EARLY FORECASTING CORN YIELD USING GROUND TRUTH DATA AND VEGETATION HEALTH INDICES IN BULGARIA

FELIX KOGAN\textsuperscript{2}; ZORNITSA POPOVA\textsuperscript{1*}; RAMESH SINGH\textsuperscript{3}; PETRA ALEXANDROVA\textsuperscript{1}

\textsuperscript{1}Institute of Soil Science Agrotechnology and Plant Protection “N. Poushkarov”, Sofia, Bulgaria,
\textsuperscript{2}Center for Satellite Application and Research, NESDIS, NOAA, Washington DC, USA
\textsuperscript{3}Chapman University, Schmid College of Science & Technology, Orange, California, USA

Abstract


Weather-related maize crop yield losses due to the transition from a planned state to market economy and the increasing climate uncertainties and drought aggravation have been a concern for farmers and policy-makers in Bulgaria since 1990. This paper discusses the possibilities to use operational satellite-based vegetation health (VH) indices for modelling maize crop yield relative to semi-early A1 and late A2 cultivar technology for early warning of drought-related grain losses. The indices were tested in Pleven oblast (Gorni Dabnik) and Burgas oblast (Sadievo) that represent main grain productive regions of North-West and South-East Bulgaria. Correlation and regression analysis were applied to model maize gain yield observed in the experimental fields of Gorni Dabnik and Sadievo from VH indices during 1982-1991. Strong correlations between Pleven maize grain yield relative to semi-early A1 maize varieties and VH indices were found during the critical period of maize development, which starts in May (week 16) and ends in June (week 23) for technology A1B1. For the late cultivar technology A2B1, the critical period of maize starts in June (week 22) and ends much latter in August-Sept (weeks 32 and 41). Relative to Burgas, for corn late cultivar A2, strong correlations of yield deviations from the trend produced by the A2B1 technology d\(Y_i\) with VH indices occur during week 27 and week 28 (July). Several models were constructed where VH indices could serve as independent variables (predictors). Thus, drought-related corn yield losses relative to semi-early and late cultivars could be predicted in Pleven oblast and Burgas oblast in advance of harvest and official grain production statistic is released.

Key words: Corn yield forecasting, long-term field experiments, satellite – base vegetation health indices, Correlation and Regression analysis

Introduction

During the first 15 years of the twenty-first century and last two decades of the previous century, South East Europe including Bulgaria, like most of the other regions of the world, experienced the impact of increased climate variability, rising temperatures and increased frequency of droughts (Gregoric (Ed.) 2012; Kogan and Guo, 2016; Popova et al., 2012; 2014; 2015). Droughts of severe-to-exceptional intensity covered 7-16% of world land (Kogan et al., 2013). These droughts had adverse consequences for societal sustainability worldwide since they reduced agricultural production, caused shortage of food and much related harm. In Bulgaria, due to the combined impacts of transition from a planned state to a market economy with increasing climate uncertainties and droughts frequency, maize crop yield losses have been a concern for

Corresponding author: zornitsa_popova@abv.bg
farmers and policy-makers since 1990.

The climate in the most plains of Bulgaria is moderate to transitional continental with semi-arid features. Plovdiv (La 42°09' Lg 24°45', Alt 160 m) and Stara Zagora (La 42°25', Lg 25°39', Alt 169 m) in the Thracian Lowland and Sandanski (La 41°34', Lg 23°17', Alt 206 m) experience the warmest and driest climate, while Sofia (La 42°15', Lg 25°45', Alt 555 m) is one of the coolest and wettest agricultural region in this country (Map 1).

The summer is wetter in the Danube plain than in the Thrace. However, some territories around Pleven (La 43°25', Lg 24°36', Alt 134 m), Silistra (La 44°07', Lg 27°16', Alt 16 m) and Varna (La 43°12', Lg 27°55', Alt 39 m) are drought prone (Alexandrov (Ed.), 2011; Popova, (Ed.) 2012).

To cope with the situation of crop yield losses, in our previous studies, vulnerability assessment to agricultural drought was carried out by using climate data trend test analyses, simulations with the soil water balance, irrigation scheduling and crop yield evaluation WINISAREG model (Pereira et al., 2003; Stewart et al., 1997) and application of standard precipitation index SPI (McKee, 1993; Pereira et al., 2010) over the period 1951-2004 (Popova et al., 2012; 2014; 2015). The model was validated using independent data sets relative to long-term field experiments with late and semi-early maize hybrids and soils of different total available water (TAW) in various locations (Popova, Eneva, Pereira, 2006; Popova, 2008; Popova and Pereira, 2011; Ivanova and Popova, 2011). Simulations were performed for the eight regions (called Oblasti in Bg), representing the varieties of climate and soil in Bulgarian plains, as referred above (Popova (Ed) 2012). Results have shown that rainfed maize is associated with great yield variability (29< Cv<72%) in this country. The most variable yields were found for southern locations (Sandanski, Stara Zagora and Plovdiv) when rainfed maize was grown on soils of low total available water (TAW). The variability of respective yields in the Danube Plain (Pleven, Silistra, Varna and Lom) proved to be much lower (30< Cv<55%) than that in the Thracian Lowland (40< Cv<70%). Considering an economical relative yield decrease threshold of potential maize productivity, resulted in 30% years of risk in Plovdiv, 20% in Sofia and 63% in Sandanski. Results for North Bulgaria have shown lower impacts, where only 10% of the years are risky in Pleven and Silistra. It has been observed that risky years increase when TAW decreases.

It has been also found that a version of seasonal SPI2 for ”July-Aug”, that is an average of the index during the period of high crop sensitivity to water stress, could be a good indicator of maize vulnerability to drought. For South Bulgaria and soils of large TAW (180 mm m⁻¹), economical losses are produced when SPI2 “July-Aug” < 0.2 in Sandanski, < -0.50 in Plovdiv and Stara Zagora and < -0.90 in Sofia. In North Bulgaria, the threshold “July-Aug” SPI2 ranges between -0.75 (Lom) and -1.5 (Pleven). The results proved that rainfed agriculture is related to high economical losses also along the Black Sea coast (Varna) and in Lom during normal years of SPI2 “July-Aug” less than +0.20.

This paper, in addition to our previous studies on vulnerability assessment of agricultural drought referred above (Popova et al, 2012; 2014; 2015), discusses the possibilities to use the cost effective operational satellite-based vegetation health (VH) indices for modelling maize crop yield well in advance for early warning of drought related risk of grain losses. The study is carried out by using datasets from long-term field experiments in Gorni Dabnik, Pleven oblast, (43°21’La; 24°21’Lg; 149 m Alt) and Sadievo, Burgas Oblast, (42°32’La; 26°03’ Lg; 154 m Alt) (Map ; Stoyanov, 2008) and NOAA1 operational space technology of satellite – base vegetation health indices (Kidwell et al., 1997). The expected output is: (a) finding whether experimental corn yields could correlate strongly with VH indices during the critical period of crop development; (b) investigating whether on such basis VH indices can be used as indicators of corn yields; and (c) building statistical models and studying their performance.

Map 1. Experimental fields of ISSNIP and meteorological stations of NIMH in Bulgaria.
Material and Methods

1. Ground Truth Data: Crop data are obtained during the so-called “balance” long-term experiment of Agro-ecology department of the “N. Poushkarov” Institute of Soil Science. The experiment was carried out with two typical corn varieties in eight representative agro-climatic regions of Bulgaria over the period 1975-1991 (Stoyanov, 2008). Observations were performed in fully irrigated plots at five levels of fertilization supply ranging from 0 to 125% of the optimum rate in fertilization treatments, named B1 (0), B2 (125%), B3 (100%), B4 (75%) and B5 (50%). The experiments provided crop data time series on annual grain yield relative to semi early A1 (Px-20 and P-37-37) and late A2 (H708) corn cultivars.

2. Satellite data: The satellite data represent solar radiation reflected or emitted from the land surface measured by the Advanced Very High Resolution Radiometer AVHRR. These data, named global vegetation index (GVI), are collected by NOAA. They are available from 1981 through present. The data set used in this study was developed by sampling the 4-km² global area coverage data and compositing from the daily afternoon observations to seven-day composite (Kidwell 1997). The Global Vegetation Indices (GVI) digital counts in the visible (VIS, 0.58–0.68 µm, Ch1), near-infrared (NIR, 0.72–1.1 µm, Ch2) and infrared (IR, 10.3–11.3 µm, Ch4) spectral regions are used. The VIS and NIR counts were converted into brightness (radiative) temperature (BT) reflected or emitted from the land surface measured by the NOAA AVHRR sensor. The smooth weekly NDVI or BT and their 1982–1991 absolute maximum (Amax) and absolute minimum (Amin) were used.

The vegetation condition index (VCI) characterizes greenness and vigour, and through them, the chlorophyll and moisture contents of the vegetation canopy. The temperature condition index (TCI) characterizes how hot the land surface and the conopy are. Moreover the TCI characterizes the moisture availability through the near-surface radiation and aerodynamic conditions (Jensen, 2000; Kogan et al., 2011). The vegetation health index (VHI) combines both VCI and TCI. VH indices change from 0, quantifying severe vegetation stress, to 100, quantifying favorable conditions. The average spatial values of VH indices for each week during 1982–1991 were calculated for the area of selected experimental fields at Gorni Dabnik (43°21’La; 24°21’Lg; 149 m Alt) and Sadievo (42°32’La; 26°03’ Lg; 154 m Alt), representing typical corn production conditions in North-West and South-East Bulgaria.

4. Methodology consists of: (a) choosing locations (experimental fields) representing important maize productive regions and real agricultural technologies (experimental treatments) that have produced a trend to the yield time series; (b) extracting the weather component from the values of the selected yield series and from the weekly NDVI and BT series and (c) to correlate the weather-related components of crop yield with NDVI and BT components. It is an adaptation of the methodology aiming at forecasting field crops production from satellite-based vegetation health indices (Kogan et al., 2003, 2005; 2011; 2015). However instead of the national statistic data, corn yield data series from long-term field experiments (1975-1991) are used (Stoyanov, 2008).

A relationship between the ground data and the satellite data, characterizing weather component has been searched. The data were expressed as a deviation from a standard: for yield (expressed as dY) – from the trend produced by agricultural technology B1 (unfertilized fully irrigated corn) on productivity of A1 (semi-early) and A2 (late) corn cultivars and for VH (VCI, TCI and VHI) – from normalized difference vegetation index NDVI and BT climatology. Both correlation and regression analysis of these deviations were performed to study the association of actual deviations of yield dY with VCI, TCI and VH indices. Thus the dY were correlated to each week’s VCI and TCI during 1982–1991 applying the “one-in one-out” technique (Jack Knife test) to investigate whether the deviation dY produced by agricultural technologies.
A1B1 and A2B1 correlate strongly with VH indices during a ‘critical’ period of strong corn response to changes in weather conditions. For Bulgarian lowlands such critical period normally covers “July-August” but it fluctuates according to regional climate characteristics that influence the date limiting corn flowering, yield formation and irrigation scheduling to manage droughts.

5. Combining ground observation data with Satellites data

Actual corn yield deviation from the trend dY was correlated with each week’s VCI, TCI and VHI during 1982-1991 to study how dY correlates with VH - indices during the period of strong crop response to weather conditions. Two types of dY models could be applied: (a) With the independent variables for the week with the highest Pearson correlation coefficient (eq.5) and (b) Several weeks indices with the Pearson correlation coefficient greater than 0.5. In this case the mean values for the selected weeks were used as independent variables (eq.6)

\[ dY = a_0 + b_1(VCI)_i + b_2(TCI)_j + b_3(VHI)_k + e \]  

where \( i, j \) and \( k \) is the week number for VCI, TCI and VHI, respectively; \( n, m \) and \( p \) is the number of weeks for which the mean VCI, the mean TCI and the mean VHI, respectively, are calculated; and \( e \) is the error.

The approach of cross-validation (‘leave-one-out’) is used. In this “Jack Knife Test” a single year was left out one by one from the data set, a model was built and prediction was made for the eliminated year (Kogan et al., 2015). As a result, 9 independent comparisons between the model predictions and ground observations were made.

To estimate the reliability of independent predictions, the corresponding verification model statistics were performed. Summary measures and difference measures test criteria have been applied: The first criteria includes the mean of the observed (\( O_i \)) and predicted (\( P_i \)) values, while the second criteria describes the quality of simulation by using the mean bias error (MBE, eq. (7)) and the root mean square error (RMSE, eq.(8)). They all are based on the term of \( (P_i - O_i) \) and calculated according to Willmott (1982):

\[ \text{(A) Mean bias error (MBE): } MBE = \frac{\sum(P_i - O_i)}{n}, \]  

\[ \text{(B) Root mean squared error (RMSE): } \text{RMSE} = \sqrt{\frac{\sum(P_i - O_i)^2}{n}}. \]  

The summation is done from case 1 (\( i = 1 \)) to case \( n (i = n) \).

Results and Discussions

1. Experimentally-based yield time series analyses:

As referred above, the ground truth data collected during the “balance” experiment (1975–1991, Stoyanov, 2008) were used after some graphical and statistical tests, as shown in figures 1, 2 and 3. Regarding time series graph, it could be concluded that fertilization technologies B2, B3 and B4 combined with a semi-early cultivar A1 (Px-20/P-3737), an appropriate irrigation and crop protection, have not practically produced any yield trends over the sixteen-year period. Contrarily, the agricultural technology A1B1 (same cultivar but unfertilised), produced a negative trend of yield decrease of -11 kg da⁻¹ yr⁻¹. These yields were approximated by equation (9) (Brockwell and Davis, 2000):

\[ Y_i = T_i + dY_i \]  

where \( T \) is a slowly changing function representing the deterministic component (trend) regulated by agricultural technology A1B1, \( dY \) is a random component regulated by weather fluctuations and \( i \) is the year or coded year number.

Fig. 2 compares the end-of-season yield time series 1975-1991 relative to two fully irrigated corn cultivars, a semi-early (A1B1, shown in a dashed line) versus a late one (A2B1, shown in a full line), grown without fertilization at four locations, representing the agro-climate potential for summer crops of southern and northern Bulgarian plains (Map). As it is seen in the figure, when comparing northern (Figs.2a and 2b) to southern selected locations (Figs. 2c and 2d), the role of corn cultivar was enhanced in the fertile Gorni Dabnik and Silvo Pole, the Danube plain, during severe droughts in the in the 80-ies (Slavov, Koleva, Alexandrov, 2004; Koleva and Alexandrov, 2008; Stoyanov, 2008; Popova et al., 2012; 2015).

A trend of the form \( T_i = a_0 + a_1 t_i \) was fitted to field data relative to the period 1982-1991, for which satellite – based data are available (Fig. 3).

Parameters \( a_0 \) (intercept) and \( a_1 \) (slope) have been derived by minimizing the differences \( \sum(Y_i - T_i)^2 \). Slopes were estimated for the fields of Sredetz (42°16’ La; 25°40’ Lg; 173 m Alt), Sadievo (42°32’ La; 26°03’ Lg; 154 m Alt) and Gorni Dabnici (43°21’ La; 24°21’ Lg; 149 m Alt), as -24.95, -14.95 and -16.85 kg da⁻¹ year⁻¹ for the technology A2 (a high demanding late cultivar H708) versus -7.21, 5.69 and 12.76 kg da⁻¹ year⁻¹ for the technology A1 (a semi-early cultivar P-3737) respectively (Figs. 3a 3b and 3c). Since the slopes are not
Early Forecasting Corn Yield Using Ground Truth Data and Vegetation Health Indices in Bulgaria

Sredec, Stara Zagora A1
Semi early corn hybrids all treatments

\[ y = -0.2157x + 1321.3 \]
\[ R^2 = 0.0001 \]

\[ y = -10.708x + 21550 \]
\[ R^2 = 0.4409 \]

Fig. 1. Trend analyses of annual corn yield time series relative to different fertilization technologies (B1 to B5) combined with semi-early cultivar technology (A1), 1975-1991.

2. Combining ground observation data with Satellites data

Figure 4 illustrates the dynamics of correlation coefficients for the actual yield deviation \( dY_i \) from the trend produced by two corn cultivar technologies consisting of: (a) a semi-early A1B1 and (b) a late A2B1 cultivar grown at Gorni Dabnik experimental field during 1982–1991 with each week’s VCI, TCI and VHI respectively. During mid July-September, when the corn flowering and yield formation is taking place (Allen large, the random component of the yield \( dY_i \) (eq.9) that is regulated by the weather conditions can be approximated by the difference \( dY_i = Y_i - T_i \) (Kogan et al., 2015).

Fig. 2. Comparing the annual crop yield time series relative to two corn cultivars A1 (a semi-early P-3737) and A2 (a late H708) combined with fertilization treatment B1 at the experimental fields of: a) Sredetz (42°16’La; 25°40’ Lg; 173 m Alt) and b) Sadievo (42°32’La; 26°03’ Lg; 154 m Alt), South Bulgaria, and c) Gorni Dabnic (43°21’La; 24°21’Lg; 149 m Alt) and d) Slivo pole (43°55’La; 24°21’ Lg; 25 m Alt), North Bulgaria, 1975-1991.
et al., 1998), correlations of \( dY \) with \( VCI, TCI \) and \( VHI \) show two picks of significant correlations.

For the semi-early cultivar technology A1B1 (Fig.4a) strong correlations of \( dY \) however are found earlier in May and June: with VHI (CC=0.60; Partial CC =0.89) that occurs during week 16 (May) and during week 21 (end of June) with TCI (CC=0.594; Partial CC =-0.036). The highest correlation coefficient with VCI is practically 0.5 during week 23 (CC=0.47; but Partial CC=-0.286) and below 0.5 during week 17(CC=-0.34; Partial CC=-0.9) (Fig. 4a).

Thus the Regression summary of the tests performed for technology A1B1 at Gorni Dabnic, Plevn region (Fig. 5) indicates the four calculated variables VCI17, VCI23, VHI16 and TCI21 (the week’s number given in subscript), the intercept \( a_0 \) and the four slope coefficients \( a_i \) of linear regression:

\[
dY_i=0.439973-0.010243 VCI_{17}+0.00134397 VCI_{23}+0.0223175 VHI_{16}+0.000153034 TCI_{21}\tag{10}
\]

The relationship between the actual (dY) and estimated (EdY) deviation from the trend is very strong with correlation coefficient CC=0.95 while the Yield Independent test results in CC=0.84 (Fig. 5).

For the corn late cultivar technology A2B1 (Figs. 4b and 6), differently to the semi-early cultivar technology A1B1 (Figs. 4a and 5), strong correlations of \( dY \) with VCI (CC=-0.53 and CC=0.57; Partial CC=-0.70 and Partial CC=0.73) occur latter during week 32 and during week 41 (August–September), and much earlier with TCI (CC=0.60; Partial CC=-0.475) in week 17 (May).

Regression summary of the tests carried out for the late corn cultivar technology A2B1 at Gorni Dabnic field shows the four calculated variables for the respective weeks VCI_{22}, VCI_{32}, VCI_{41} and TCI_{17}, the slope coefficients \( a_i \) and the intercept \( a_0 = 0.933 \) of linear regression eq.11:

\[
dY_i=0.932589 + 0.00477279 VCI_{22} -0.00333117 VCI_{32}+0.00704330 VCI_{41}-0.00272812 TCI_{17}\tag{11}
\]

The relationship between the actual (dY) and estimated (EdY) deviation from the trend of A2B1 is still strong with CC=0.87 while the Yield Independent test results in CC =0.63 (Fig.6) that is slightly lower than those of technology A1B1 (Fig. 5).

Figure 7 shows the correlation coefficient of \( dY_i \) for agricultural technology A2B1 (a late cultivar H708, unfertilized) with each week’s VCI, TCI and VHI computed for Sadievo experimental field, Burgas region, where the climate is influenced by the southern Black Sea.

For corn late cultivar (Figure 7), the correlation of \( dY_i \) with VCI (CC=0.33 and CC=0.41; Partial CC=-0.75 and Partial CC=0.77) occurs only during week 27 and week 28 (July), while with TCI (CC=0.52 and CC=0.47; Partial CC=0.52

Figure 3. Trends of the annual yield time series relative to corn cultivars A1 (semi-early P-3737) and A2 (late H708) relative to: a) Sredetz and b) Sadievo, Southeast Bulgaria and c) Gorni Dabnic experimental field, Northwest Bulgaria, 1982-1991.
and Partial CC=−0.39) correlation is found during week 5 and week 6 (January).

Regression summary of the tests carried out for the technology late corn cultivar A2B1 at Sadievo field, Burgas Region (Figure 8) let to four calculated variables for the respective weeks VCI\(_{27}\), VCI\(_{28}\), TCI\(_5\) and TCI\(_6\), the intercept \(a_0 = 1.045\) and the slope coefficients \(a_1\) of linear regression eq.10:

\[
dY = 1.04511 - 0.01926 VCI_{27} + 0.020634 VCI_{28} + 0.0064522 TCI_{05} - 0.0043098 TCI_{06}
\]  

(12)

Fig. 4. Dynamics of the Pearson Correlation coefficient between the actual deviations of yield \(dY\) relative to agricultural technology: a) A1B1 (a semi-early cultivar P-3737) and b) A2B1 (a late cultivar H708) and the vegetation health indexes VCI, TCI and VHI, unfertilized corn, Gorni Dabnik, Pleven region.

Fig. 5. Regression summary and graphs of the tests performed for agricultural technology A1B1 (a semi-early cultivar, P-3737, unfertilized corn), Gorni Dabnic, Pleven region, 1982-1991.
Figure 6. Regression summary of the tests performed for agricultural technology A2B1 (a late cultivar H708, unfertilized corn), Gorni Dabnic, Pleven, 1982-1991.

Fig. 8. Regression summary of the tests performed for agricultural technology A2B1 (a late cultivar), Sadievo, Burgas region, 1982-1991.
Fig. 7. Dynamics of the Pearson Correlation coefficient between the actual deviation of yield $dY$ and the vegetation health indexes VCI, TCI and VHI relative to agricultural technology A2B1 (a late cultivar H708, unfertilized), Sadi-evo, Burgas Region, 1982-1991.

Conclusions

In this paper, three satellite-based globally universal VH indices characterising vegetation greenness and vigour (VCI), moisture and thermal conditions (TCI) and vegetation health (VHI) were used as yield predictors of two corn cultivars (a semi-early and a late one) in the experimental fields of Gorni Dabnik, North-West Bulgaria, and Sadievo, South-East Bulgaria. The regions were Pleven and Burgas respectively and the first one is a major grain producer in this country. Previously this technique was applied and showed good results to model different crops (wheat, corn, sorghum, rice, etc.) in USA, Russia, Kasahstan, China and other countries. In this case study, the VH proxies was limited to the case of statistical modelling of crop yield relative to unfertilised corn that used to be a common agricultural technology during the transition from a state-planned to a market economy in this country. The study has shown very good results of dependent validation test (CC of 87 and 95%) and good results of independent validation (CC of 63 - 84%). The developed models were quite accurate and reliable in prediction of corn grain yield before official statistics of grain harvest is released. From the three indices characterizing moisture (VCI), thermal (TCI) and vegetation health (VHI) conditions, the first and the third were the best in the study but all three were good predictors of corn yield. The article also showed that there is potential for VH application in modelling corn yield in a larger regional and country scale. Further investigation of yield losses predictors might include combining satellite data with national statistics harvested maize yield data. The VH indices and data are delivered every week to http://www.star.nesdis.noaa.gov/smcVH/index.php.

Acknowledgements

Grateful acknowledgement to Fulbright Exchange Scholar Program (CIES) for awarded Visiting Scholar Grant on “Economical and physical aspects of climate change impacts on agriculture – Assessing vulnerability to drought and irrigation demands” to Prof. Zornitsa Popova and to Prof. Ramesh Singh for hosting and providing facilities at School of Earth and Environmental Sciences, Chapman University, Orange, Southern California. Thanks to Dr. Felix Kogan, NOAA, Camp
Springs, USA for the nice cooperation.

References


