SPAD, Yield and Protein were the outputs of network. In order to train network, after defining the inputs and outputs to the network, there was a need to define the optimum number of neurons in the hidden layer. For this purpose, 5 numbers of neurons defined as an initial number of neurons. After running the network with 5 neurons, 5 neurons added to primary neurons and training process was repeated. It
was observed that during the training process of the network with increase of neuron numbers on hidden layer, the error of network decreases, based on performance plot of network. After each repetition 5 neurons were added to the previous neurons number. This action continued until the error reduction was converted into a horizontal line. This number of neurons selected as the optimum number of neurons on hidden layer.

**ANFIS Modellings**

The adaptive network-based fuzzy inference system (ANFIS) is a combination of artificial neural network (ANN) and fuzzy inference system (FIS) that was introduced to overcome the disadvantages of ANN and FIS [70]. ANFIS uses a feed-forward network to search for fuzzy decision rules that perform well on a given task during the training process. The ANFIS structure consists of five layers. As an example we assume the ANFIS model with two inputs and one output. For a first order Sugeno fuzzy model, two fuzzy if–then rules are considered:

**Rule 1:** If \( x \) is \( A_1 \) and \( y \) is \( B_1 \); then \( f_1 = p_1x + q_1y + r_1 \)

**Rule 2:** If \( x \) is \( A_2 \) and \( y \) is \( B_2 \); then \( f_2 = p_2x + q_2y + r_2 \)

where \( A_1, A_2 \) and \( B_1, B_2 \) are the fuzzy sets for inputs \( x \) and \( y \), respectively, \( p_1, q_1, r_1 \) and \( p_2, q_2, r_2 \) are the parameters of the output function that are determined during the training process ANFIS [71-73].

In this study, in order to test and train the ANFIS network, modeling was performed using ANFIS toolbox in MATLAB R2012a and Sugeno-type fuzzy inference systems were used in the modeling process. In order to classify the input data and make the rules, grid partition method is utilized, because of a few input variables. Two different types of input member functions (MFs) including trapezoidal (Trap MF) and Gaussian (Gauss MF) were used to model the network.
A linear function was used as output MFs and the hybrid learning algorithm was employed to model the predicted values. The validation and comparing performance of RBF and ANFIS models were checked out using the comparing parameters such as correlation coefficient (r) and the root mean square error (RMSE) as follow.

\[
R = \left(1 - \frac{\sum_{i=1}^{n} (z_i - z'_i)^2}{\sum_{i=1}^{n} z_i^2}\right)^{1/2}
\]

(9)

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (z_i - z'_i)^2}
\]

(10)

Where \(z\) is the target value and \(z'\) is the predicted value by ANFIS network [74]. The root mean square error (RMSE) is frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed (target values) and Pearson correlation (r) is a measure of the linear correlation between two variables (here are predicted and target variables) and is as a measure of the degree of linear dependence between two variables.
4 Results and discussion

Primary results of Data

Based on the first stage of materials and methods, this section presents the results of experimental treatments on trial data. This results includes different years, different levels of nitrogen application and strategies with a wide range of variation within the investigated crop. Variables (Table 4.1). This wide range in the investigated crop variables was planned in order to make the relationship between plant performance and reflectance measurements. Based on the Fig.2.1 one of the reasons that reflectance in the visible region in 2011 increased compared to 2012 might be the difference in soil background [75]. Another reason might be the higher SPAD value (hence, nitrogen concentrations of leaves and stems) in 2010, as shown in Table 2.1, as an increase in chlorophyll concentration causes increased reflectance in the visible regions and movement of the red edge to longer wavelengths [76], the position of the red edge in 2010 (around 720) was different from that in other years (around 700 nm).

<table>
<thead>
<tr>
<th>Crop variables</th>
<th>Year</th>
<th>n</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPAD (\text{-})</td>
<td>2011</td>
<td>56</td>
<td>43.1</td>
<td>1.38</td>
<td>39.4</td>
<td>45.3</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>40</td>
<td>42.6</td>
<td>1.62</td>
<td>38.2</td>
<td>46</td>
</tr>
<tr>
<td>Yield (Kg ha\text{-}1)</td>
<td>2011</td>
<td>56</td>
<td>6864</td>
<td>966</td>
<td>5516</td>
<td>8549</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>40</td>
<td>6708</td>
<td>615</td>
<td>5450</td>
<td>8154</td>
</tr>
<tr>
<td>Protein (%)</td>
<td>2011</td>
<td>56</td>
<td>11.4</td>
<td>0.99</td>
<td>8.86</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>40</td>
<td>11.6</td>
<td>1.55</td>
<td>9.62</td>
<td>14.0</td>
</tr>
</tbody>
</table>

\(^a\) \(n\) is the number of samples

Based on Table. 4.1 the mean and standard deviation value for SPAD in 2010 is respectively the highest and lowest compared to 2011 and 2012, for yield variable 2010 has highest mean value and 2012 has the low standard deviation and for
protein variable 2010 has the lowest standard deviation and 2012 has the highest mean value. This results can effect on the performance of the prediction model.

**Evaluation of RBF models**

In this study, the RBF method of artificial neural network was developed for modelling a network to predict the growth status of winter wheat based on reflectance measurements. 25 and 30 number of neurons were the optimum number of neurons on hidden layer for 2011 and 2012 data sets, respectively. 70 percent of data randomly selected as training data and the rest of data selected as testing data by network. The outputs of network was extracted and performance of network was calculated on predicting growth statues on two years for four growth stage (GS36, GS37, GS45 and GS60) using presented comparing parameters on materials and methods. As an initial result, Fig. 4.1 indicates the results of RBF network on the modeling dataset for two years on GS36, the predicted values of crop variables were plotted against the actual data.
The plots show high determination coefficient and high correlation coefficient of predicted values against actual values for 2011 and 2012 datasets. These results represent proximity of predicted and actual values. In order to display the results statistically, the calculated results tabulated on Table 4.2. High R value and low RMSE value, increases the accuracy of network.
<table>
<thead>
<tr>
<th>Year</th>
<th>Year</th>
<th>SPAD (-)</th>
<th>Yield (kg/ha)</th>
<th>Protein (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>r</td>
<td>RMSE</td>
<td>r</td>
</tr>
<tr>
<td>2011</td>
<td>GS36</td>
<td>0.998</td>
<td>0.104</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>GS37</td>
<td>0.998</td>
<td>0.142</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>GS45</td>
<td>0.998</td>
<td>0.170</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>GS60</td>
<td>0.995</td>
<td>0.669</td>
<td>0.997</td>
</tr>
<tr>
<td>2012</td>
<td>GS36</td>
<td>0.998</td>
<td>0.340</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>GS37</td>
<td>0.999</td>
<td>0.072</td>
<td>0.987</td>
</tr>
<tr>
<td></td>
<td>GS45</td>
<td>0.985</td>
<td>0.809</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>GS60</td>
<td>0.997</td>
<td>0.408</td>
<td>0.999</td>
</tr>
</tbody>
</table>

As shown on Table 4.3, for each two years from GS36 to GS60, there is a relative decreasing on predicting accuracy and correlation due to decreasing correlation coefficient and increasing RMSE value for SPAD, yield and protein except GS 37 on 2012. Based on Table 4.3, due to proximity of the results of correlation coefficient, the factor of RMSE will be better than correlation coefficient to compare the results. The results show a good relationship between actual crop variables and predicted values for validation datasets. Based on Table 4.2, GS36 of 2011 has high correlation coefficient (0.998, 0.998 and 0.998 for SPAD, Yield and Protein, respectively) and low RMSE (0.104, 75.59 and 0.075 for SPAD, Yield and Protein, respectively) compared to other growth stages and on 2012, GS37 has high correlation (0.999 and 0.997) and low RMSE (0.072 and 0.046) for SPAD and Protein, respectively and GS36 with correlation value of 0.999 and RMSE value of 27.22 has the best resubliming the other growth stages. This results of prediction model were obtained without pre-processing operations on data sets.

**Evaluating of ANFIS models**

70% of the data were used to generate the model, and the remaining (30%) were used for prediction. The initial ANFIS model was generated by grid partition
method. Fuzzification of input data, was performed by two different types of MFs. After training process the ANFIS models were tested using independent data set. The outputs of network was extracted and performance of network was calculated on predicting growth statues on two years for four growth stage (GS36, GS37, GS45 and GS60) using presented comparing parameters on materials and methods. As an initial result, Fig. 4.4 indicates the results of ANFIS on the modeling dataset for two years. The relationship between target and predicted values by using two different types of MFs was indicated on Fig. 4.4 and Fig. 4.5 as an initial and schematic result, the predicted values of crop variables were plotted against the actual data.
Fig. 4.4. Plots for predicted and actual values of growth status for 2011 datasets. Left plots indicates results of Gaussian MFs and right plots indicates results of Trap MFs.
Based on the results of Fig. 4.4 and Fig. 4.5, the obtained results of Gaussian membership functions have the high correlation and linear relationship compared to the obtained results from trap membership functions which reflects the high ability and high accuracy of Gaussian MFs for learning compared to Trap MFs. Table 4.3 presents the statically results of ANFIS performance.
### Table 4.3. Performance indices (R and RMSE) for ANFIS models.

<table>
<thead>
<tr>
<th>Year</th>
<th>MF Type</th>
<th>SPAD ((^{-}))</th>
<th>Yield (kg/ha)</th>
<th>Protein (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(r)</td>
<td>RMSE</td>
<td>(r)</td>
</tr>
<tr>
<td>2011</td>
<td>Gauss</td>
<td>GS36</td>
<td>0.9657</td>
<td>0.4523</td>
</tr>
<tr>
<td></td>
<td>Trap</td>
<td></td>
<td>0.6819</td>
<td>1.0887</td>
</tr>
<tr>
<td></td>
<td>Gauss</td>
<td>GS37</td>
<td>0.9287</td>
<td>0.832</td>
</tr>
<tr>
<td></td>
<td>Trap</td>
<td></td>
<td>0.7929</td>
<td>1.5669</td>
</tr>
<tr>
<td></td>
<td>Gauss</td>
<td>GS45</td>
<td>0.9997</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>Trap</td>
<td></td>
<td>0.9079</td>
<td>1.4751</td>
</tr>
<tr>
<td></td>
<td>Gauss</td>
<td>GS60</td>
<td>0.9952</td>
<td>0.652</td>
</tr>
<tr>
<td></td>
<td>Trap</td>
<td></td>
<td>0.982</td>
<td>1.2762</td>
</tr>
<tr>
<td>2012</td>
<td>Gauss</td>
<td>GS36</td>
<td>1</td>
<td>0.0021</td>
</tr>
<tr>
<td></td>
<td>Trap</td>
<td></td>
<td>0.9815</td>
<td>0.9750</td>
</tr>
<tr>
<td></td>
<td>Gauss</td>
<td>GS37</td>
<td>1</td>
<td>0.0144</td>
</tr>
<tr>
<td></td>
<td>Trap</td>
<td></td>
<td>0.9276</td>
<td>1.7632</td>
</tr>
<tr>
<td></td>
<td>Gauss</td>
<td>GS45</td>
<td>0.969</td>
<td>1.2126</td>
</tr>
<tr>
<td></td>
<td>Trap</td>
<td></td>
<td>0.9002</td>
<td>2.1393</td>
</tr>
<tr>
<td></td>
<td>Gauss</td>
<td>GS60</td>
<td>0.9851</td>
<td>0.9926</td>
</tr>
<tr>
<td></td>
<td>Trap</td>
<td></td>
<td>0.9626</td>
<td>1.5212</td>
</tr>
</tbody>
</table>

Based on the results of table 4.3, Gaussian MFs has the high correlation coefficient and low RMSE values compared to Trap MFs, it means that the output of ANFIS network on Gaussian MFs has the high accuracy and low difference with target values compared to Trap MFs. This claim is true for each two years (2011 and 2012) on each four stages (GS36, GS37, GS45 and GS60) for all dependent variables (SPAD, Yield and Protein). For example, the results of GS36 on 2011 are as follow.

For SPAD, yield and protein, Gaussian MFs has high correlation (0.9657, 0.9382 and 0.9609, respectively) and low difference between output and target values based on the RMSE values (0.4523, 525.685 and 0.3464, respectively) compared to Trap MFs. This result is also true for the rest of the year and growth stages, without exception, but in some cases has high intensity (GS36 on 2011) and in some cases has low intensity (GS60 on 2011 and 2012). Therefore, due to the best
results of Gaussian MFs compared to Trap MFs, it was chosen as the best type of membership function and the network was trained with Gaussian function.

### Comparing ANFIS and RBF models

In order to compare the results of two networks we need statics and attributable information. Fig. 4.3 and Fig. 4.4 take a graphically, initial and simple result of comparing two methods. Based on the results of Fig. 4.3, Fig. 4.4 and Fig. 4.5, it can be taken that the linear relationship of RBF results is stronger than ANFIS results due to its high values of determination coefficient and correlation coefficient and also about ANFIS results, Gaussian MFs showed high linear relationship between target and output values compared to Trap MFs. To prove this claim, the static results extracted and tabulated on Table 4.4 Because of the extent of output and the impossibility of comparing parameters one by one, these results tabulated in terms of the average value of comparative parameters for each network separately on table. 4.4.

<table>
<thead>
<tr>
<th>Year</th>
<th>Net</th>
<th>SPAD</th>
<th>Yield</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>r</td>
<td>RMSE</td>
<td>r</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>ANFIS</td>
<td>Gauss</td>
<td>0.972</td>
<td>0.507</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trap</td>
<td>0.866</td>
<td>1.352</td>
</tr>
<tr>
<td></td>
<td>RBF</td>
<td>-</td>
<td>0.997</td>
<td>0.271</td>
</tr>
<tr>
<td>2012</td>
<td>ANFIS</td>
<td>Gauss</td>
<td>0.988</td>
<td>0.555</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trap</td>
<td>0.943</td>
<td>1.599</td>
</tr>
<tr>
<td></td>
<td>RBF</td>
<td>-</td>
<td>0.995</td>
<td>0.407</td>
</tr>
</tbody>
</table>

As shown on Table. 4.4, RBF network with high correlation coefficient (0.997, 0.997 and 0.995 in 2011 for SPAD, yield and protein, respectively and 0.995, 0.995 and 0.997 in 2012) and low RMSE (0.271, 103.315 and 0.111 in 2011 for SPAD, yield and protein, respectively and 0.407, 105.482 and 0.096 in 2012) shows high
accuracy, high linear relationship and high performance, compared to ANFIS. Between two types of membership functions of ANFIS model developing, based on Table 4.4, Gaussian MFs with high correlation coefficient (0.972, 0.971 and 0.983 in 2011 for SPAD, yield and protein, respectively and 0.988, 0.990 and 0.993 in 2012) and low RMSE (0.507, 393.388 and 0.166 in 2011 for SPAD, yield and protein, respectively and 0.555, 172.077 and 0.159 in 2012) has high performance an high accuracy compared to Trap MFs.

From the other hand, based on the results of Table 3, due to the gap in the numbers of Table 4.3 for ANFIS output, the results show significant instability in the provision and unlike ANFIS, RBF in addition to high performance (Table 4.4) shows instability in presenting the results (Table 4.2). According to the results, the final decision is to select and to recommend the RBF method among three learning methods (RBF, ANFIS with gauss MFs and ANFIS with Trap MFs) as the best and precise predictor method of SPAD, Yield and protein using wavelength as the only independent input of network.

Rasooli sharabian et al. [16] on a study about determining the important wavelength using multivariate analysis including partial least squares regression (PLSR) and stepwise multiple linear regression (SMLR) procedures for prediction of winter wheat growth status and grain yield, reported strong relationships between predicted and actual crop variables. The best prediction model selected by SMLR on maximum data normalization for determination coefficient (R²) and root mean square error (RMSR) were 0.84, 1.94 for SPAD, 0.87, 301 for grain yield, and 0.80, 0.786 for protein content. On a study by Gupta et al. [77] Two types of feed forward back propagation neural network (FFBPNN) models were developed for the estimation of rice crop growth variables namely FFBPANN-I and FFBPANN-II model. The FFBPANN-I model was developed with one input neuron (HH- or VV- polarized scattering coefficient) and one output neuron (biomass or leaf area index or plant height or chlorophyll content) while the FFBPANN-II model was developed with two input neurons (HH- and VV polarized scattering coefficient) and four output neurons (biomass, leaf area index,
plant height and chlorophyll content). Results indicated a good estimating performance for HH- and VV polarized scattering. In general, correlation coefficient, and root mean square error between observed and FFBPANN simulated values of crop variables was reported as 0.993 and 0.118 for HH-Polarization, 0.982 and 0.157 for VV- polarization and 0997 and 0.057 for combination of HH and VV polarizations. On a other study by Liu et al. [78] a back propagation (BP) neural-network model was developed to estimate chlorophyll concentration in rice under heavy metal stress on three experiment farms located in Changchun, Jilin Province, China with level II pollution, with level I pollution and with safe level. The correlation coefficient ($R^2$) between the measured chlorophyll concentration and the predicted chlorophyll concentration was 0.9014, and the root mean square error (RMSE) was 2.58.

The obtained results of present study is following the previous researches. The positive point of this study is to use different methods (RBF and ANFIS) for predicting growth status of winter wheat. Based on the results of the other predicting methods that was conducted in predicting winter wheat status, it can be said the present methods have high ability in prediction of variables. This is due to the nature of soft computing techniques.

Important wavelength were obtained using regression analysis on SPSS software using the output values of the best predictor (RBF network). Some wavelengths [(410, 450, 550, 580, 660, 665, 720, 740, 800, 930, 990, 1010, 1110, 1120, 1150, 1240 and 1340 nm), (410, 450, 550, 580, 660, 720, 740, 800, 930, 990, 1010, 1110, 1120, 1150, 1240 and 1340 nm) and (410, 450, 550, 580, 660, 720, 740, 800, 930, 990 and 1010)] were identified by RBF as significant wavelengths for SPAD, grain yield and protein content, respectively for 2012 and [(410, 420, 450, 510, 530, 560, 590, 640, 680, 690, 710, 730, 750, 770, 780, 940, 1000, 1010, 1070, 1170, 1130, 1290 and 1350), (410, 420, 450, 510, 530, 560, 590, 640, 680, 690, 710, 730, 750, 770, 780, 940, 1000, 1010, 1070, 1130, 1170, 1290 and 1350) and (450, 510, 530, 560, 590, 640, 680, 690, 710, 730, 750, 770, 1010, 1070, 1170, 1290 and 1350)] were identified for SPAD, grain yield and protein content, respectively for 2011.
5 Conclusion

Remote sensing has great potential for several applications because it enables wide area, non-destructive, and real-time acquisition of information on Eco physiological plant conditions. Remotely sensed data, obtained either by satellite, aircraft or grand-based platforms, can provide a set of detailed and spatially distributed data on plant growth and development. Remote sensing can be used to delineate crop biophysical parameters, and several ground-based biophysical parameters, that can be related to the remotely sensed canopy using empirical methods. On the other hand, the growth response of vegetation in relation to measured or predicted climatic variables can be monitored by multispectral vegetation indices resulting from canopy reflectance in a relatively wide waveband. Vegetation indices such as the perpendicular vegetation index and the normalized difference have been developed to monitor vegetative growth at all stages. Vegetation indices are mostly ratios or linear combinations of signals from radiometer bands. These indices provide more highly correlated relationships than individual bands with vegetation parameters including green leaf area index, wet and dry biomass, percent cover by vegetation, plant height, fraction of leaf chlorosis, and leaf water content.

Field in season measurements including SPAD value, canopy reflectance using an active N-sensor embedded on the tractor as a ground-based platform with an RTK-GPS and solar sensor, spectral reflectance data using a portable spectroradiometer, height of the crop, nitrogen content of leaves, protein content of grain and grain yield were done after the stem elongation.

Precision farming (site-specific management, prescription farming, and variable rate application technology) is an information and technology-based agricultural management system to identify, analyze, and manage site-soil spatial and temporal variability within fields for optimum profitability, sustainability, and protection of the environment. This implies the concept of using information about variability
in site and climatic characteristics to manage specific sites in a field by using management practices. The main components of precision farming are global navigation satellite system (GNSS), geographic information systems (GIS), remote sensing (RS), and variable rate application (VRA). Precision agriculture requires reliable technology to acquire accurate information on crop conditions. Based on this information, the amount of fertilizers and pesticides for the site-specific crop management can be optimized. A ground-based optical sensor and instrumentation system was developed to measure real-time crop conditions.

The goal of this study was to model the SPAD, yield and protein of plant using wave length using soft computing methods. ANFIS and RBF were selected as the predictor of system and were trained and tested. Results indicated that the obtained results of RBF method with high average correlation coefficient (0.997, 0.997 and 0.996 in 2011 for SPAD, yield and protein, respectively and 0.994, 0.995 and 0.997 in 2012) and low average RMSE (0.271, 103.315 and 0.111 in 2011 for SPAD, yield and protein, respectively and 0.407, 105.482 and 0.096 in 2012) has the high accuracy and high performance compared to ANFIS models, then it was selected as the best predictor. Eventually, RBF network was proposed as the estimator network for studied outputs based on related input and was used to obtain the important wavelength.
References


[58] M. Mostafaei, H. Javadikia, L. Naderloo, Modeling the effects of ultrasound power and reactor dimension on the biodiesel production yield: Comparison of prediction
abilities between response surface methodology (RSM) and adaptive neuro-fuzzy inference system (ANFIS), Energy, 115 (2016) 626-636.


Final Report of Research Project

Prediction of Crop Growth Monitoring by Using Spectral Data and Soft Computing on Wheat

Researchers:
Vali Rasooli Sharabiani
Yousef Abbaspour Gilandeh
Sina Faiz-o-llah zade Ardabili
Sama Amid

Department of Biosystem Engineering
Faculty of Agriculture and Natural Resources

This Research Project Has Been Financially Supported by the Office of Vice Chancellor for Research

Date:
July 2017