95. Canopy anomaly classification using hybrid machine learning

B. Maestrini[1,*], J.M. Michielsen[1], S. Boersma[1], S. Wan[2], A. Beniers[1], G. Kessel[1], F. Hollewand[1], M. Shamsi[1] and M. Afonso[1]

1 Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen The Netherlands;
** Bernardo.maestrini@wur.nl*
2 Heilongjiang Academy of Agricultural Sciences, Harbin, Nangang, P.R. China

Abstract

An hybrid machine learning model that detects anomalies in canopy reflectance and identifies their cause in potato crops was developed. The model was created by training a long-short-term-memory neural network on synthetic crop NDVI data generated by coupling a potato growth model (Tipstar) and a canopy reflectance model (PROSAIL). The following stress factors were considered: no stress (control), N deficiency, presence of weeds, poor potato emergence. The model was validated against field experimental data – collected for two years – where the different stresses were imposed. The balanced model accuracy on the experimental dataset ranged between 50 and 100%, depending on the year and the stress factor.

Keywords: canopy anomaly classification, hybrid machine learning, potatoes, PROSAIL

Introduction

Anomaly detection is defined as "the problem of finding patterns in data that do not conform to expected behavior […] The importance of anomaly detection is due to the fact that anomalies in data translate to significant (and often critical) actionable information in a wide variety of application domains" (Chandola *et al.*, 2009). In the domain of open field crops management, detecting anomalies practically translates into detecting anomalies in the canopy, since canopy information can be automatically retrieved through several sensors (satellites, tractor mounted cameras, hand-held sensors). Detecting anomalies (i.e. deviation from potential growth) is a relevant topic since they are caused by stress factors, thus identifying and classifying stresses represent the first step in the smart crop management (measure, act, monitor) cycle.
While most anomaly classification are carried out at leaf scale, this study focuses on classification at canopy scale. An example of anomaly classification at canopy scale is the model by Khanna *et al.* (2019) that takes as input spatiotemporal spectral data (i.e. time series of hyperspectral images), extracts plant traits such as height and canopy reflectance and classifies the following stress factors: drought, nitrogen (N) deficiency and weed.
The aim of this study is to develop a model that takes as input a time series of multispectral measurements of canopy, e.g. from satellite, and classifies the time series in one of four possible causes of anomaly: no anomaly (control), N stress, low emergence, presence of weeds. The model makes use of time series of Normalized Differential Vegetation Index (NDVI), since this is a well-known and robust vegetation index.
The expectations about the shape of vegetation index time series prior experimental and modelling work were the following: (1) weeds increase NDVI throughout the season, except when the NDVI is saturated in the middle of the season, (2) lower tuber emergence would decrease NDVI at the beginning of the season and (3) N stress would decrease NDVI in the middle of the season, as base soil and tuber N supply is usually sufficient to cover the very early stages of the crop growth.

Materials and methods

The model was built using an hybrid machine learning approach where process-based models (crop growth models in this case) and machine learning are integrated (Maestrini *et al.*, 2022). The workflow was the following: a synthetic dataset of time series of crop canopy NDVI was generated by imposing the stresses on simulation scenarios produced using the crop growth model Tipstar (van Oort *et al.*, 2024) coupled to the radiative transfer model PROSAIL (Jacquemoud *et al.*, 2009). A long-short term memory neural network (LSTM) was trained using the synthetic dataset, its input is the NDVI time series of variable length and the output an anomaly cause class (no stress, weed, N stress, low emergence). The model was tested against observations from an ad-hoc two years field experiment.

Synthetic dataset

The scenarios used for the simulations of the synthetic dataset were imposed in the Tipstar-PROSAIL coupled models. Tipstar (van Oort *et al.*, 2024) is a dynamic potato growth model with daily time steps that includes N and water limitations and simulates water movement in the soil using a tipping bucket approach. Leaves development and senescence is simulated using a leaf cohort approach, where leaves die as they reach the thermal time set by leaf age max parameter and photosynthesis is modelled using a radiation use efficiency coefficient that is a function of daily PAR and temperature (Rodriguez *et al.*, 1999). Tipstar takes, as input, time series of weather, soil physical and chemical parameters, management, e.g. N fertilization and irrigation events, emergence date and density, and cultivar earliness, and returns time series of crop state variables such as yield, LAI, soil moisture. The weather data used for the simulations were obtained from the Weather Company (https://www.weathercompany.com/) and the soil physics characteristics (van Genuchten parameters; van Genuchten, 1980) were derived from Bofek 2020 Dutch soil maps (Heinen *et al.*, 2022). PROSAIL is a radiative transfer model that takes as input both leaf level parameters (e.g. chlorophyll content, water thickness) and canopy level parameters (leaf area index, leaf angle distribution). The R implementation of PROSAIL was used (version 2.2.2; Feret and de Boissieu, 2023). PROSAIL is a model resulting from the coupling of a leaf optical properties model and a canopy radiative transfer. In this study PROSAIL by coupling the models PROSPECT-PRO and 4SAIL2 was used. Some of the PROSAIL parameters were made a function of the output of Tipstar. The parameters that were varied were: leaf area index (output of Tipstar), ground coverage, calculated as a function of LAI following Hui *et al.*, (2013), solar zenith angle (calculated under the assumption that observations were taken at 11 am local time), leaf chlorophyll content, calculated as a function of leaf N content (an output of Tipstar), following Vos and Bom (1993), brown fraction, calculated as the ratio between dead LAI and alive LAI (both output of Tipstar). The remaining parameters, were set following Roosjen *et al.* (2018) who used PROSAIL to simulate the reflectance of a potato crop canopy in the Netherlands, whereas for the characteristics of brown leaves the parameters provided in PROSAIL examples of R implementation were used (Feret & de Boissieu, 2023) and bare soil reflectance was set to the default wet soil reflectance spectra. The hemispherical-directional reflectance factor in viewing direction simulation (rdot) was used.
A factorial experiment with 8 sites×44 years (1980–2023)×4 treatments×2 varieties, for a total of 2800 runs was simulated. The 8 Dutch sites correspond to the experimental farms of Wageningen University and Research, and the simulated varieties were a late and an early maturity variety, with earliness score 3 and 8 respectively. The earliness score refers to the score used to describe the phenology of Dutch potato varieties (Tiemens-Hulscher *et al.*, 2013),which is a measure of canopy cover duration. Model inputs were varied using inputs described in Table 1.

Table 1. Model input variability by treatment.

	N fert (kg N/ha)	No. of irrig events	Amount irrigation/event (mm)	Planting density (plants/ha)
Control	U (min=170, max=300)	Pois (λ=6)	logN (μ=5, σ=1.5)	U (min=40000, max=50000)
N stress	U (min=0, max=130)	Pois (λ=6)	logN (μ=5, σ=1.5)	U (min=40000, max=50000)
Rainfed	U (min=170, max=300)	0	0	U (min=40000, max=50000)
Low emergence	U (min=170, max=300)	Pois (λ=6)	logN (μ=5,σ=1.5)	U (min=40000, max=50000) N (μ=0.5, SD=0.1)
Weeds	U (min=170, max=300)	Pois (λ=6)	logN (μ=5,σ=1.5)	U (min=40000, max=50000)

The letter before the parenthesis indicates the distribution from which the parameter was drawn; U indicates a uniform distribution, Pois a Poisson distribution, log*N* a log normal distribution. The different distributions were chosen to reflect the domain of the input, for example the number of irrigation events needs to be a positive integer thus the Poisson distribution was chosen. For each of the 2800 simulation a number from each of the 4 distributions described in the column was drawn depending on the simulation treatment.

The difference in earliness was simulated by modifying the *leaf age max* parameter, a parameter that regulates the thermal life span of a cohort of leaves. While the treatments of no stress (control), lack of N (N), poor emergence and water stress could be simulated by modifying the input to the crop growth model, the presence of weeds treatment could not be simulated through Tipstar, as weeds competition is a process that is not included in the model. To simulate the effect of weeds on NDVI, the weed leaf area index (LAI) simulated value was added to the crop LAI. The LAI of weeds was simulated using a second-degree polynomial, that had, at the beginning of the season, a minimum LAI between 0 and 0.5, in the middle of the season, an LAI between 0.2 and 2, and, at the end of the season, an LAI between 0 and 0.5. The effect of weeds on crop growth, i.e. competition for resources, was not simulated.

LSTM anomaly classifier

The architecture of the neural network trained to classify canopy was a long-short term memory (LSTM) neural network. The trained LSTM takes as input, the time series of the relative difference between the NDVI of a control field and the NDVI of the observed field (dNDVIr) calculated as:

$$\text{dNDVIr}=((\text{NDVI\_target} - \text{NDVI\_control})/\text{NDVI\_control})*100 \quad (1)$$

The control field would represent a field that the farmer considers free of anomalies (potential growth). The LSTM model was a many-to-one classifier, meaning that it would take, as input, daily dNDVIr and return one label for the entire time series. The duration of the time series input is variable, so that a farmer could use the time series at various periods of the season. The time series fed to the model are initialized at emergence, so that the first day after emergence would represent the first element of the time series. The LSTM was composed of 3100 parameters and trained on a set of 2000 time series observation, it was validated on 220 synthetic time series and tested on 560 time series. It is important to note that validation on the 220 synthetic time series was only for the sake of preventing model overfitting during the training procedure. A stationary zero mean random noise with a standard deviation of 2% was added to the control treatments during training to simulate the fact that controls where not always 0 since the dNDVIr was calculated for each ground truth plot using as baseline the average of the control for a combination of year and cultivar (otherwise it would be unrealistically always 0).

The validation on the ground-truth observations was carried out independently after model training. The model was trained for 100 epochs and batch size of 100 samples.

Ground truth dataset for validation

A ground truth dataset for validation was collected through a specific field experiment in 2023 and in 2024. In both cases, two cultivars were cultivated on a sandy soil in Wageningen (NL). The sowing distance was 25 cm and the row distance 75 cm. The cultivars cultivated were Avenger (late starch cultivar) and Frieslander (early consumption cultivar). Both years were wet years so the rainfed treatment did not differ from the irrigated treatment as irrigation was actually not needed, thus it was considered an additional control.

All plots received 45 Mg/ha of cattle slurry (estimated N concentration of 4 kg N/Mg, short-term nitrogen fertilizer replacement value 50% Schröder *et al.*, 2007) and 24 kg N/ha in the form of urea (Nutri-Carb N, NTS). The sowing rate was 53 000 seed-potatoes/ha and, in the low emergence plots, 2/3 of the standard amount was sown. For the weed treatment, about half of the normal herbicide dosage was applied until canopy cover. In 2023, the potatoes were planted on 24 of May. The early cultivar (Frieslander) emerged on 12 June, whereas the late cultivar (Avenger) emerged on 26 June. The emergence of the late cultivar was late and poor because potatoes had to be brought back to the shelter since, right before planting, a long rainy period started. This is a problem that generally affected a lot of potato farmers in the NL in 2023. In 2024 potatoes were planted on 28 May and emerged on 14 June.

The plots reflectance (used to calculate NDVI) was measured approximately every 10 days (see Figure 1). The measurements of canopy reflectance were carried out using a TEC5 HandySpec Spectrometer (TEC5 Steinbach, Germany), providing no image but hyperspectral canopy signatures, and a multispectral camera, Altum 5 (AgEagle, Wichita, KS, USA), mounted on a drone. The images from the drone were stitched using Metashape (Agisoft, St. Petersburg, Russia) and NDVI values averaged by plot.

Results

The synthetic data were reasonably close to the observations, as shown for absolute NDVI of control treatment (Figure 1) and for the relative NDVI differences (dNDVIr; Figure 2). There is a discrepancy in NDVI time series of control treatment for the late cultivar of 2023 (Figure 1, 2nd plot from the left) that was probably due to the emergence problem described in the material and methods section.

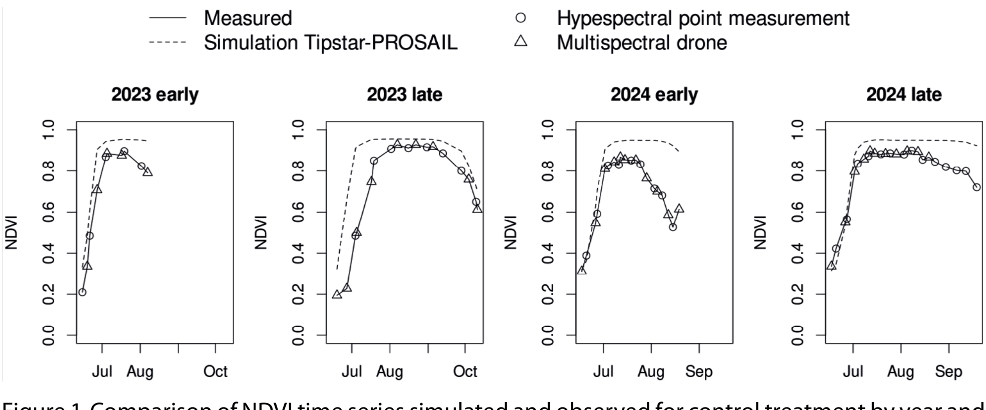

Figure 1. Comparison of NDVI time series simulated and observed for control treatment by year and cultivar.

The relative differences in NDVI time series between treatments reflected the expectations (Figure 2) as described in the introduction section. Weeds showed a higher NDVI at the beginning and at the end of the season, low N treatments had a lower NDVI throughout the season and low emergence treatments had a low NDVI at the beginning of the season, but caught up with the control treatment as the season proceeded. A similar pattern was observed for the synthetic data (Figure 2).
The model was evaluated using balanced accuracy, which is the average between specificity and sensitivity calculated by class. The balance accuracy measure was chosen since there were more controls than the other treatments in our experiment. This was implemented using the R package caret (Kuhn, 2008). The balanced accuracy of the predictions is reported in Table 2.

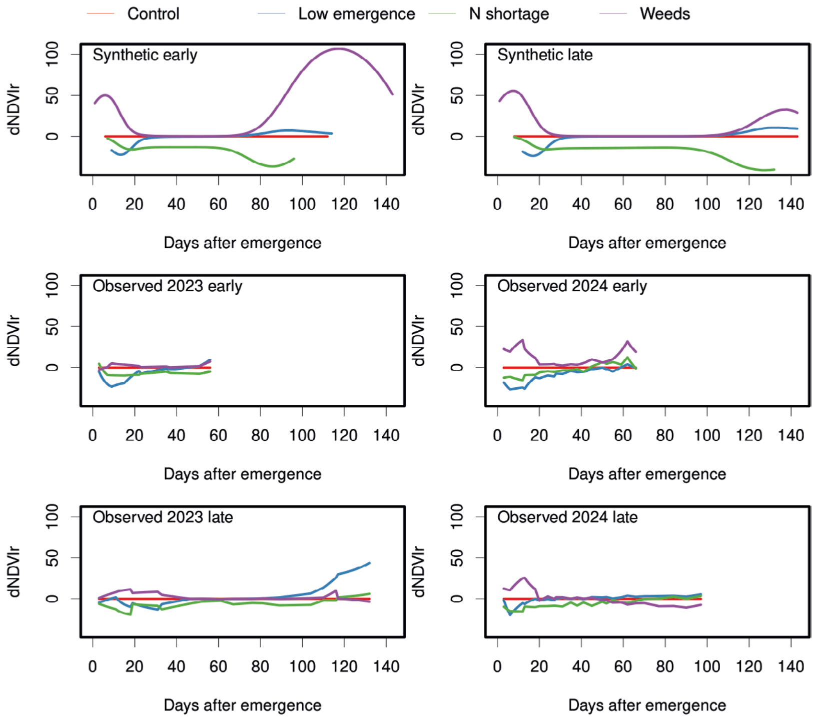

Figure 2. dNDVIr (relative difference NDVI) of synthetic and observed dataset by treatment and cultivar earliness.

Table 2. Balanced accuracy by class, year and earliness.

Year	Cultivar	Weeds	N shortage	Low emergence	Control	Average
2023	Early	0.80	0.80	0.50	0.5	0.65
2024	Early	0.88	0.67	0.79	0.5	0.71
2023	Late	0.90	1.00	0.80	0.5	0.80
2024	Late	0.71	0.79	0.62	0.5	0.66
Average		0.82	0.82	0.68	0.5	0.70

Balanced accuracy is the average between sensitivity (true positive rate) and specificity (true negative rate) calculated by class.

Discussion

The model has an accuracy (cases correctly classified over the total number of cases) that varied between 0.5 (50%) and 1 (100%). The misclassification by class (Table 2) indicates that the most difficult classes to predict were control and low emergence. It is possible that the control treatments were harder to predict because, in the synthetic data, the dNDVIr control was on average 0 (despite having a stationary zero mean noise) whereas in the observed dataset, this could have been non-stationary noise, meaning that a control plot could have consistently lower throughout the season than the average of the control plots. Low emergence was difficult to predict because NDVI is lower than the control only at the beginning of the season and it can be confounded with lack of N, which instead shows its effect throughout the season.

The main challenge to the success of this approach is the construction of a synthetic dataset that is representative of reality. At the moment, the model overestimates the duration of canopy growth specially for the early cultivar. Once a more representative synthetic dataset will be established, it will be possible to move away from the current approach that requires the farmer to indicate a stress-free field (i.e. a control field). The models themselves or their coupling also require further improvements like a weed growth feedback mechanism on crop growth (weeds compete with crop for resources), or a function to predict ground coverage (a parameter of PROSAIL) that includes plant density among the predictors. The current neural network structure can predict multiple stresses or even stress severity at the same time (since it has 4 terminal nodes). However, for the sake of simplicity in this first round, it was trained as a classifier.

Conclusions

A hybrid machine learning approach was applied to the problem of anomaly classification in agriculture and results are encouraging despite the model is far from being precise enough to be deployed in practice. The model has two main advantages, it uses time series of NDVI (a robust vegetation index that can be estimated from satellite images globally) and it can classify the canopy anomalies also in the middle of the season.

Acknowledgements

The project was funded by the Dutch ministry for Agriculture (project number KB-38-001-002)

References

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: a survey. ACM Computing Surveys, 41(3), 15. https://doi.org/10.1145/1541880.1541882

Feret, J.-B., & de Boissieu, F. (2023). PROSAIL: leaf and canopy radiative transfer model and inversion routines. Available online at https://gitlab.com/jbferet/prosail

Heinen, M., Mulder, H.M., Bakker, G., Wösten, J.H.M., Brouwer, F., Teuling, K., & Walvoort, D.J.J. (2022). The Dutch soil physical units map: BOFEK. Geoderma, 427, 116123. https://doi.org/10.1016/j.geoderma.2022.116123

Hui, J., Liu, J., Song, B.-T., & Cong-Hua, X. (2013). Impact of plant density on the formation of potato mimitubers derived from microtubers and tip-cuttings in plastic houses. Journal of Integrative Agriculture, 6, 12. https://doi.org/10.1016/S2095-3119(13)60321-4

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., & Ustin, S.L. (2009). PROSPECT + SAIL models: a review of use for vegetation characterization. Remote Sensing of Environment, 113(Suppl. 1), S56–S66. https://doi.org/10.1016/j.rse.2008.01.026

Khanna, R., Schmid, L., Walter, A., Nieto, J., Siegwart, R., & Liebisch, F. (2019). A spatio temporal spectral framework for plant stress phenotyping. Plant Methods, 15(1), 13. https://doi.org/10.1186/s13007-019-0398-8

Kuhn, M. (2008). Building predictive models in R using the Caret package. Journal of Statistical Software, 28(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Maestrini, B., Mimić, G., van Oort, P.A.J., Jindo, K., Brdar, S., van Evert, F.K., & Athanasiados, I. (2022). Mixing process-based and data-driven approaches in yield prediction. European Journal of Agronomy, 139, 126569. https://doi.org/10.1016/j.eja.2022.126569

Rodriguez, D., Van Oijen, M., & Schapendonk, A.H.M.C. (1999). LINGRA-CC: A sink-source model to simulate the impact of climate change and management on grassland productivity. New Phytologist, 144(2), 359–368. https://doi.org/10.1046/j.1469-8137.1999.00521.x

Roosjen, P.P.J., Brede, B., Suomalainen, J.M., Bartholomeus, H.M., Kooistra, L., & Clevers, J.G.P.W. (2018). Improved estimation of leaf area index and leaf chlorophyll content of a potato crop using multi-angle spectral data – potential of unmanned aerial vehicle imagery. International Journal of Applied Earth Observation and Geoinformation, 66, 14–26. https://doi.org/10.1016/j.jag.2017.10.012

Schröder, J.J., Uenk, D., & Hilhorst, G.J. (2007). Long-term nitrogen fertilizer replacement value of cattle manures applied to cut grassland. Plant and Soil, 299(1–2), 83–99. https://doi.org/10.1007/s11104-007-9365-7

Tiemens-Hulscher, M., Eisinger, E., & Lammerts Van Bueren, E. (2013). Potato breeding: a practical manual for the potato chain. Aardappelwereld, The Hague.

van Genuchten, M.Th. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal, 44(5), 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

van Oort, P.A.J., Maestrini, B., Pronk, A.A., Vaessen, H., & van Evert, F.K. (2024). A simulation study to quantify the effect of sidedress fertilisation on N leaching and potato yield. Field Crops Research, 314, 109425. https://doi.org/10.1016/J.FCR.2024.109425

Vos, J., & Bom, M. (1993). Hand-held chlorophyll meter: a promising tool to assess the nitrogen status of potato foliage. Potato Research, 36(4), 301–308. https://doi.org/10.1007/BF02361796