DOI: 10.32615/ps.2019.017 PHOTOSYNTHETICA 57 (1): 226-230, 2019

BRIEF COMMUNICATION

Machine learning in determination of water saturation deficit in wheat leaves on basis of Chl *a* fluorescence parameters

K. RYBKA[*,+], M. JANASZEK-MAŃKOWSKA[**], P. SIEDLARZ[*], and D. MAŃKOWSKI[*]

Plant Breeding and Acclimatization Institute-National Research Institute, IHAR-PIB Radzików, 05-870 Błonie, Poland[*]
Warsaw University of Life Sciences-SGGW, Nowoursynowska 166, 02-787 Warsaw, Poland[**]

Abstract

Water saturation deficit (WSD) is a parameter commonly used for detection of plant tolerance to temporary water shortages. However, this parameter does not meet criteria set for screening. On the other hand, measurement of chlorophyll (Chl) *a* fluorescence is a fast and high-throughput method. This work presents the application of learning systems to set up a model between WSD and Chl *a* fluorescence parameters allowed for development of a new screening test. Multilayer perceptron (MLP) was trained to predict WSD values on the basis of Chl *a* fluorescence. The best MLP consisted of three inputs: maximal quantum yield of PSII photochemistry, approximated number of active PSII reaction centres per absorption, and measure of forward electron transport, three hidden nodes and one output (WSD). The MLP precision was 82% with a correlation coefficient of 0.98. Continuous improvement of MLP structure and model adaptation to new data takes place.

Additional key words: artificial intelligence, drought, photosynthesis, *Triticum aestivum*.

Requirements of continuous yield increase arising from the need to feed the growing human population result in difficult challenges for breeders to produce more of better yielding varieties. Since the abiotic stresses, such as salinity, drought, heat or frost, affect approximately 40% of the world land area and can limit, even by half, the agricultural production worldwide, the scientific investigations on improvement of plant stress tolerance can help in handling that difficult breeding task (Rybka and Nita 2015, Miazek *et al.* 2017). Most of types of plant tolerance to environmental stresses are of qualitative character (QTL) and are strongly modified by environmental conditions. This is why modern research on such complex phenomena requires approaches using high computing power for modelling of multi-parameter data, generated using new technologies. In plant breeding, great expectations are associated with phenomics, which already allow for automatic detection of physiological state of plants in the field and which enable building of new crop breeding systems in the next future (Brown *et al.* 2014, Grosskinsky *et al.* 2015, Krajewski *et al.* 2015, Vadez *et al.* 2015).

We believe that inclusion of water saturation deficit (WSD), as a feature of a field screening, can be valuable for breeding progress. WSD refers to the degree of tissues dehydration, phenomenon which occurs as one of the first symptoms of plant responses to various environmental stimuli, such as soil salinity, both drought, and flooding, and temperature over or under optimal ranges (Grudkowska and Zagdańska 2010). Plant wilting is an easy to notice as a physiological reaction, associated with complex biochemical processes leading to induction of the stress tolerance or the programmed cell death occurring under extreme situations in tissues of stress-susceptible plants (Kosová *et al.* 2015, Miazek *et al.* 2017). WSD fluctuations as a simple phenotypic trait meets the requirements for being an indicator of plant

Received 12 January 2018, *accepted* 11 July 2018.
[+]Corresponding author; e-mail: k.rybka@ihar.edu.pl
Abbreviations: $(1 - V_j)/V_j$ – measure of forward electron transport; AI – artificial intelligence; ANN – artificial neural networks; BFGS – Broyden-Fletcher-Goldfarb-Shanno learning algorithm; Chl – chlorophyll; ETE – extra training set; F_v/F_0 – maximal efficiency of water photolysis on the PSII donor site; F_v/F_m – maximal quantum yield of PSII photochemistry; GE – global error; HL – hidden layer; IL – input layer; MLP(s) – multilayer perceptron(s); OL – output layer; PI_{ABS} – performance index; *r* – correlation coefficient; RC/ABS – approximated number of active PSII reaction centres per absorption; TE – test set; TF(s) – transfer function(s); TR – training set; VA – validation set; VCdim – Vapnik-Chervonenkis dimension [-]; WSD – water saturation deficit [%].
Acknowledgements: The authors express their thanks to prof. Vasilij Goltsev from Sofia University, Bulgaria, for the discussion on the fluorescence parameters on which the model was built. The research is supported by the National Centre for Research and Development, Poland, as PBS3/B8/19/2015 grant. The training data set was generated within the NN 304267540 grant supported by The National Science Centre, Poland.

stress tolerance (Noctor *et al.* 2015, Gietler *et al.* 2016, Miazek *et al.* 2017). Unfortunately, mere WSD detection does not meet requirements of the screening test due to the low bandwidth. However, detection of Chl *a* fluorescence, in total measurement speed if one sample per 15 s, meets those terms, as far as the scope of manual phenotyping is concerned (Strasser *et al.* 1995, 2004; Stirbet *et al.* 2018). Since both traits are affected by environmental stimuli, simple reasoning resulted in effort of indirect determination of WSD based on modelling of fluorescence data.

In general, data collected in time of plant screening for breeding purposes are of qualitative and quantitative character. Processing of these data by building models to describe and analyse the physiological state of plants is hindered by the nature of field data, data structure, variability, and quantity as well as by interactions and non-linear dependencies occurring between variables of interest (Singh *et al.* 2016). Practice shows that artificial neural networks (ANN) are efficient tools to cope with the modelling of such complex systems even though regular patterns of behaviour are not clearly defined. ANN deal successfully with noisy and incomplete data and have an ability to generalise obtained knowledge. For this reason, they are increasingly being used for modelling of complex, multi-dimensional issues or between-variable dependencies that are difficult to formalise. Among ANNs multilayer perceptrons (MLPs) are easy to apply and their ability to set up optimal parameters allow approximate solutions for extremely complex problems (Belue and Bauer 1993). MLP is a feed-forward, supervised-trained ANN consisted of input (IL), hidden (HL), and output layer (OL), each one built from single nodes called artificial neurons. Input signals are multiplied by weights in each layer and then transformed by transfer functions (TF) set for HL and OL. MLP training consists in iterative weights adaptation in order to provide the best fit of output to experimental data. Training algorithm adapts weights iteratively according to information about gradient of objective function and the direction of its minimisation. While MLP usually consists of a single IL and single OL with number of nodes resulting from the number of input and output variables respectively, number of HLs may be arbitrary, but usually one HL is enough to approximate each constrained continuous function with any low error (Hecht-Nielsen 1987, Cybenko 1989). The number of hidden nodes is crucial for providing the best generalisation ability which is the function of all node connections corresponding to the model complexity (capacity) described by Vapnik-Chervonenki´s dimension (VCdim) (Hush and Horne, 1993). Generalisation ability is dependent on the number of training cases to VCdim ratio. Practice shows that this ratio should be at least 10 (Janaszek *et al.* 2011).

In this paper, we attempted to provide a multilayer perceptron to predict WSD on the basis of Chl *a* fluorescence parameters.

WSD and corresponding Chl *a* fluorescence data were collected as part of programmes on the effects of water shortages on crop yields, run in Plant Breeding and Acclimatization Institute–National Research Institute, Poland (latitude: N52°12′50.017″, longitude: E20°38′38.744″). Data collected on 10-d-old seedlings of spring wheat (*Triticum aestivum* L.) cultivars (from experiments of post-registration varieties evaluation) and advanced breeding lines [from single seed descent (SSD) breeding programs] were chosen as a general data set used for MLP training and testing**.** Seedlings grew as it follows: overnight germination of seeds following surface sterilization in 1% sodium hypochlorite; rolling sets of 25 seeds into filter paper strips (25 × 5 cm), prior to placing into plastic boxes filled to half with Knopp solution supplemented with Hoagland's micronutrients; 10-d growth of seedlings in a climatic chamber (day/night temperature of 18/14°C, photoperiod of 16/8 h, PPFD of 150 µmol m^{-2} s^{-1}, and 70–80% air humidity). For gravimetric determination of WSD, the nutrient solution drainage for 24–48 h was done in order to induce differences in tissues hydration (Gietler *et al.* 2016, Miazek *et al.* 2017). To check the versatility of MLP, additional data were used as an extra test set (ETE). Those data were obtained in the greenhouse measurements during LED lamps usefulness testing for SSD breeding. Plants were grown in the greenhouse of *Plant Breeding Strzelce Ltd.*, *Co. Group IHAR* in Strzelce, Poland (latitude: N 52°18′20.944″ longitude: E 19°24′7.534″) in propagation trays at temperature for night/day of 16/22 ± 3°C and humidity of 70–80% and were lighted for 8/24 h by PPFD 120–150 µmol m^{-2} s^{-1} from 13 LED lamps of different spectra (based on both: red–blue and white LEDs) and HPS as a control. Plants under such cultivation differed markedly from plants grown in the field and were about 30–40 cm high with narrow, delicate leaves. Data of WSD and Chl *a* parameters were obtained during the earing stage on penultimate leaves (Stefański *et al.* 2018). In all experiments, the WSD was determined gravimetrically in three replicas (Miazek *et al.* 2017). Chl *a* fluorescence, in all experiments, was measured on 20-min dark-adapted leaf blades on abaxial side (ten replica/case), using *PocketPEA* portable fluorometer (*Hansatech Instruments*, King's Lynn, Norfolk, UK). Fluorescence was induced by saturating, red actinic light with energy of 3,500 µmol(photon) m^{-2} s^{-1}, and first 3 s of transient fluorescence covering more than its exponential growing part were registered. Following parameters of Chl *a* fluorescence were used for modelling: maximal quantum yield of PSII (F_v/F_m), approximated number of active PSII reaction centres (RC/ABS), measure of forward electron transport [$(1 - V_j)/V_j$], performance index (PI_{ABS}), and maximal efficiency of water photolysis on the PSII donor site (F_v/F_0) (Żurek *et al.* 2014).

The WSD prediction was realised by MLP consisted of three layers. Single MLP input node corresponded to one input variable and the output node corresponded to WSD. General data set used for this study included 237 cases, which were divided into a training set (TR) that covered 90% of cases presented during whole training process and test set (TE) which included 10% of cases, unused in MLP training. Within TR, a validation set (VA) of 20% cases was allocated to control the training process. Another extra TE (ETE) was allocated for final verification of the best MLP and included 14 cases obtained from independent

experiment with wheat grown in a greenhouse under different LED illuminations. The MLP task was to map input variables onto one output using weights adaptation and TF to obtain high correlation coefficient (r) between experimental and predicted WSD values as well as the lowest global error (GE) given as:

$$GE = \sqrt{\sum_{i=1}^{n}\left(z_i - y_i\right)^2 / \sum_{i=1}^{n} z_i^2} \quad (1)$$

where n denotes number of test cases, z_i means WSD experimental value, y_i stands for MLP output and the nominator under the root denotes sum of squares of residuals. Since neither the structure of the best MLP nor the best combination of TFs were known, MLPs with different number of inputs as well as 4 different TFs (identity, logistic, exponential, and hyperbolic tangent) were tested to select model that met the requirements. During training inputs and outputs were normalised to obtain values in range 0–1. All tested MLPs were implemented in *Statistica 13.1* (*Dell Inc*., 2016) and trained according to quasi-Newtonian Broyden-Fletcher-Goldfarb-Shanno (BFGS) training algorithm. Weights in the first training cycle were randomly initiated and during whole training process weight reduction was realized in the range of 0.0001–0.001 in both HL and OL. After choosing the best MLP model, its sensitivity analysis was made to simulate how reduction of input variables affects the predictive capabilities of the model. The sensitivity measure is the ratio of residual sum of squares calculated if a single variable is excluded from a set of MLP inputs to the residual sum of squares calculated if all input variables are considered. The greater the ratio the more sensitive model is to the lack of excluded variable.

To account for the initial weight variance, each combination of input variables with set of TFs was tested by m = 2,000 MLPs, giving a total of 62,000 tested models. From each m^{th} group of MLPs, the best five were chosen in terms of quality measured as r between experimental and approximated WSD values included in the TE. As a result, we identified 155 models with r in the range of 0.91–0.98. Then, the best models left were grouped according to the number of input variables (i) and in each group i = (1, 2, ..., 5) one model was identified with the lowest GE (Table 1). Each of top five models processed data with logistic TF in the HL and exponential TF in the OL. Their GE ranged from 0.18 to 0.26 and r values were in the range of 0.96–0.98. Based on GE, r, and model complexity, the best MLP was selected. The winning MLP was based on $(1 - V_j)/V_j$, RC/ABS, and F_v/F_m parameters used to approximate WSD values (Table 1). The best MLP was characterised by GE of 0.18 with corresponding r of 0.98, both determined for the TE (Fig. 1*A*). The best MLP generalisation ability was additionally verified on ETE (Fig. 1*A*). Obtained results were above the expectations (GE = 0.14). The reason was low variation of input variables and WSD values included in ETE. It was discovered that low WSD values and corresponding Chl *a* parameters were oversampled in general data set and thus also in TR. Additionally, those Chl *a* fluorescence parameters have a high variation whereas corresponding WSD values have lower variability and consequently disrupted the training process. Nonetheless, such small difference in GE between TE and ETE indicates that MLP acquired generalisation abilities during training and this fact encourages further analysis and model improvements. In order to check MLP stability, additional 2,000 models were built using the same 237 cases but different, randomly selected combinations of test cases. Each model had the same proportion of cases in TR, VA, and TE, as well as the same structure and TFs as the previously chosen MLP. As a result, three MLPs, characterised by the highest quality, were chosen. Obtained results might be considered similar since GE varied from 0.2 to 0.17, whereas correlations ranged from 0.96 to 0.97 (Fig. 1*B*). This confirmed that the best MLP was stable and its parameters (weights) were not accidental. It should be pointed that MLP with GE = 0.17 (better than the MLP chosen previously) has lower quality (r = 0.97). It is interesting and noteworthy fact that the model included three parameters describing the functional activity at three main stages of energy transformation within PSII: (*1*) the light energy harvesting (the parameter RC/ABS describes a relative concentration of active PSII reaction centers in total pigments harvesting the light energy for PSII); (*2*) the energy trapping (the parameter F_v/F_m relates to maximum efficiency of the light reactions and efficiency of energy trapping by PSII reaction centers); (*3*) electron transport in ETC (the parameter $[(1 - V_j)/V_j]$ reflects potential efficiency of secondary electron transfer leading to stabilization of charges primarily separated in PSII reaction center) (Strasser *et al.* 1995, 2004; Goltsev *et al.* 2016, Paunov *et al.* 2018, Stirbet *et al.* 2018). In general,

Table 1. The characteristic of top five multilayer perceptrons. MLP structure – number of nodes in input - hidden - output layer; complexity – number of modified weights (input – hidden and hidden – output connections); GE – global error; r – correlation coefficient.

Input variables	MLP structure	complexity	GE	r
$(1 - V_j)/V_j$	1 - 3 - 1	6	0.26	0.96
$(1 - V_j)/V_j$, RC/ABS	2 - 7 - 1	21	0.21	0.97
$(1 - V_j)/V_j$, RC/ABS, F_v/F_m	3 - 3 - 1	12	0.18	0.98
$(1 - V_j)/V_j$, RC/ABS, F_v/F_m, F_v/F_0	4 - 3 - 1	15	0.19	0.97
$(1 - V_j)/V_j$, RC/ABS, F_v/F_m, F_v/F_0, PI_{ABS}	5 - 2 - 1	12	0.20	0.97

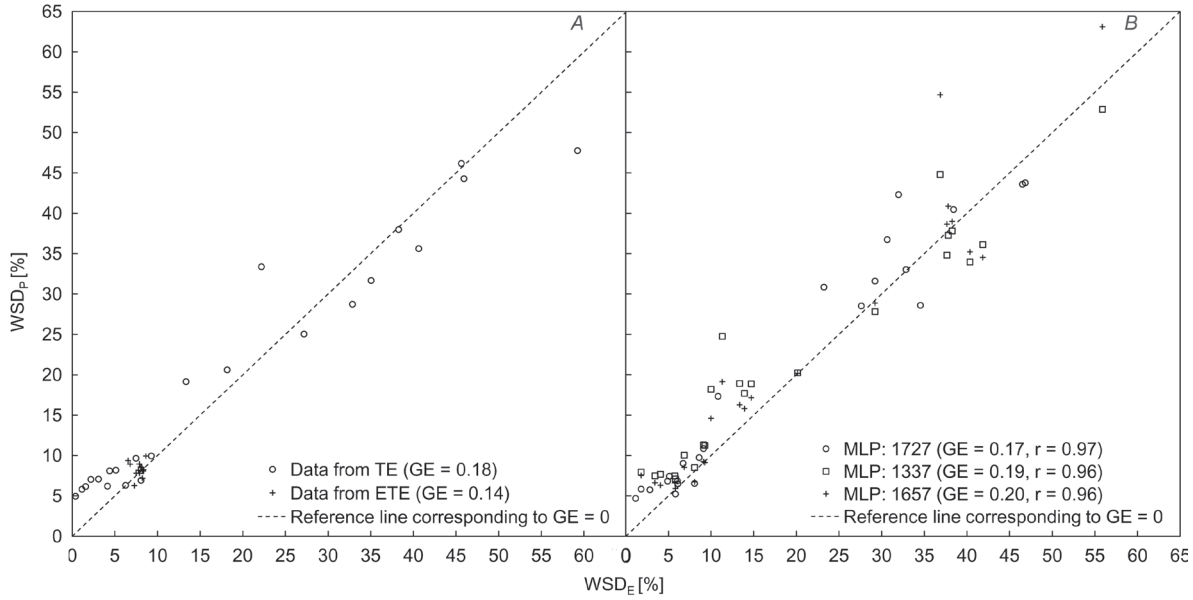

Fig. 1. Real data and data predicted by MLP. WSD – water saturation deficit [%], lower index E – experimental data (original) from test or extra test set, lower index P – data predicted by MLP (MLP output), GE – global error, TE – test set, ETE – extra test set, dotted line corresponds to ideal data mapping, *i.e.* global error of 0 (zero) – full accuracy. Deviations from the *dotted line* represent mapping imperfection (inaccuracy) expressed as global error higher than 0.

plant tissues dehydration leads to changes in hydration and polarisation of cell macromolecules which on hand results in conformational modifications and finally causes their lower biological efficiency (Malferrari *et al.* 2013, Ball 2017). All these phenomena have been described and are experimentally studied in more detail (Kościelniak and Biesaga-Kościelniak 2006, Grzesiak *et al.* 2007, Rapacz 2007, Ashraf and Harris 2013). Goltsev *et al.* (2012) studied the influence of stepwise reduction of relative water (RWC) content in detached primary leaves of 20–25 old pea plants on changes in photosynthetic electron transport, based on the data of raw fluorescence [prompt (PM) and delayed (DF)] and modulated reflection (MR) of light at 820 nm as well as using parameters of the JIP test. The best model for RWC estimation was based on the raw data of PM, DF, and MR curves (Goltsev *et al*. 2012).

Our model combines three variables on the basis of their ability to explain WSD. The sensitivity analysis provided information on inputs hierarchy within the best neural model. Obtained results showed that among three MLP inputs variable $(1 – V_j)/V_j$ was the carrier of the most important information on WSD variability. The residual sum of square ratio calculated for $(1 – V_j)/V_j$ indicated that its exclusion from the input dataset diminished model accuracy about 8.24 times compared to the MLP with all three inputs. In case of RC/ABS exclusion, model accuracy decreased 4.08 times whereas F_v/F_m removal lowered model prediction ability 1.44 times. These results indicated that for WSD prediction, $(1 – V_j)/V_j$ is about twice as important as RC/ABS and almost six times more important than F_v/F_m.

Since general data set did not cover entire range of WSD and additionally middle and high values of this variable were underrepresented, further research is being carried out to collect data accounting for the greatest possible variability of WSD and corresponding Chl *a* fluorescence characteristics. It is also very important to identify all factors, which may modify Chl *a* fluorescence parameters, and in result disrupt a proper WSD prediction. Moreover, attempts of neural model adaptation will continue to obtain its adequate structure ensuring high accuracy and sensitivity.

References

Ashraf M., Harris P.J.C.: Photosynthesis under stressful environments: An overview. – Photosynthetica **51**: 163-190, 2013.

Ball P.: Water is an active matrix of life for cell and molecular biology. – P. Natl. Acad. Sci. USA **114**: 13327-13335, 2017.

Belue L.M., Bauer K.W. Jr.: Determining input features for multilayer perceptrons. – Neurocomputing **7**: 111-121, 1995.

Brown T.B., Cheng R., Sirault X.R.R. *et al*.: TraitCapture: genomic and environment modelling of plant phenomic data. – Curr. Opin. Plant Biol. **18**: 73-79, 2014.

Cybenko G.: Approximation by superpositions of a sigmoidal function. – Math. Control Signal. **2**: 303-314, 1989.

Dell Inc.: Dell Statistica (data analysis software system), ver. 13. – software.dell.com, 2016.

Gietler M., Nykiel M., Zagdańska B.M.: Changes in the reduction state of ascorbate and glutathione, protein oxidation and hydrolysis leading to the development of dehydration intolerance in *Triticum aestivum* L. seedlings. – Plant Growth Regul. **79**: 287-297, 2016.

Goltsev V.N., Kalaji H.M., Paunov M. *et al*.: Variable chlorophyll fluorescence and its use for assessing physiological condition of plant photosynthetic apparatus. – Russ. J. Plant Physl+ **63**: 869-893, 2016.

Goltsev V., Zaharieva I., Chernev P. *et al.*: Drought-induced modifications of photosynthetic electron transport in intact leaves: Analysis and use of neural networks as a tool for a

rapid non-invasive estimation. – Biochim. Biophys. Acta **1817**: 1490-1498, 2012.

Grosskinsky D.K., Svensgaard J., Christensen S., Roitsch T.: Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. – J. Exp. Bot. **66**: 5429-5440, 2015.

Grudkowska M., Zagdańska B.: Acclimation to frost alters proteolytic response of wheat seedlings to drought. – J. Plant Physio. **167**: 1321-1327, 2010.

Hecht-Nielsen R.: Kolmogorov's mapping neural network existence theorem. – In: Proceedings of the International Conference on Neural Networks. Pp. 11-14, IEEE CS Press, 1987.

Hush D., Horne B.: Progress in supervised neural networks. – IEEE Signal Processing Magazine **332**: 8-39, 1993.

Janaszek M., Mańkowski D.R., Kozdój J.: MLP artificial neural networks in predicting the yield of spring barley. – Biuletyn IHAR **259**: 93-112, 2011.

Kosová K., Vítámvás P., Hlaváčková I. *et al*.: Responses of two barley cultivars differing in their salt tolerance to moderate and high salinities and subsequent recovery. – Biol. Plantarum **59**: 106-114, 2015.

Kościelniak J., Biesaga-Kościelniak J.: Photosynthesis and non-photochemical excitation quenching components of chlorophyll excitation in maize and field bean during chilling at different photon flux density. – Photosynthetica **44**: 174-180, 2006.

Krajewski P., Chen D., Ćwiek H., *et al.*: Towards recommendations for metadata and data handling in plant phenotyping. – J. Exp. Bot. **66**: 5417-5427, 2015.

Malferrari M., Mezzetti A., Francia F., Venturoli G.: Effects of dehydration on light-induced conformational changes in bacterial photosynthetic reaction centers probed by optical and differential FTIR spectroscopy. – BBA-Bioenergetics **1827**: 328-339, 2013.

Miazek A., Nykiel M., Rybka K.: Drought tolerance depends on the age of the spring wheat seedlings and differentiates patterns of proteinases – Russ. J. Plant Physiol. **64**: 333-340, 2017.

Noctor G., Lelarge-Trouverie C., Mhamdi A.: The metabolomics of oxidative stress. – Phytochemistry **112**: 33-53, 2015.

Paunov M., Koleva L., Vassilev A. *et al.*: Effects of different metals on photosynthesis: cadmium and zinc affect chlorophyll fluorescence in durum wheat. – Int. J. Mol. Sci. **19**: 787, 2018.

Rapacz M.: Chlorophyll *a* fluorescence transient during freezing and recovery in winter wheat. – Photosynthetica **45**: 409-418, 2007.

Rybka K., Nita Z.: Physiological requirements for wheat ideotypes in response to drought threat. – Acta Physiol. Plant. **e37**: 1-13, 2015.

Singh A., Ganapathysubramanian B., Singh A.K., Sarkar S.: Machine learning for high-throughput stress phenotyping in plants. – Trends Plant Sci. **21**: 110-124, 2016.

Stefański P., Siedlarz P., Matysik P. et al.: The usefulness of light sources based on diodes characterized by a continuous spectrum of white light enriched with a blue band in cereal breeding. – Biuletyn IHAR **283**: 1-11, 2018. [In Polish]

Stirbet A., Lazár D., Kromdijk J., Govindjee: Chlorophyll *a* fluorescence induction: Can just a one-second measurement be used to quantify abiotic stress responses? – Photosynthetica **56**: 86-104, 2018.

Strasser R.J., Srivastava A., Govindjee. Polyphasic chlorophyll *a* fluorescence transient in plants and cyanobacteria. – Photochem. Photobiol. **61**: 32-42, 1995.

Strasser R.J., Tsimilli-Michael M., Srivastava A.: Analysis of the chlorophyll *a* fluorescence transient. – In: Papageorgiou G.C., Govindjee (ed.): Chlorophyll *a* Fluorescence: A Signature of Photosynthesis. Advances in Photosynthesis and Respiration, Vol. 19. Pp. 321-362. Springer, Dordrecht 2004.

Vadez V., Kholová J., Hummel G. *et al.*: LeasyScan: a novel concept combining 3D imaging and lysimetry for high-throughput phenotyping of traits controlling plant water budget. – J. Exp. Bot. **66**: 5581-5593, 2015.

Żurek G., Rybka K., Pogrzeba M. *et al.*: Chlorophyll *a* fluorescence in evaluation of the effect of heavy metal soil contamination on perennial grasses. – PLoS ONE **9**: e91475, 2014.