Multicollinearity Analysis of Phenology Metrics From Modis Satellite Imagery for Maize and Soybean in Croatia

Analiza multikolinearnosti fenoloških metrika iz modis satelitskih snimki za kukuruz i soju u Hrvatskoj

Radočaj, D., Jurišić, M.

Poljoprivreda / Agriculture

ISSN: 1848-8080 (Online)
ISSN: 1330-7142 (Print)

https://doi.org/10.18047/poljo.31.2.9

Fakultet agrobiotehničkih znanosti Osijek, Poljoprivredni institut Osijek

Faculty of Agrobiotechnical Sciences Osijek, Agricultural Institute Osijek

ISSN 1330-7142
UDK=005.642.4+621.397.444:633.15'34
https://doi.org/10.18047/poljo.31.2.9

MULTICOLLINEARITY ANALYSIS OF PHENOLOGY METRICS FROM MODIS SATELLITE IMAGERY FOR MAIZE AND SOYBEAN IN CROATIA

Radočaj, D., Jurišić, M.

Original scientific paper
Izvorni znanstveni članak

SUMMARY

This study analyzed the phenology metrics of maize and soybean in Croatia across the years 2021 and 2022 based on Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data, focusing on median dates of occurrence and their variability. The total number of filtered parcels from the Paying Agency for Agriculture, Fisheries and Rural Development database with area larger than 25 ha was 108 and 463 for maize, as well as 25 and 163 parcels for soybean in the years 2021 and 2022, respectively. Median maturity and peak growth stages occurred later in 2021 than in 2022 for both crops, suggesting potentially broader climatic impacts on growth duration. The correlation matrices for the study indicated strong relationships between maturity, senescence, and peak phases but not dormancy which had low levels of correlation with other measures. The presence of high and significant multicollinearity was underlined in the study by the analysis of variance inflation factors used when analyzing phenological parameters, especially with limited sample sizes. These findings suggest that dormancy and peak phenology can be useful for a characterization of crop yield and cropland conditions.

Keywords: MODIS, enhanced vegetation index, peak, dormancy, variance inflation factor

INTRODUCTION

The study of periodic biological events in agriculture is known as phenology and is central to understanding how our crop plants respond to changing environments. Such technologies as remote sensing, especially based on satellite imagery (such as Moderate Resolution Imaging Spectroradiometer, MODIS) have made it possible to monitor vegetation dynamics of a large spatial scale over prolonged time frames (Frappart et al., 2020). Phenological metrics from remote sensing can be used as crop yield predictors (Ji et al., 2021), as the results from a previous study suggested that there is significant correlation between the phenological stages with maize yield, and thus the development of predictive model for the prediction of maize yield before harvest is of advantage to decision making processes (Ji et al., 2021). Those remote sensing derived phenology metrics were used to monitor different farming systems and practices (Radočaj et al., 2020). A previous study analyzed changes in vegetation cover over time, analyzing its impact in response to agricultural interventions and adapt strategies particularly in places where ground data collection is expensive or hard to collect (Lebrini et al., 2021).

Multicollinearity, or more than one predictor in a regression model being highly correlated, can make it difficult to identify the real relationships of variables in the presence of remote sensing data for the purposes of phenology analyses (Ullah et al., 2019). Remote sensing derived phenology metrics are not only essential for crop growth and yield prediction but also an important consideration in cropland suitability studies (Radočaj & Jurišić, 2022). Thus, a need arises for a single metric to represent all suitability values on a parcel level in areas with large crop yield data, which is rarely available on a larger area. The difficulty is due to the possible multicollinearity amongst several phenological metrics and could obscure the relationship

Dorijan Radočaj, PhD (dradocaj@fazos.hr), Prof. Mladen Jurišić – Josip Juraj Strossmayer University of Osijek, Faculty of Agrobiotechnical Sciences Osijek, Vladimira Preloga 1, 31000 Osijek, Croatia

of each indicator to the joint contribution to land suitability. An analysis of each pair of phenological metrics is conducted to identify the most relevant metric, based on multicollinearity, which is also accurately reflective of the suitability of a particular agricultural area for crop production. Therefore, the goal of this study was to identify and reduce the potential of confounding effects due to correlated metrics, using a multicollinearity analysis of phenology metrics from MODIS imagery on maize and soybean in Croatia for two study years.

MATERIALS AND METHODS

The study area included maize and soybean parcels from the Paying Agency for Agriculture, Fisheries and Rural Development database with area larger than 25 ha for years 2021 and 2022 (fig. 1).

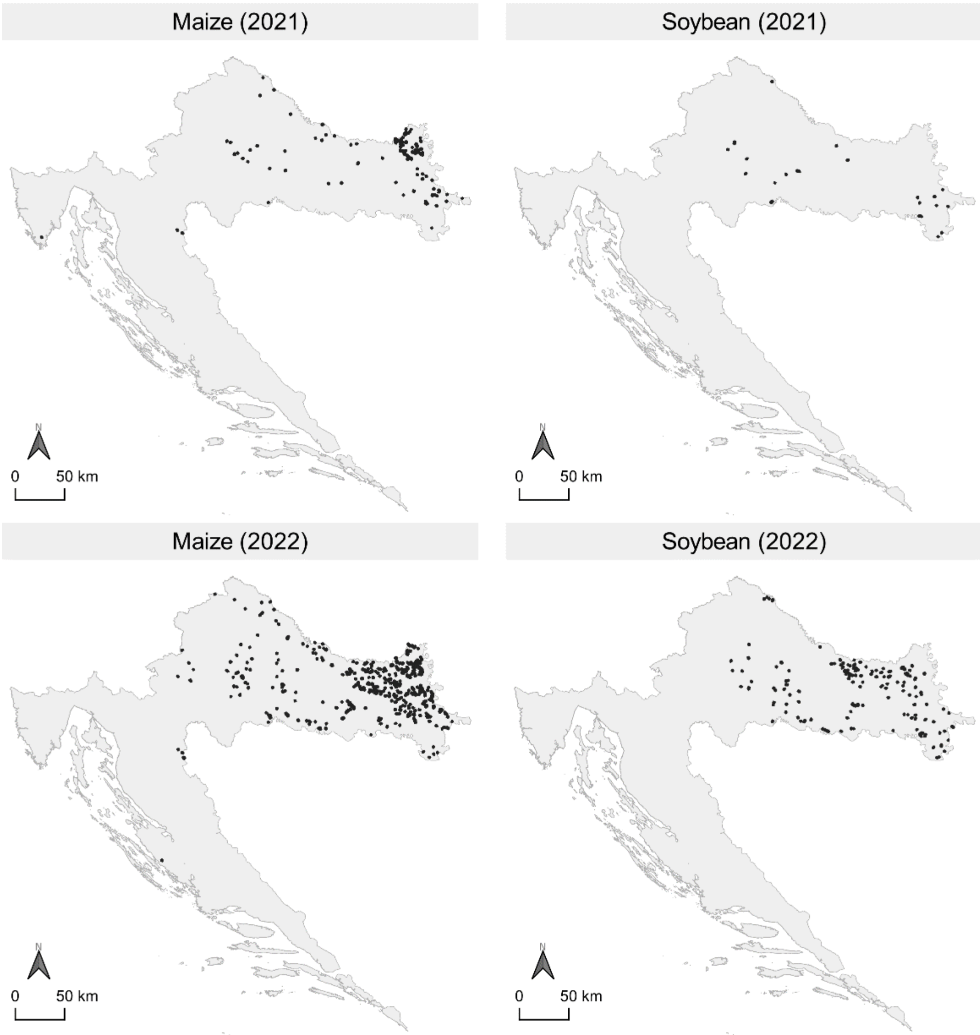

Figure 1. Study area with the representation of maize and soybean parcels in years 2021 and 2022.
Slika 1. Područje istraživanja prikazom referentnih čestica kukuruza i soje u 2021. i 2022. godini

The area threshold was selected to include parcels which could cover area of pixel from MODIS phenology data or larger. The total number of filtered parcels was 108 and 463 for maize, as well as 25 and 163 parcels for soybean in the years 2021 and 2022, respectively.

The study utilized the MCD12Q2.006 Land Cover Dynamics Yearly Global data product from the Moderate Resolution Imaging Spectroradiometer (MODIS) to analyze vegetation phenology metrics including greenup, mid-greenup, maturity, peak greenness, senescence, mid-greendown, and dormancy (Table 1). This dataset provides global land surface phenology metrics at a spatial resolution of 500 m and covers the time period since 2001. The data is derived from time series of the two-band Enhanced Vegetation Index (EVI2), which was calculated from the MODIS Nadir Bidirectional Reflectance Distribution Function-Adjusted Reflectance. The analysis was conducted using *Google Earth Engine*, a cloud-based platform that allows for efficient processing of large geospatial datasets (Salinero-Delgado et al., 2022).

Table 1. Definition of evaluated phenology metrics from MCD12Q2.006 data (Friedl et al., 2022)

Tablica 1. Definicija evaluiranih fenoloških metrika iz MCD12Q2.006 podataka (Friedl et al., 2022)

Phenology metrics *Fenološke metrike*	**Definition** *Definicija*
Greenup	Date when EVI2 first crossed 15% of its amplitude *Datum kada je EVI2 prvi put prešao 15 % svoje amplitude*
Mid-Greenup	Date when EVI2 first crossed 50% of its amplitude *Datum kada je EVI2 prvi put prešao 50 % svoje amplitude*
Maturity	Date when EVI2 first crossed 90% of its amplitude *Datum kada je EVI2 prvi put prešao 90 % svoje amplitude*
Peak	Date when EVI2 reached the maximum *Datum kada je EVI2 dosegnuo maksimum*
Senescence	Date when EVI2 last crossed 90% of its amplitude *Datum kada je EVI2 posljednji put prešao 90 % svoje amplitude*
Mid-Greendown	Date when EVI2 last crossed 50% of its amplitude *Datum kada je EVI2 posljednji put prešao 50 % svoje amplitude*
Dormancy	Date when EVI2 last crossed 15% of its amplitude *Datum kada je EVI2 posljednji put prešao 15 % svoje amplitude*

The multicollinearity was conducted by correlation matrices using the Spearman correlation coefficient, and the Variance Inflation Factor (VIF). It was done by assessing greenup, mid-greenup, maturity, senescence, mid-greendown and dormancy based on peak, which is the most common phenology metric used in previous studies, since it has presence of greater than 39 percent in vegetation phenology studies in 2024. The Spearman correlation coefficient was used in identifying the strength and direction of relationship between two ranked variables and without the assumption of proportionality in the two variables, being appropriate for monotonic relationships. A correlation matrix derived from these coefficients measured the relationships between peak and other considered phenology metrics, with larger absolute correlation coefficients indicating the possibility of multicollinearity. Further, the VIF values were computed for each independent variable of the regression model for which peak was set as the dependent variable. VIF measured the degree to which the variance of a regression coefficient increased because of multicollinearity with other predictors, VIF of 1 was taken to imply no correlation while values greater than 5 was considered critical levels of multicollinearity that may pose serious threat to the reliability of the regression estimates.

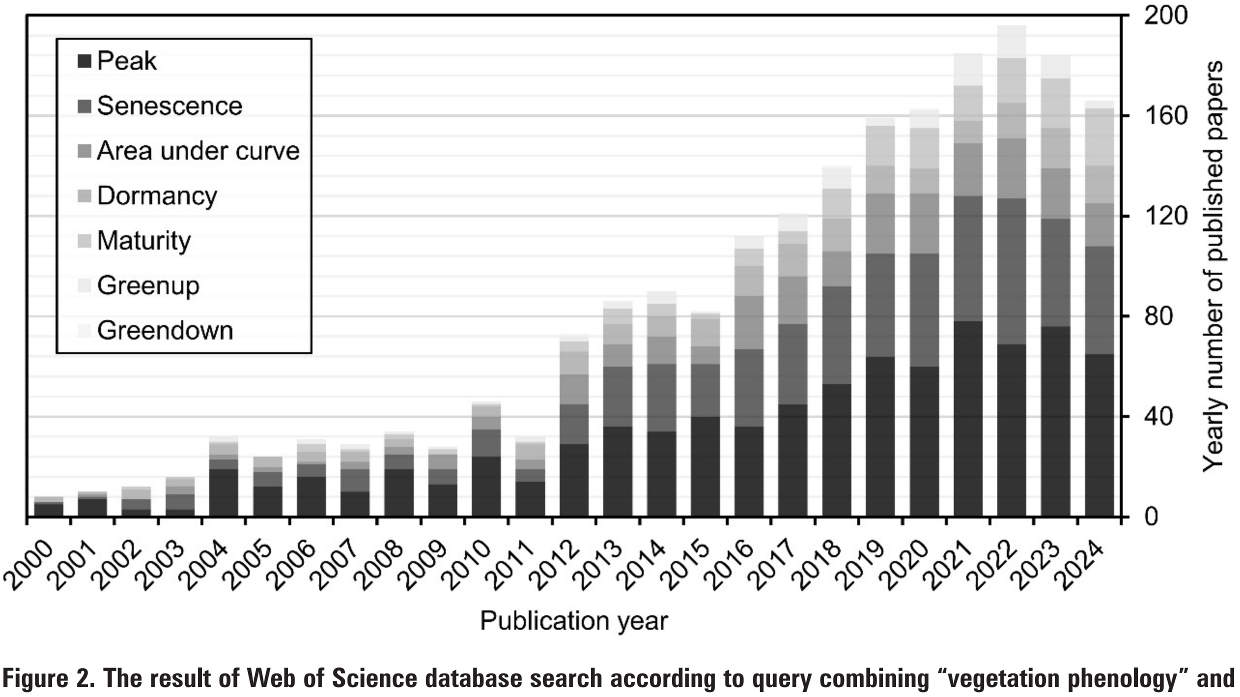

Figure 2. The result of Web of Science database search according to query combining "vegetation phenology" and phenology metrics.

Grafikon 2. Rezultat pretraživanja baze Web of Science prema upitu koji kombinira "vegetation phenology" i fenološke metrike

RESULTS AND DISCUSSION

Table 2. summarized phenology metrics for maize and soybean across the years 2021 and 2022, with median dates of occurrence (expressed in Day of Year, DOY) and their corresponding coefficients of variation (CV). Greenup, the initial growth phase, occurred relatively consistently for maize in 2021 and 2022, while soybean had slightly later median greenup in 2021 than in 2022. The soybean's mid greenup was slightly earlier in 2021 and then in 2022, suggesting variability in growth conditions, which was also confirmed by median maturity values. The median peak phenology stage for both maize and soybean occurred slightly later in 2021 than in 2022, with 21 and 25 days later for maize and soybean, respectively. This observation was further supported with the relationship between median senescence, mid-greendown and dormancy values for both crops in 2021 and 2022. The coefficients of variation across these metrics suggest varying levels of consistency within each crop's phenological responses, with lower CV values indicating more stable phenological timing, which was generally more present in the latter vegetation growth stages, as represented by senescence, mid-greendown and dormancy.

Table 2. Descriptive statistics of evaluated phenology metrics for maize and soybean ground truth parcels in the years 2021 and 2022

Tablica 2. Deskriptivna statistika promatranih fenoloških metrika za poljoprivredne čestice kukuruza i soje u 2021. i 2022. godini

Phenology metrics *Fenološke metrike*	**Maize (2021)** *Kukuruz (2021)*		**Maize (2022)** *Kukuruz (2022)*		**Soybean (2021)** *Soja (2021)*		**Soybean (2022)** *Soja (2022)*	
	Median (DOY)	CV	Median (DOY)	CV	Median (DOY)	CV	Median (DOY)	CV
Greenup	92	0.314	94	0.284	98	0.234	92	0.253
Mid-Greenup	135	0.172	133	0.161	143	0.159	131	0.144
Maturity	179	0.117	164	0.100	186	0.118	165	0.100
Peak	202	0.097	181	0.084	209	0.100	184	0.087
Senescence	225	0.081	200	0.077	230	0.074	205	0.076
Mid-Greendown	254	0.068	228	0.084	259	0.034	235	0.073
Dormancy	283	0.070	267	0.112	291	0.042	276	0.083

The phenology metrics in the current study for both maize and soybean in 2021 and 2022 are comparable to studies done earlier showing that phenology of crops are highly dependent on environmental factors. This constancy in times of greenup for maize in both years indicated to a predictable response to climatic factors. Thus, the study supports He et al. (2020), who observed that phenophases are important for the management of crops because they may differ from one year to another due to climate changes. The small variation observed in mid-greenup and maturity dates for soybean suggest vulnerability of the growth conditions to changes in weather patterns, a phenomenon captured by previous studies about the effects of temperature/moisture availability on crops as they develop (Liu & Dai, 2020). Therefore, the later peak phenology identified in both crops during 2022 might be an example of wider climactic changes that affect the growth duration, as shown by other studies suggesting that changes in phenological timing may impact yield (Fatima et al., 2020). In addition, the CV values indicated that while the initial growth phases appear to be most heterogeneous, while senescence and dormancy phases seem to be most homogeneous. This agrees with Menzel et al (2006), as it was established that latter phenophases are not influenced as much by environmental conditions as the former, being more suitable for evaluation of crop status and yield.

Figure 3 presents correlation matrices for all evaluated phenology metrics for maize and soybean in 2021 and 2022. In all four analyzed datasets, maturity and senescence produced very high Spearman's correlation coefficient with peak, while dormancy was the only one which consistently had low correlation with peak, as well as with all other metrics except dormancy. Therefore, peak and dormancy should be evaluated according to their correlation with crop yield, representing likely candidates for proxy metric to cropland suitability assessment. Moreover, mid-greendown had relatively low correlation with both greenup and mid-greenup, suggesting varying duration of the overall vegetation period for both maize and soybean in both study years. These observations were confirmed with VIF values displayed in Figure 4, in which dormancy was the only phenology metric which did not result in multicollinearity with peak in all four datasets. Additionally, greenup for both maize and soybean in 2021 produced VIF lower than 5, as well as mid-greendown for soybean in 2021, but this particular observation might be biased due to low sample count for soybean in 2021.

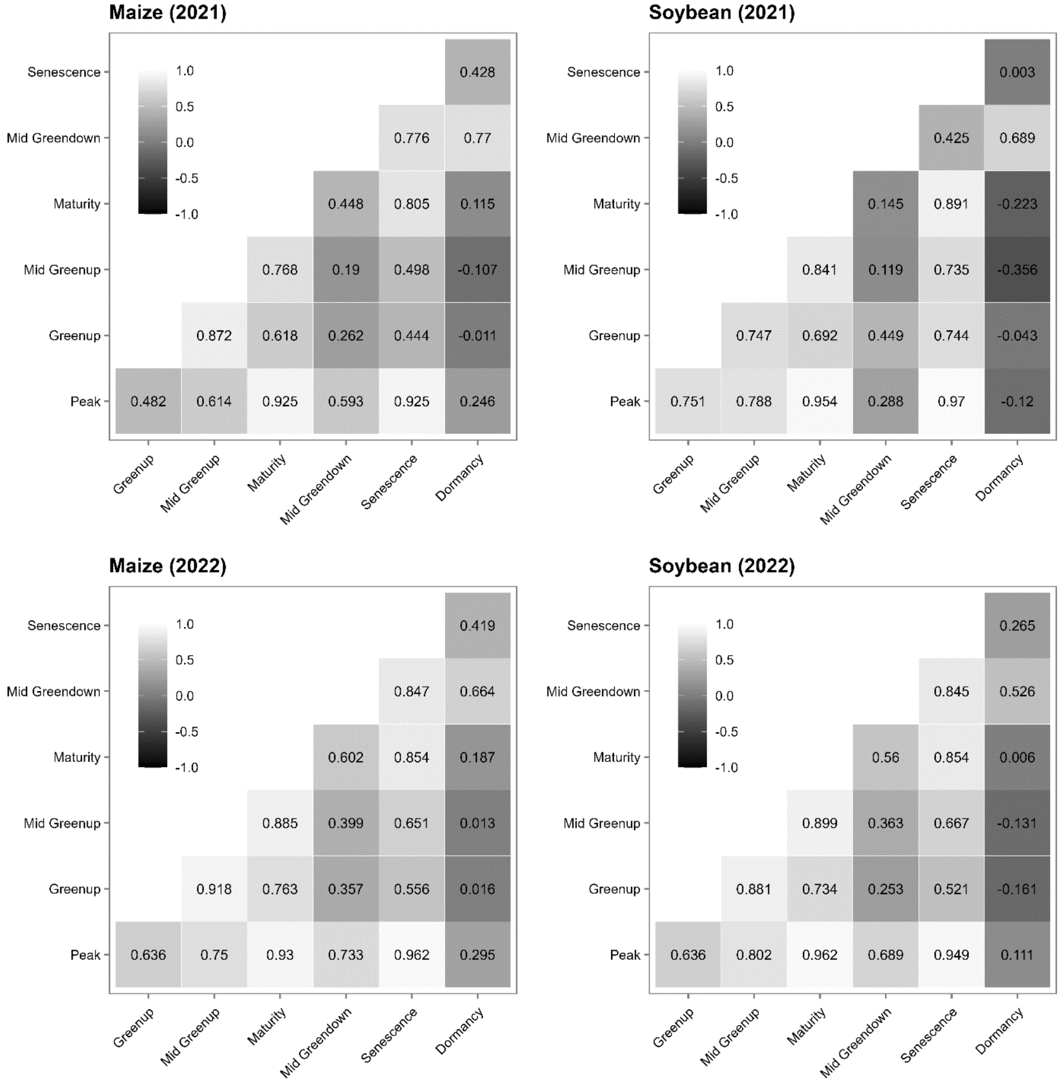

Figure 3. Correlation plots of vegetation phenology metrics according to Spearman's correlation coefficient

Grafikon 3. Korelacijski dijagrami fenoloških metrika vegetacije prema Spearmanovu koeficijentu korelacije.

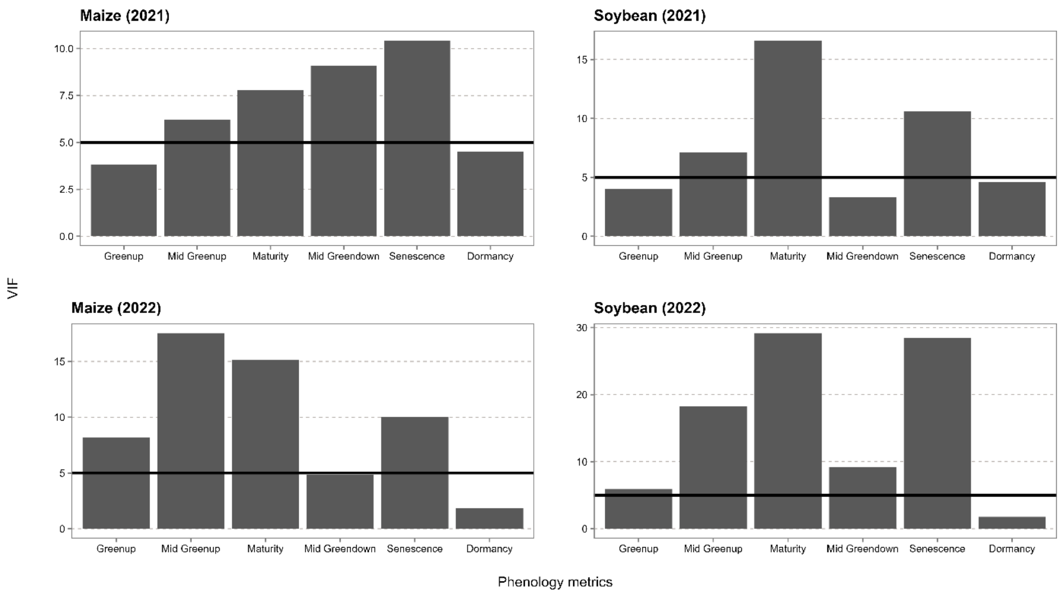

Figure 4. VIF values representing multicollinearity of evaluated vegetation phenology metrics according to vegetation peak.

Grafikon 4. VIF vrijednosti koje predstavljaju multikolinearnost korištenih fenoloških metrika vegetacije u odnosu na vrhunac vegetacije.

The results show the matrices of correlation coefficients between phenological indicators for both crops, which have strengthened the conclusions reached in other research works pointing to interconnection of phenological events. The high Spearman's correlation coefficients between maturity, senescence and peak gives the evidence that these stages are related in accordance with research done by Magney et al. (2016), establishing the late season phenological events to be critical in yield determination. On the other hand, the low and less significant correlation of dormancy with peak and other metrics agrees with the study by Beil et al. (2021), which indicate that dormancy is less sensitive to the preceding growth phases and more of a stable measure of prevailing abiotic status. The VIF values below 2.0 for greenup and mid-greendown in 2021 indicated the need to consider sample size in statistical analyses according to Park et al. (2021), as small samples may provide unbiased observations on phenological data.

CONCLUSIONS

A study of the phenology metrics for 2021 and 2022 for maize and soybean based on MODIS satellite data is provided. The main observations from this study include that maize had consistent greenup timing, whereas soybean had slightly different phenological phases across years, probably due to changes in environmental conditions. Notably, later median dates for maturity and peak growth than in 2022 were observed for both crops in 2021, potentially indicating effects of broader climatic trends on growth duration. The results presented here indicate that the CV values differed in the extent to which they showed variation in phenological responses and that this agrees with literature in that the later stages of senescence and dormancy showed greater consistency than earlier stages. The multicollinearity analysis also showed strongly interconnected phenological events, especially between age at maturity, senescence, and peak stages. In contrast, dormancy has a low correlation with other metrics, suggesting that it reflected more stable environmental responses. This indicates that dormancy and peak phenology may be viable metrics for evaluation of crop yield and the cropland suitability. In addition, VIF values were analyzed for the effect of multicollinearity on phenological metrics and the need to take into account multicollinearity when evaluating phenological metrics was highlighted, particularly in studies with small sample sizes. Future studies should increase samples sizes and explore other environmental variables that may aid in phenological development of these key crops.

REFERENCES

1. Beil, I., Kreyling, J., Meyer, C., Lemcke, N., & Malyshev, A. V. (2021). Late to bed, late to rise—Warmer autumn temperatures delay spring phenology by delaying dormancy. *Global Change Biology*, *27*(22), 5806–5817. https://doi.org/10.1111/gcb.15858
2. Fatima, Z., Ahmed, M., Hussain, M., Abbas, G., Ul-Allah, S., Ahmad, S., Ahmed, N., Ali, M. A., Sarwar, G., Haque, E. ul, Iqbal, P., & Hussain, S. (2020). The fingerprints of climate warming on cereal crops phenology and adaptation options. *Scientific Reports*, *10*(1), 18013. https://doi.org/10.1038/s41598-020-74740-3
3. Frappart, F., Wigneron, J.-P., Li, X., Liu, X., Al-Yaari, A., Fan, L., Wang, M., Moisy, C., Le Masson, E., Aoulad Lafkih, Z., Vallé, C., Ygorra, B., & Baghdadi, N. (2020). Global Monitoring of the Vegetation Dynamics from the Vegetation Optical Depth (VOD): A Review. *Remote Sensing*, *12*(18), Article 18. https://doi.org/10.3390/rs12182915
4. Friedl, M., Gray, J., & Sulla-Menashe, D. (2022). *MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V061* [Dataset]. NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MCD12Q2.061
5. He, L., Jin, N., & Yu, Q. (2020). Impacts of climate change and crop management practices on soybean phenology changes in China. *Science of The Total Environment*, *707*, 135638. https://doi.org/10.1016/j.scitotenv.2019.135638
6. Ji, Z., Pan, Y., Zhu, X., Wang, J., & Li, Q. (2021). Prediction of Crop Yield Using Phenological Information Extracted from Remote Sensing Vegetation Index. *Sensors*, *21*(4), Article 4. https://doi.org/10.3390/s21041406
7. Lebrini, Y., Boudhar, A., Laamrani, A., Htitiou, A., Lionboui, H., Salhi, A., Chehbouni, A., & Benabdelouahab, T. (2021). Mapping and Characterization of Phenological Changes over Various Farming Systems in an Arid and Semi-Arid Region Using Multitemporal Moderate Spatial Resolution Data. *Remote Sensing*, *13*(4), Article 4. https://doi.org/10.3390/rs13040578
8. Liu, Y., & Dai, L. (2020). Modelling the impacts of climate change and crop management measures on soybean phenology in China. *Journal of Cleaner Production*, *262*, 121271. https://doi.org/10.1016/j.jclepro.2020.121271
9. Magney, T. S., Eitel, J. U. H., Huggins, D. R., & Vierling, L. A. (2016). Proximal NDVI derived phenology improves in-season predictions of wheat quantity and quality. *Agricultural and Forest Meteorology*, *217*, 46–60. https://doi.org/10.1016/j.agrformet.2015.11.009
10. Menzel, A., Sparks, T. H., Estrella, N., & Roy, D. B. (2006). Altered geographic and temporal variability in phenology in response to climate change. *Global Ecology and Biogeography*, *15*(5), 498–504. https://doi.org/10.1111/j.1466-822X.2006.00247.x
11. Park, D. S., Newman, E. A., & Breckheimer, I. K. (2021). Scale gaps in landscape phenology: Challenges and opportunities. *Trends in Ecology & Evolution*, *36*(8), 709–721. https://doi.org/10.1016/j.tree.2021.04.008
12. Radočaj, D., & Jurišić, M. (2022). GIS-Based Cropland Suitability Prediction Using Machine Learning: A Novel Approach to Sustainable Agricultural Production. *Agronomy*, *12*(9), Article 9. https://doi.org/10.3390/agronomy12092210
13. Radočaj, D., Jurišić, M., Gašparović, M., & Plaščak, I. (2020). Optimal Soybean (*Glycine max* L.) Land Suitability Using GIS-Based Multicriteria Analysis and Sentinel-2

Multitemporal Images. *Remote Sensing*, *12*(9), Article 9. https://doi.org/10.3390/rs12091463

14. Salinero-Delgado, M., Estévez, J., Pipia, L., Belda, S., Berger, K., Paredes Gómez, V., & Verrelst, J. (2022). Monitoring Cropland Phenology on Google Earth Engine Using Gaussian Process Regression. *Remote Sensing*, *14*(1), Article 1. https://doi.org/10.3390/rs14010146
15. Ullah, M. I., Aslam, M., Altaf, S., & Ahmed, M. (2019). Some new diagnostics of multicollinearity in linear regression model. *Sains Malaysiana*, *48*(9), 2051–2060.

ANALIZA MULTIKOLINEARNOSTI FENOLOŠKIH METRIKA IZ SATELITSKIH SNIMAKA MODIS-A ZA KUKURUZ I SOJU U HRVATSKOJ

SAŽETAK

Ova studija analizirala je fenološke metrike kukuruza i soje u Hrvatskoj tijekom 2021. i 2022. godine na temelju satelitskih podataka spektroradiometra umjerene rezolucije (MODIS), usredotočujući se na srednje datume pojavljivanja i njihovu varijabilnost. Ukupan broj filtriranih čestica iz baze Agencije za plaćanja u poljoprivredi, ribarstvu i ruralnome razvoju površine veće od 25 ha bio je 108 i 463 za kukuruz te 25 i 163 parcele za soju u 2021. i 2022. godini. Srednje faze zrelosti i vršne faze rasta detektirane su 2021. kasnije negoli 2022. za oba usjeva, što ukazuje na potencijalno šire klimatske utjecaje na trajanje rasta. Koeficijenti varijacije otkrili su veću stabilnost u kasnijim fenološkim fazama, kao što su starenje i mirovanje, što je u skladu s postojećom literaturom o njihovoj otpornosti na varijabilnost okolišnih uvjeta. Korelacijske matrice pokazale su jake odnose između zrelosti, starenja i vršnih faza, dok je mirovanje pokazalo niske korelacije s drugim metrikama. Analiza faktora inflacije varijance (VIF) istaknula je važnost analize multikolinearnosti pri procjeni fenoloških metrika, posebno u studijama s ograničenim veličinama uzorka. Ova opažanja sugeriraju da mirovanje i vršna fenologija mogu poslužiti kao vrijedni pokazatelji za procjenu prinosa usjeva i studija o pogodnosti zemljišta za usjeve.

Ključne riječi: MODIS, pojačani vegetacijski indeks, vrh, dormantnost, faktor inflacije varijance

(*Received on January 28, 2025; accepted on September 11, 2025* – Primljeno 28. siječnja 2025.; prihvaćeno 11. rujna 2025.)