Euphytica

Towards an objective evaluation of persistency of *Lolium perenne* swards using UAV imagery.

--Manuscript Draft--

Manuscript Number:	EUPH-D-18-00066
Full Title:	Towards an objective evaluation of persistency of <i>Lolium perenne</i> swards using UAV imagery.
Article Type:	Original Article
Keywords:	breeding; persistency; vegetation index; close remote sensing
Abstract:	<p>Perennial ryegrass (<i>Lolium perenne</i>) is a perennial crop used in temperate regions as forage. In <i>L. perenne</i> breeding programs, persistency is an important trait. Poor persistency results in sward degradation and associated yield and nutritive value losses. Breeders assess persistency of selections using visual scoring in field plots during the 2nd or 3rd growing season. This evaluation system is easy and cheap but is not free from human bias. As an alternative we have developed a methodology to assess persistency of <i>L. perenne</i> breeding materials based on Vegetation Indices (VIs) derived from Unmanned Aerial Vehicle (UAV) imagery. The VIs Excess Green (ExG2), Green Lead Index (GLI) and Normalized Green Intensity (GCC) were found to provide consistent results for flights carried out under different light conditions and were validated by ground reference information. The correlation between the VIs and the ground reference was 0.889. To test the implementation of the method we compared the ExG2 value based approach to selection with a visual score based selection methodology as applied by two breeders. The breeding decisions of Breeder A agreed well with decisions based on ExG2 values (74.6%), but those of Breeder B displayed a lower agreement (54.0%). In contrast, agreement between decisions based on different flights was very high (91.6%). The methodology was validated for general applicability. In summary, the results demonstrate that basing persistency selection in <i>L. perenne</i> breeding programs on ExG2 values from UAV imagery is likely to be more objective in comparison to the currently-used visual scoring method.</p>

Towards an objective evaluation of persistency of *Lolium perenne* swards using UAV imagery

Sward persistency evaluation with a UAV

Irene Borra-Serrano^{1,4}, Tom De Swaef¹, Jonas Aper¹, An Ghesquiere¹, Koen Mertens², David Nuyttens², Wouter Saeys³, Ben Somers⁴, Jürgen Vangeyte², Isabel Roldán-Ruiz^{1,5}, Peter Lootens^{*1}

Corresponding author: * peter.lootens@ilvo.vlaanderen.be

¹ Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Plant Sciences Unit, Caritasstraat 39, 9090 Melle (Belgium)

² Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Technology and Food Science Unit, Burg. van Gansberghelaan 115, 9820 Merelbeke (Belgium)

³ KU Leuven, Department of Biosystems, MeBios, Kasteelpark Arenberg 30, 3001 Leuven (Belgium)

⁴ KU Leuven, Division of Forest, Nature and Landscape, Celestijnenlaan 200E, 3001 Leuven (Belgium)

⁵ Ghent University Department of Plant Biotechnology and Bioinformatics, Technologiepark 927, 9052 Ghent (Belgium)

Abstract

Perennial ryegrass (*Lolium perenne*) is a perennial crop used in temperate regions as forage. In *L. perenne* breeding programs, persistency is an important trait. Poor persistency results in sward degradation and associated yield and nutritive value losses. Breeders assess persistency of selections using visual scoring in field plots during the 2nd or 3rd growing season. This evaluation system is easy and cheap but is not free from human bias. As an alternative we have developed a methodology to assess persistency of *L. perenne* breeding materials based on Vegetation Indices (VIs) derived from Unmanned Aerial Vehicle (UAV) imagery. The VIs Excess Green (ExG2), Green Lead Index (GLI) and

Normalized Green Intensity (GCC) were found to provide consistent results for flights carried out under different light conditions and were validated by ground reference information. The correlation between the VIs and the ground reference was 0.889. To test the implementation of the method we compared the ExG2 value based approach to selection with a visual score based selection methodology as applied by two breeders. The breeding decisions of Breeder A agreed well with decisions based on ExG2 values (74.6%), but those of Breeder B displayed a lower agreement (54.0%). In contrast, agreement between decisions based on different flights was very high (91.6%). The methodology was validated for general applicability. In summary, the results demonstrate that basing persistency selection in *L. perenne* breeding programs on ExG2 values from UAV imagery is likely to be more objective in comparison to the currently-used visual scoring method.

Keywords: RPAS; breeding; persistency; vegetation index; close remote sensing

1. INTRODUCTION

Perennial ryegrass (*Lolium perenne* L.) is one of the dominant forage grasses in livestock farming in northwestern Europe in terms of seed sales (Burgon et al., 1998) and land area covered (Wilkins and Humphreys, 2003). Pastures and hay fields containing *L. perenne* are installed and subsequently exploited for multiple years. Excellent persistency is desired because reseeding of grasslands is expensive, sacrifices grass production during the establishment period and often results in substantial nutrient losses to drainage water (Wilkins and Humphreys, 2003). Therefore, persistency is an important evaluation criterion in perennial ryegrass breeding programs along with forage yield, quality, and disease resistance (Wilkins and Humphreys, 2003; Popay and Hume, 2011).

Although a formal definition of persistency is not straightforward, in a perennial forage crop it can be defined as the capacity to survive and produce optimal yields over several growing seasons (Herrmann et al., 2008; Taylor and Quesenberry, 1996). In perennial ryegrass, persistency predominantly affects the yield in the second and subsequent harvest years (Wilkins and Humphreys, 2003). Plant mortality creates open spaces and allows undesirable species to become established in the sward. This results in losses in productivity and nutritive value (Popay and Hume, 2011). Persistency is a complex trait influenced by intrinsic biological properties determining the degree of perenniality together with biotic and abiotic interactions (Casler and van Santen, 2010; Humphreys et al., 2010). Biotic stresses can include pathogens and pests that can cause loss of vigor and mortality in forage plants. Abiotic stresses such as low temperature, heat, drought and flooding can also affect persistency. In general terms, diploid perennial ryegrass varieties are considered to be more persistent than tetraploids. Diploids are longer-lived and form swards of good density (Humphreys et al., 2010), while tetraploid ryegrass plants produce larger leaves and tillers than diploids, resulting in more open swards. However, persistency displays a high variation within each of these groups. In recent years, tetraploid varieties are catching up with diploid varieties in terms of persistency (Chaves et al., 2009).

66

67 In perennial ryegrass breeding programs, enhanced persistence or survivability over the long term is
68 essential. Persistency is usually scored visually through field observations: ground cover is visually
69 evaluated seven days after a cut in the second or third year after sowing, at the end of the production
70 trials. The breeder rates the vegetation cover with scores ranging from 1 to 5 or 1 to 9 (Luscier et al.,
71 2006), with the highest value assigned to the most persistent swards according to greater soil surface
72 coverage and vigor of the plants. The purpose is to select plant material that demonstrates greater
73 production stability and adaptation to the environment. Persistency thus differs from sod density,
74 which is the density of the sward during production. The persistency score is often used as a negative
75 selection criterion: populations that fall below a certain threshold are discarded. Visual scores are
76 characterized by their low cost in terms of equipment and ease of performance. However, visual
77 ratings have a low repeatability and could be biased by the observer (Luscier et al., 2006) and can be
78 time consuming especially in large breeding trials. Therefore, the availability of a high throughput,
79 objective methodology to estimate persistency in perennial ryegrass swards would offer clear
80 advantages for the breeder, potentially allowing a quicker breeding progress for this trait. Digital
81 imaging can remove the bias related to the observer and is also a quick and non-invasive method for
82 collecting data on vegetation cover (Liu et al., 2012).

83

84 In combination with recent advances in digital photography and image analysis software, remote or
85 proximal digital imaging provides opportunities for improved vegetation data collection and analysis
86 (Luscier et al., 2006; Walter et al., 2012). This is also reflected in the multiple studies in which analysis
87 systems have been developed to quantify grass cover based on digital images collected on the ground
88 using tripods (for examples see: Liu et al., 2007; Lootens et al., 2016; Luscier et al., 2006; Lynch et al.,
89 2015; and Richardson et al., 2001). In contrast, a large number of studies demonstrate the potential of
90 Unmanned Aerial Vehicles or “drones” (UAVs) for vegetation mapping and monitoring (Araus and
91 Cairns, 2014; Caturegli et al., 2016; Salamí et al., 2014). This remote sensing technology exhibits high

spatial and temporal resolution, flexibility in terms of flight planning and image acquisition scheduling, versatility, reduced cost and immediate accessibility of the information in comparison to other remote sensing platforms (Xiang and Tian, 2011). Moreover, ground-based digital imaging is limited by scale at (sub)plot level, while UAVs allow simultaneous imaging and characterization of a large number of plots (Araus and Cairns, 2014). Disadvantages of UAV imagery compared to on-ground imagery are that UAV images might have a lower spatial resolution (although <1cm Ground Sampling Distance can be achieved with appropriate flight settings); a delay in the decision-making process due to a longer time required for the data processing (considering the time required for all the steps involved from the preparation of the flight to the reporting of results); or the inability to fly due to poor flight conditions or regulation restrictions.

Depending on the application, different approaches for digital image analysis can be followed. For applications in which explicit spatial information on vegetation cover has to be determined (without the necessity to recognize individual plants or plant parts) spectral vegetation indices (VIs) are commonly used (Guijarro et al., 2015; Z. Y. Liu et al., 2007; Meyer and Neto, 2008; Woebbecke et al., 1995). VIs are algorithms that combine information from two or more spectral bands; these are mathematic calculations using light reflectance at different wavelengths (Rasmussen et al., 2016). VIs constitute a simple and convenient means to extract information from the remotely sensed data contained in spectral reflectance. Their ease of use facilitates the processing and analysis of large amounts of data (Govaerts et al., 1999). Vegetation indices can use bands from different spectral regions. Indices such as NDVI or OSAVI that include the near-infrared region might be more informative but are less human intuitive as human eye is not sensitive to that region of the spectrum (Meyer and Neto, 2008). The indices that involve red, green and blue bands (visual region) emphasize colors such as plant greenness. Authors such as Rasmussen et al. (2016), Meyer and Neto (2008) or Torres-Sanchez et al. (2014) demonstrate different uses for which those indices are suitable related to

vegetation cover. Advantages of using RGB cameras are their easier and quicker handling being more intuitive for their use and processing and lower price.

The aim of this work was to develop a reliable methodology based on UAV-derived VIs to estimate persistency of *L. perenne* selections in a breeding context. We used a field trial established before the start of this investigation to estimate persistency for a set of diploid selections and a set of tetraploid selections from the *L. perenne* breeding program at Flanders Research Institute for Agriculture, Fisheries and Food (ILVO; Melle, Belgium) according to standard procedures (Humphreys et al., 2010). First, robust VIs were identified that rendered stable results over flights carried out under different environmental conditions. Then one VI was selected that showed strong correlation with ground-truth data generated using manually-acquired detailed images that were processed through supervised classification approaches. Third, the level of agreement between the VI values obtained from the UAV images and traditional visual methods of scoring for persistency was estimated through a negative selection of the 50% least persistent populations. The developed methodology was then validated by applying it to an independent field trial. Finally, the advantages and limitations of the developed methodology were discussed in relation to the state of the art evaluation methods.

2. MATERIAL AND METHODS

2.1 Field trials and plant material

For the development of the methodology, we used an experimental field (Field Trial 1; FT1). FT1 was approximately 1 ha in area, located in Melle (Belgium). On that field, 237 *L. perenne* selections were evaluated for persistency in the context of breeding activities. The geographic coordinates (UTM, zone 31N, WGS-84) of the bottom right corner of FT1 were X = 554310 m and Y = 5648345 m. FT1 was sown on April 25th 2013 and was mown five times per year in 2014 and 2015 at a height of 5 cm. The plots were fertilized in 2014 with 256 kg N·ha⁻¹, 273 kg K·ha⁻¹ and 18 kg P·ha⁻¹. In 2015, the plots were

fertilized with 260 kg N·ha⁻¹, 295 kg K·ha⁻¹ and 23 kg P·ha⁻¹. The trial consisted of two sections, corresponding to 102 diploid and 135 tetraploid selections. Each section was divided into four blocks. Within each block, plots of 1.35 x 5.8 m (7.83 m²) were installed, with each selection represented by two or three replicates in a randomized complete block design.

To investigate the general applicability of the methodology, it was validated for a second independent field trial, Field Trial 2 (FT2), which comprised 234 *L. perenne* selections. The geographic coordinates (UTM, zone 31N, WGS-84) of the bottom right corner of FT2 were X = 554216 m and Y = 5648109 m. FT2 was sown on April 14th 2014 and was mown five times per year in 2015 and 2016 at a height of 5 cm. The plots were fertilized in 2015 with 300 kg N·ha⁻¹, 295 kg K·ha⁻¹ and 23 kg P·ha⁻¹. In 2016, the plots were fertilized with 312 kg N·ha⁻¹, 314 kg K·ha⁻¹ and 25 kg P·ha⁻¹. FT2 consisted of two sections, corresponding to 138 diploid (four blocks) and 96 tetraploid selections (three blocks). Plot size was the same as in FT1, and each selection was represented by two or three replicates in a randomized complete block design.

2.2 Data acquisition of FT1

2.2.1 UAV-derived orthophotos

On March 18th and March 22nd 2016 (7 and 11 days after mowing, respectively) a set of nadir overlapping images (60% forward-longitudinal-lap and 60% side-lap) was captured by a sensor mounted on a 12-rotor UAV (model Onyxstar Hydra-12, Altigator, Belgium). For imagery we used a conventional RGB (red-green-blue or visible spectral range) camera (model Sony α6000, Sony Corporation, Tokyo, Japan) equipped with a 35 mm lens (f/1.8, Sony Corporation, Tokyo, Japan). The camera was set to operate in manual mode to avoid different settings in successive images. Exposure time (shutter time), F-stop (aperture) and sensitivity to light (ISO) were adjusted in the field just before the start of each flight and the image trigger was activated by the UAV according to the flight route.

169

170 Per date two flights were carried out, one in the morning around 11:30 (local time) and a second in
171 the afternoon around 14:45, at 50 m flight altitude. Different weather conditions were experienced:
172 the first date was cloudy and the second date was clear sky and sunny, resulting in different light
173 conditions and solar radiation levels during the acquisition of the images. The solar radiation recorded
174 during each of the flights (chronologically) was 175.4 W/m², 75.6 W/m², 552.7 W/m² and 505.5 W/m².

175

176 These four flights rendered four sets of images. The flight route was programmed using the software
177 MikroKopter v2.14b (HiSystems GmbH, Germany) and automated to ensure that the whole
178 experimental field was covered with the necessary overlap to build the orthophoto. Raw images were
179 used as they enable image corrections without data loss. The raw images were pre-processed in
180 Lightroom 6.5 (Adobe Systems Incorporated, San Jose, CA, USA) to correct for white balance and
181 exposure based on a gray card (Grey card, Novoflex) used as a color neutral reference. The gray card
182 was placed on the ground of the field trial during the flights. After corrections were implemented, the
183 raw images were converted to lossless tif files. Agisoft Photoscan Professional Edition (Agisoft LLC, St.
184 Petersburg, Russia) was used for mosaicking and building orthophotos (example provided in Fig. 1A).
185 To create geo-referenced ortho-mosaicked images, five artificial Ground Control Points (GCPs) were
186 geo-referenced using a RTK GPS (model Stonex S10 GNSS, Stonex SRL, Lissone, Italy) and identified in
187 the images. The four resulting ortho-mosaicked images had a spatial resolution of 0.52 cm defined
188 automatically by the software based on the camera parameters (sensor size and focal length). The
189 geo-referencing error of the orthophoto ranged from 0.04 to 0.09 m. This reference error represents
190 the distance between the coordinates of the GCP measured by the GPS and the estimated position in
191 the orthophoto (root mean square error). It was used to decide if an additional refinement of the
192 marker location was required or if a GCP should be excluded. The error obtained in this experiment
193 was less than 0.10 m and was thus considered acceptable.

194

2.2.2 Ground-truth assessment

To estimate and compare the performance to assess persistency by different VIs derived from UAV images (see section 2.2.3 below), a ground-truth reference dataset was generated for a subset of 30 plots distributed randomly across the trial (Fig. 1B). In each of these 30 plots on-ground images of 0.30 m x 0.50 m frames, which were geo-referenced with RTK GPS, were captured on March 18th 2016 using a Nikon D90 camera (Nikon Corporation, Tokyo, Japan) mounted on a tripod at a height of 1.2 m in nadir position. The frame was placed in the middle of the plot to avoid border effects and at approximately the same location for all the plots to avoid bias due to positioning of the frame (Fig. 1C). To avoid and ensure that the shade of the tripod was beside the frame area and not on the image, the tripod was placed opposite to the sun. In figures 1C and 1D the positioning can be seen and in Figure 4A two examples of frame-images are shown.

These 30 on-ground images were processed using the supervised classification method 'Spectral Angle Mapper' (SAM) (Kruse et al., 1993) in ENVI/IDL 5.0 (Exelis Inc., Boulder, CO, USA). We defined three classes: 'green grass', 'bare soil' and 'old/dead grass' (not green). Training and validation pixels, called regions of interest (ROIs), were randomly selected for each class in the frame images ensuring that the set of pixels used for validation was different from that used for training. Training ROIs were used as the basis to conduct the classification. Validation ROIs were used to assess the accuracy of the classification by computation of a confusion matrix and Kappa coefficient. To this end, each validation ROI was located in the classification image and the class assigned was compared to the one given by the operator during the selection of pixels for this analysis. Classification match (or coincidence) was determined through the confusion matrix.

Once the procedure had been validated, the class 'green grass' was used to estimate the percentage of area covered by green grass (A-GG) in each frame. In this study, only 'green grass' was considered,

as we only used vegetation indices that highlight and differentiate green vegetation from other features (bare soil or old grass).

2.2.3 Vegetation indices

A set of 13 vegetation indices (VIs) proposed by different researchers working on RGB imagery were selected as candidates for the persistency evaluation (Table 1). For each individual plot of the experiment an average value of each of these indices was determined from UAV orthophotos, excluding the border (15 cm buffer) to avoid influence from pixels that do not belong to the plots.

In addition, we located the 30 georeferenced on-ground frames used as ground-truth (Section 2.2.3) on the orthophoto, defined a polygon with the same dimensions and calculated the average value of the different VIs for the selected frame area. The location of the frames and the values of the VIs for plots and frames were determined in QGIS 2.8.7-Wien (QGIS Geographic Information System. Open Source Geospatial Foundation Project). The VI values and green grass fractions (A-GG) estimated for each frame were used for regression analysis to check whether the VIs were delivering correct information on the green cover fraction.

2.2.4 Visual scores of persistency

Two experienced breeders visually scored the plots on March 17th 2016 (six days after mowing, and one day before the first UAV flight) using the 1-9 persistency scoring system applied routinely in breeding, with a score of 1 representing low persistency and 9 high persistency. Breeder A scored the diploid plots and Breeder B scored the tetraploid plots, according to standard practice in the ILVO breeding program. In addition, they were requested to score in parallel a subset of 90 plots randomly chosen across the experiment (including both diploid and tetraploids).

This allowed us to evaluate (i) the repeatability of the scores assigned at different times by the same breeder (plots scored two times, 36 plots for diploids and 54 for tetraploids by the same breeder), (ii) the agreement between scores given to the same plot by two different breeders (90 plots from second score), and most importantly (iii) to compare these visual scores with UAV derived data, as described in section 2.2.5.

2.2.5 Data handling and statistical analyses

To facilitate the understanding of the different steps done to develop the methodology the workflow is schematically illustrated in Figure 2, where the data used in each part are indicated.

First we identified VIs that rendered consistent and stable results for different flights (and consequently under different solar radiation conditions) on the same day using a subset of 90 plots (the same set to which visual scores were assigned twice as described in section 2.2.4 above). We only considered comparisons of UAV flights within one day (morning and afternoon), as the plants' condition may have changed between the two flight dates. Therefore, in this study, a suitable index is one that renders the same value for different flights on the same day, meaning that the VI value is as stable as possible and minimally influenced by the time of flight or flight conditions. The selection was based on Pearson correlation values and their significances, and on the intercept and slope of the linear regression. The following criteria were applied for selection of candidate VIs: correlation higher than 0.9 and significant ($p < 0.001$), intercept between -0.1 and 0.1, and slope between 0.9 and 1.1 to avoid offsets. For selection, the VI had to fulfil the criteria for both dates investigated.

The second step (for the VIs selected in the previous step) was to determine whether the VI values derived from UAV flights were representative for the ground-truth values obtained using 30 on-ground images of small frames (0.3 m x 0.5 m) within the plots. We calculated the correlations between the VI values derived from the frame area in the orthophoto and the percentage of area

covered by green grass (A-GG) in the on-ground photo frames and built different regression models. Multiple linear regression models were constructed using backward elimination, forward selection and stepwise selection to obtain the indices or combinations of indices that better represented A-GG. All the statistical analyses were carried out in RStudio 0.99.879 (RStudio: Integrated Development Environment for R, RStudio Inc., Boston, MA, USA). Because the performance of different VIs was similar and end results were highly correlated (see section 3.1.3), we chose one of them for further analyses.

For the VI selected in the previous step, correlation coefficients were calculated for all pair-wise combinations of flights using data from a subset of 90 plots. This was done to assess the effect of the time interval between mowing and UAV flight. UAVs cannot operate under certain weather conditions thus the chosen date is not always viable. The consistency of scores assigned to the same plot by each breeder at different times and the agreement of scores assigned by different breeders were estimated using the same subset of 90 plots ($n_{2x} = 34$, $n_{4x} = 56$). Finally, we tested the practical implementation of the procedure developed in the previous steps by comparing it with the current methodology to estimate persistency in perennial ryegrass. In practice, ILVO breeders perform a negative selection by the elimination of the 50% of populations with the lowest persistency score. We simulated this type of selection using the best performing vegetation index related to persistency and results were compared to selections based on breeder scores. The comparison was carried out for diploids and tetraploids separately with the full set of plots, as is usually done in practical breeding. This allowed us to determine the potential impact of shifting from a system based on visual scoring to a system based on UAV derived images and VIs. For the selection based on breeder scores we followed the standard procedure in the ILVO breeding program. This is based on the processing of the data in Trialwizard (Beerepoot Consultancy BV, The Netherlands), which performs corrections based on ‘moving mean’ and ‘nearest-neighbor’ analysis (taking into account the experimental design and the replicates present in the field), rendering one single persistency score per selection. For the UAV-derived data,

the averages of the VI values estimated per plot were processed in a similar way as the persistency scores, using Trialwizard to obtain a single index value per population. To establish the limits of the 50% negative selection, the median of score and index values were determined.

2.3 Application of the methodology to FT2

FT2 was used to test the general applicability of the developed methodology. The two abovementioned experienced breeders were responsible for the persistency scoring in this field. The breeder's persistence rating and the UAV flight were performed on October 10th 2016, six days after mowing. A set of nadir images was collected with the same sensor and procedure described in section 2.2.2 at 55 m altitude around 14h00 (80% forward-longitudinal-lap and 60% side-lap). In this case, we only considered the VI selected in previous steps. Data analysis was as described in section 2.2.5.

3. RESULTS

3.1 Development of the methodology: FT1

3.1.1 Selection of robust UAV-based VIs

Based on the VIs obtained for a subset of 90 plots, the stability of the index values between the two flights carried out on the same day was calculated (Table 2). Only three of the investigated VIs fulfilled the pre-defined criteria (see section 2.2.5): ExG2, GCC and GLI. These VIs showed highly significant correlations between flights carried out on the same day, an intercept close to 0 and a slope around 1 (Fig. 3). These results demonstrate that values obtained under different light conditions for these three UAV-based VIs were highly stable.

3.1.2 Correspondence between UAV-based VIs and ground-truth

The next step was to investigate whether the three UAV-based VIs selected in the previous step delivered accurate information regarding green cover. For this comparison we used pictures of a

subset of 30 on-ground frames to generate a ground-truth dataset. These 30 frames were considered representative for the state of the corresponding entire plots: correlation values higher than 0.845 (Supplementary data 1) were calculated between the VIs derived from the area of the frames in the correct location on the orthophoto (VI_UAV_Frame) and the value of the VI of the plots where the frames were placed (VI_UAV_Plot). Subsequently, the on-ground images of the frames were suitable to relate and validate the use of vegetation indices to estimate persistency.

The supervised classification of the pixels in the 30 on-ground frames was performed with a mean overall accuracy of 85% and Kappa coefficient of 0.78. The confusion matrix and the Kappa coefficient were calculated for each frame separately, rendering different accuracies for each frame. The classification matrix of an average plot (Table 3) shows that 95.9% of the green grass is classified as green grass. The classification of old grass and bare soil was less straightforward, although 78.9 and 88.6% of the bare soil and old grass, respectively, were correctly classified. Especially bare soil and old grass were less easy to identify separately, but they were confused less frequently with green grass (Table 3).

As the classification gave a high mean overall accuracy, the estimation of the percentage of area covered by green grass (class 1), bare soil (class 2) and old grass (class 3) was executed. In Fig. 4A-D two frames are shown together with their classifications and percentage of area covered by green grass (A-GG) considered as ground-truth.

Correlations between the VI_UAV_Frame and A-GG were calculated for the 13 VIs (results not shown). The selected VIs based on the criteria established previously for stability (section 3.1.1) resulted here in the highest correlations: $r = 0.889$, which supported their use to assess vegetation cover and hence persistency. Multiple linear regression models using all 13 or a selection of VIs were performed, but did not lead to improved correlations. Therefore, the three UAV-based VIs were considered the most

appropriate to fit linear regressions for our purposes. In Fig. 5A-C the linear regression is illustrated for the morning flight on the 18th March for ExG2, GLI, GCC and the area of the frame covered by grass (A-GG).

3.1.3 Choice of one representative UAV-based VI

Because the scatter plots and regression lines for the three UAV-based VIs selected show high similarities (Fig. 3, Fig. 5) and their values are strongly correlated ($r = 0.981$), one can question if these are actually different or if they capture the same information. Therefore, the relation between ExG2, GCC and GLI was evaluated using a random sample of 1000 RGB values for which the indices were calculated (Fig. 6). ExG2 and GCC are related linearly and ExG2 and GLI curvilinearly. However, as the range of the values derived from the orthophotos (indicated by the dashed lines in Fig. 6) was much smaller than in the random sample, all pair-wise relations became approximately linear. Therefore, we can conclude that the three selected VIs delivered approximately the same information in this experiment. For that reason and for the sake of simplicity, further analyses were carried out using ExG2 exclusively, as this index is the most widely used of the three.

3.1.4 Inter-day variability

The next step was to calculate correlations between the four ExG2 datasets (Table 4) to estimate if flights performed at a five-day interval deliver similar information. In all cases, the correlations were ≥ 0.871 , even though between the dates, the plants had grown and the weather conditions were different. This shows that during this period of the year, flights separated by five days delivered similar results in this experiment, adding flexibility to the implementation of a methodology based on the use of visual imaging from UAVs when the planning of flights is constrained by climatological conditions.

3.1.5 Comparison of the persistency assessment though UAV-based VIs and breeder scores

The results of the usual procedure of negative selection were compared to the decisions that would have been taken if UAV-based ExG2 values had been used (Fig. 7). First, based on the repeated scores given by the two breeders to the set of 90 plots we estimated the within-breeder consistency and the agreement of scores assigned by the different breeders (Table 5). Both breeders assigned consistent scores when they evaluated the same plot two times, with correlation values of 0.754 and 0.760 for Breeder A and B, respectively. However, it should be noted here that the agreement between the scoring by different breeders was only 0.243 (Table 5). The two breeders involved in this experiment probably assess persistency using different criteria. Moreover, Breeder B evaluates the plots using a narrower range of scores than Breeder A (results not shown), which may also affect the inter-breeder consistency.

Second, we compared the decisions of Breeder A and Breeder B (separately) to ExG2 data from the flight closest to the scoring (morning of 18th of March) and of the flight that was separated by the longest interval from the scoring (afternoon of 22nd March). Note that in these comparisons we took the scores of Breeder A for diploid populations and of Breeder B for tetraploid populations. For the selection of the 50% worst populations (negative selection) the scores of Breeder A and the ExG2 values of the flight closest to the scoring displayed a degree of agreement of 74.6% (Fig. 7A). Correspondingly, for 25.4% of the populations a different decision would have been made using the criteria of Breeder A and the ExG2 values. Corresponding values for Breeder B were lower: 54% for the negative selection (Fig. 7C). Results obtained taking into consideration the afternoon flight of 22nd of March were comparable (Fig. 7B and Fig. 7D for breeders A and B, respectively). Taken together, these results indicate that the agreement between decisions taken based on the UAV-based EXG2 data and the visual scoring by a breeder is comparable to the agreement between the scoring done by different breeders.

Finally, to complete the analysis, we also compared the consistency of the decisions that would be taken using information from different flights (Fig. 7 and Table 4). As expected, a high degree of agreement was found between flights, with higher values for comparisons involving flights on the same date (91.6%, Fig. 7E) than when the flights with the lowest degree of agreement (Table 4) were compared (85.6%, Fig. 7F).

3.2 Validation on an independent test field

In FT2, the negative selection of the 50% least persistent selections showed an agreement of 60.9% between the ExG2 values (UAV-flight derived) and the score of Breeder A (Fig. 8A). The degree of agreement with the scores of Breeder B was 58.4% (Fig. 8B). These values are similar to those obtained for FT1, except that the agreement value for Breeder A was slightly lower in this case. The range of scores assigned by Breeder A to the selections investigated in FT2 was lower than for FT1, whereas the opposite was true for Breeder B. This probably reflects a different range of variation in the selections investigated in FT1 and FT2.

4. DISCUSSION

In this study, we propose a methodology based on the use of vegetation indices derived from visual UAV imagery to estimate persistency in diploid and tetraploid perennial ryegrass in the context of breeding activities. The developed methodology has resulted in an accurate method to measure ground cover as a tool to determine the persistency of different accessions.

4.1 Three UAV-based VIs deliver robust results under different environmental conditions

The use of vegetation indices is widespread (Guijarro et al., 2015; Z. Y. Liu et al., 2007; Meyer and Neto, 2008; and others). As outdoor light conditions can differ greatly, it is important to use indices that are as insensitive as possible to these differences. Atmospheric and lighting conditions change throughout a day, resulting in significant differences between the light intensities at which different

images are acquired (Guijarro et al., 2015). In this study, the first criterion to select suitable VIs was therefore their robustness against changes in lighting conditions. Three of the 13 tested VIs tested (EXG2, GLI and GCC) delivered robust data over flights performed at different times on the same day and thus under different light conditions (from 75.6 W/m² to 552.7 W/m²). Not surprisingly, these are all normalized indices; the non-normalized ExG1 investigated did not pass the stability test and was one of the VIs with the lowest performance. This is in agreement with the results reported by Woebbecke et al. (1995) who tested the use of non-normalized and normalized RGB VIs in shaded and unshaded surfaces and concluded that the normalized ones were less variable than the non-normalized ones. The effect of the change in illumination over time on information derived from on-ground digital imagery in a forest scenario was studied by Sonnentag et al. (2012). They concluded that it was not possible to remove the effect entirely on ExG1 and GCC. The effect of spatial and temporal factors on the classification accuracy of vegetation in wheat fields using VIs from UAV imagery was also investigated by Torres-Sánchez et al. (2014), who reported that ExG2 and VEG were the most suitable indices.

Furthermore, the three selected VIs delivered consistent data even when flights were separated by five days and were done under contrasting light environments. The fact of delivering consistent data even after five days is important, as adverse climatological conditions can prevent flying at pre-defined days after mowing. In this experiment, the plants had resumed growth after mowing, but flights executed within a window of five days still delivered consistent VI results and consistent breeding decisions in terms of populations being selected or eliminated. This finding indicates the practical applicability of our procedure to evaluate persistency in realistic situations. Note, however, that the consistency mentioned here refers to relative values, as the ExG2 values were slightly different for the two dates investigated. We cannot determine whether this was due to plant growth or incomplete correction for the light environment despite the application of white balance and exposure correction.

4.2 Three UAV-based VIs provide a good estimation of the area covered by green grass

Results presented in section 3.1.2 demonstrate that UAV-based EXG2, GLI and GCC values represent reliable estimates of area covered by green grass ($r^2 = 0.791$), and hence could be used to estimate persistency, as the assessment of this trait is done through visual scores based on vegetation cover. In an attempt to further improve these results, different multiple linear regressions were carried out involving some or all of the indices, but no significant improvement was achieved. This is not surprising, given the linear relationship demonstrated for the range of RGB values considered in this study in pair-wise comparisons of the three VIs. Therefore, a linear regression with any of the three selected indices fit our purposes. As EXG2 is one of the most widely cited and used vegetation indices (Guijarro et al., 2015; Mao et al., 2003; Meyer and Neto, 2008; Sonnentag et al., 2012; Torres-Sánchez et al., 2014) we based further developments and testing of the procedure on this index. Results for GLI or GCC are expected to be similar.

4.3 The visual scoring system (UAV-based VIs) consistency compared to the visual scoring by breeders

All pair correlations between the four ExG2 datasets were ≥ 0.871 (Table 4), meaning that similar and consistent information were delivered by the repeated flights. This creates room for flexibility when implementing the methodology.

In this experiment, two experienced breeders performed the visual scoring of the plots. The assignment of visual scores was done according the current practice in the perennial ryegrass breeding program at ILVO. Both breeders involved in this experiment were trained to assign scores of persistency; their training is reflected in the high 'within-breeder' consistency estimated here ($r = 0.754$ between consecutive scorings by the same breeder). On the other hand, it was found that they probably use slightly different criteria for visual evaluation, as the between-breeder correlation was low ($r = 0.243$). When interviewing the breeders about their criteria, clear differences emerged. Breeder B scored the plots using a narrower range of values than Breeder A and took the amount of

bare soil present in the plots into account, while Breeder A did not consider this aspect. This was confirmed by the higher correlation between the percentage of bare soil (determined through a supervised classification) and the visual scores for Breeder B ($r = -0.331$; $p < 0.001$) than for Breeder A ($r = -0.025$; non-significant).

Visual scoring has the advantage of being cheap, quick, and easy to perform, but contains a risk of delivering biased information. Furthermore, even experienced breeders within the same organization do not always agree on a consistent system for assigning scores. Indeed, visual scores are affected by the observer: the assessments require experience and a certain level of skill, but a certain level of subjectivity cannot be avoided (Milberg et al., 2008). Studies to estimate the observer bias affecting visual estimations (Hatton et al., 1986; Horst et al., 1984; Kennedy and Addison, 1987) have shown variances from 13% to 90%.

Regardless, our results demonstrate that using UAV-based VIs to estimate persistency would allow a higher level of objectivity and flexibility by eliminating the human influences on the evaluation.

4.4 Implementation of the UAV-based procedure for the evaluation of persistency is a realistic option

Finally, we tested the feasibility of shifting to a system to determine persistency in perennial ryegrass using UAV-based VI data. ExG2 data were compared to scoring data in terms of selection as is done in a real breeding situation. Negative (worst 50%) selection was carried out based on the visual scores and the UAV based vegetation index values. The results of the selection were then compared. The highest levels of agreement between visual scoring and ExG2 data were found for Breeder A with values of 74.6% and 60.9% for FT1 and FT2, respectively. The level of agreement was lower for the visual scorings by Breeder B (54.0% and 58.4%, respectively). We can therefore conclude that the UAV-based approach using ExG2 can reproduce the visual scoring system used by Breeder A for diploid selections quite well, while it was less successful in reproducing the visual scoring system of

Breeder B for tetraploid selections. The latter might be improved by incorporating information on the percentage bare soil, as Breeder B also included this information in his scoring system.

Regarding the objectivity and reproducibility, even when selections based on the two flights with the lowest correlation were compared, levels of agreement of 85.6% for negative selections was found. Consequently, the UAV-based VI procedure presented here would provide the same end result: the same selection of populations, independent of the flight used. Thus, our results clearly demonstrate that an evaluation system based on UAV-based VIs would significantly increase the reproducibility and objectivity of evaluations of persistency in perennial ryegrass.

In summary, we have developed a procedure using a UAV-based VI that provides useful information on the persistency of perennial ryegrass selections sown in plots. This system is ready for implementation in practical breeding. In future research, extra variables should be considered to further refine the system, e.g. the amount of bare soil. Moreover, during the breeding experiment, each generation of selection for all-round performance takes at least three years (Wilkins and Humphreys, 2003). Breeders typically score multiple parameters, not only persistency, to develop improved varieties. Consequently, their decisions to select an accession or not is not only based on persistency, but also on other parameters such as yield. Therefore, it could be interesting to monitor a breeding field trial over the three years to test whether the selection accuracy can be improved or to deliver other extra useful information that can be extracted from UAV imagery.

In conclusion, we have demonstrated that UAV-based vegetation indices can assist breeders to evaluate persistency of *L. perenne* selections at plot level using a UAV equipped with a visual camera. ExG2 was revealed to be a stable normalized vegetation index, it corresponded very well with the true value of persistency, and it allowed assessment of the populations in a way that was consistent with the breeder evaluations based on a negative selection procedure.

529

530 5. ACKNOWLEDGEMENTS

531 The research reported in this article was carried out at ILVO as part of the ISense project (Isense.farm
532 and www.ilvo.vlaanderen.be) without any specific grant from funding agencies in the public,
533 commercial, or not-for-profit sectors. The authors wish to thank Thomas Vanderstocken and Filip De
534 Brouwer for piloting the UAV flights and Miriam Levenson for English language corrections.

535

536 6. REFERENCES

537

538 Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: the new crop breeding frontier.

539 Trends Plant Sci. 19, 52–61. doi:10.1016/j.tplants.2013.09.008

540 Burgon, A., Bondesen, O.B., Verburgt, W.H., Hall, A.G., Bark, N.S., Robinson, M., Timm, G., Fairey, D.T.,

541 Hampton, J.G., 1998. The forage seed trade. Forage seed Prod. Vol. 1 Temp. species. 271–286.

542 Casler, M.D., van Santen, E., 2010. Breeding Objectives in Forages, in: Boller, B., Posselt, U.K.,

543 Veronesi, F. (Eds.), Fodder Crops and Amenity Grasses. Springer New York, New York, NY, pp.

544 115–136. doi:10.1007/978-1-4419-0760-8_5

545 Caturegli, L., Corniglia, M., Gaetani, M., Grossi, N., Magni, S., Migliazzi, M., Angelini, L., Mazzoncini, M.,

546 Silvestri, N., Fontanelli, M., Raffaelli, M., Peruzzi, A., Volterrani, M., 2016. Unmanned aerial

547 vehicle to estimate nitrogen status of turfgrasses. PLoS One 11, 1–13.

548 doi:10.1371/journal.pone.0158268

549 Chaves, B., De Vlieghe, A., Van Waes, J., Carlier, L., Marynissen, B., 2009. Change in agronomic

550 performance of *Lolium perenne* and *Lolium multiflorum* varieties in the past 40 years based on

551 data from Belgian VCU trials. Plant Breed. 128, 680–690. doi:10.1111/j.1439-0523.2009.01621.x

552 F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, A.F.H.G., 1993.

553 The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging
554 Spectrometer Data. Remote Sens. Environ. doi:10.1016/0034-4257(93)90013-N

555 Govaerts, Y.M., Verstraete, M.M., Pinty, B., Gobron, N., 1999. Designing optimal spectral indices: A
556 feasibility and proof of concept study. Int. J. Remote Sens. 20, 1853–1873.
557 doi:10.1080/014311699212524

558 Guerrero, J.M., Pajares, G., Montalvo, M., Romeo, J., Guijarro, M., 2012. Support Vector Machines for
559 crop/weeds identification in maize fields. Expert Syst. Appl. 39, 11149–11155.
560 doi:10.1016/j.eswa.2012.03.040

561 Guijarro, M., Riomoros, I., Pajares, G., Zitinski, P., 2015. Discrete wavelets transform for improving
562 greenness image segmentation in agricultural images. Comput. Electron. Agric. 118, 396–407.
563 doi:10.1016/j.compag.2015.09.011

564 Hatton, T.J., West, N.E., Johnson, P.S., 1986. Relationships of the Error Associated with Ocular
565 Estimation and Actual Total Cover. J. Range Manag. 39, 91–92. doi:Doi 10.2307/3899697

566 Herrmann, D., Boller, B., Studer, B., Widmer, F., Kölliker, R., 2008. Improving persistence in red clover:
567 Insights from QTL analysis and comparative phenotypic evaluation. Crop Sci. 48, 269–277.
568 doi:10.2135/cropsci2007.03.0143

569 Horst, G.L., Engelke, M.C., Meyers, W., 1984. Assessment of Visual Evaluation Techniques1. Agron. J.
570 76, 619–622. doi:10.2134/agronj1984.00021962007600040027x

571 Humphreys, M., Feuerstein, U., Vandewalle, M., Baert, J., 2010. Ryegrasses, in: Boller, B., Posselt, U.K.,
572 Veronesi, F. (Eds.), Fodder Crops and Amenity Grasses. Springer New York, New York, NY, pp.
573 211–260. doi:10.1007/978-1-4419-0760-8_10

574 Kennedy, K. a., Addison, P. a., 1987. Some Considerations for the Use of Visual Estimates of Plant
575 Cover in Biomonitoring. J. Ecol. 75, 151–157.

576 Liu, Y., Mu, X., Wang, H., Yan, G., 2012. A novel method for extracting green fractional vegetation
 577 cover from digital images. *J. Veg. Sci.* 23, 406–418. doi:10.1111/j.1654-1103.2011.01373.x
 578 Liu, Z.Y., Huang, J.F., Wu, X.H., Dong, Y.P., 2007. Comparison of vegetation indices and red-edge
 579 parameters for estimating grassland cover from canopy reflectance data. *J. Integr. Plant Biol.* 49,
 580 299–306. doi:10.1111/j.1744-7909.2007.00401.x
 581 Lootens, P., Ruttink, T., Rohde, A., Combes, D., Barre, P., Ruiz, I.R., 2016. High - throughput
 582 phenotyping of lateral expansion and regrowth of spaced *Lolium perenne* plants using on - field
 583 image analysis. *Plant Methods* 1–14. doi:10.1186/s13007-016-0132-8
 584 Lusnier, J.D., Thompson, W.L., Wilson, J.M., Gorham, B.E., Dragut, L.D., 2006. Using digital photographs
 585 and object-based image analysis to estimate percent ground cover in vegetation plots. *Front.*
 586 *Ecol. Environ.* 4, 408–413. doi:10.1890/1540-9295(2006)4[408:UDPAOI]2.0.CO;2
 587 Lynch, T.M.H., Barth, S., Dix, P.J., Grogan, D., Grant, J., Grant, O.M., 2015. Ground Cover Assessment of
 588 Perennial Ryegrass Using Digital Imaging. *Agron. J.* 107, 2347. doi:10.2134/agronj15.0185
 589 Mao, W., Wang, Y., Wang, Y., 2003. Real-time detection of between-row weeds using machine vision.
 590 2003 ASAE Annu. Meet. 0300, 1.
 591 Meyer, G.E., Neto, J.C., 2008. Verification of color vegetation indices for automated crop imaging
 592 applications. *Comput. Electron. Agric.* 63, 282–293. doi:10.1016/j.compag.2008.03.009
 593 Milberg, P., Bergstedt, J., Fridman, J., Odell, G., Westerberg, L., 2008. Observer bias and random
 594 variation in vegetation monitoring data. *J. Veg. Sci.* 19, 633–644. doi:10.3170/2008-8-18423
 595 Popay, A.J., Hume, D.E., 2011. Endophytes improve ryegrass persistence by controlling insects. *Pasture*
 596 *Persistence – Grassl. Res. Pract. Ser.* 15, 149–156.
 597 Rasmussen, J., Ntakos, G., Nielsen, J., Svendsgaard, J., Poulsen, R.N., Christensen, S., 2016. Are
 598 vegetation indices derived from consumer-grade cameras mounted on UAVs sufficiently reliable

599 for assessing experimental plots? *Eur. J. Agron.* 74, 75–92. doi:10.1016/j.eja.2015.11.026

600 Richardson, M.D., Karcher, D.E., Purcell, L.C., 2001. Quantifying Turfgrass Cover Using Digital Image
601 Analysis. *Crop Sci.* 41, 1884–1888. doi:10.2135/cropsci2001.1884

602 Salamí, E., Barrado, C., Pastor, E., 2014. UAV flight experiments applied to the remote sensing of
603 vegetated areas. *Remote Sens.* 6, 11051–11081. doi:10.3390/rs61111051

604 Sonnentag, O., Hufkens, K., Teshera-Sterne, C., Young, A.M., Friedl, M., Braswell, B.H., Milliman, T.,
605 O’Keefe, J., Richardson, A.D., 2012. Digital repeat photography for phenological research in
606 forest ecosystems. *Agric. For. Meteorol.* 152, 159–177. doi:10.1016/j.agrformet.2011.09.009

607 Taylor, N.L., Quesenberry, K.H., 1996. Red clover science. Kluwer Academic Publishers, Dordrecht.

608 Torres-Sánchez, J., Peña, J.M., de Castro, A.I., López-Granados, F., 2014. Multi-temporal mapping of
609 the vegetation fraction in early-season wheat fields using images from UAV. *Comput. Electron.*
610 *Agric.* 103, 104–113. doi:10.1016/j.compag.2014.02.009

611 Wilkins, P.W., Humphreys, M.O., 2003. Progress in breeding perennial clovers for temperate
612 agriculture. *J. Agric. Sci.* 140, 129–150. doi:10.1017/S0021859605005101

613 Woebbecke, D.M., Meyer, G.E., Von Bargen, K., Mortensen, D.A., 1995. Color indices for weed
614 identification under various soil, residue, and lighting conditions. *Trans. ASAE* 38, 259–269.
615 doi:10.13031/2013.27838

616 Xiang, H., Tian, L., 2011. Method for automatic georeferencing aerial remote sensing (RS) images from
617 an unmanned aerial vehicle (UAV) platform. *Biosyst. Eng.* 108, 104–113.
618 doi:10.1016/j.biosystemseng.2010.11.003

619

621 Table 1 List of RGB vegetation indices used in this study. R= Red, G= Green and B= Blue bands.

Index	Abbreviation	Reference	Formula
Non-normalized			
Excess Green 1	ExG1	Woebbecke et al., 1995	$ExG(1) = 2 * G - R - B$
Normalized			
Color Index of Vegetation	CIVE	Kataoka et al., 2003	$CIVE = (0,441R - 0,881G + 0,385B + 18,78745) / (R + G + B)$
Excess Green 2	ExG2	Woebbecke et al., 1995	$ExG(2) = (2 * G - R - B) / (R + G + B)$
Excess Red	ExR	Meyer et al., 1998	$ExR = (1,4 * R - G) / (R + G + B)$
Green Lead Index	GLI	Gobron et al., 2000	$GLI = (2 * G - R - B) / (2 * G + R + B)$
Normalized Green Intensity	GCC	Woebbecke et al., 1995	$GCC = G / (R + G + B)$
Norm. Difference Color Index	NGRDI	Gitelson et al., 2002	$NGRDI = (G - R) / (G + R)$
Ratio Color Index	RCI		$RCI = G / R$
Vegetativen	VEG	Hague et al., 2006	$VEG = G / (R^{0,667} * B^{0,333})$
Woebbecke Index	WI	Woebbecke et al., 1995	$WI = (G - B) / (R - G)$
Combination			
Combination 1	COM1	Guijarro et al., 2011	$COM(1) = 0.25ExG(2) + 0.3ExR + 0.33CIVE + 0,12VEG$
Combination 2	COM2	Guerrero et al., 2012	$COM(2) = 0.36ExG(2) + 0.47CIVE + 0,17VEG$
Excess Green minus Exc. Red	ExGR	Neto, 2004	$ExGR = ExG(2) - ExR$

622

623

624 Table 2 Comparison of the stability of the vegetation indices (VI) over the different flights for the 18th of March
625 and the 22nd of March using a subset of 90 plots (N= 90). Pearson correlation (r) and significance level (p), with
626 bold showing $r > 0.9$ and $p < 0.001$. Intercept and slope of a linear regression, with bold values showing an
627 intercept between -0.1 and 0.1, and a slope between 0.9 and 1.1. An * marks the selected vegetation indices,
628 based on the pre-defined criteria .

Index	18 March				22 March				Selection
	r	p	intercept	slope	r	p	intercept	slope	
Non-normalized									
ExG1	0.924	0.001	3.482	0.817	0.722	0.001	16.974	0.580	
Normalized									
CIVE	0.956	0.001	-0.931	1.050	0.917	0.001	-0.849	1.045	
ExG2	0.957	0.001	-0.002	1.050	0.919	0.001	0.000	1.045	*
ExR	0.928	0.001	0.019	0.978	0.859	0.001	-0.007	1.040	
GLI	0.957	0.001	-0.001	1.047	0.920	0.001	0.000	1.040	*
GCC	0.957	0.001	-0.017	1.049	0.919	0.001	-0.015	1.045	*
NGRDI	0.937	0.001	-0.011	0.993	0.872	0.001	0.002	1.041	
RCI	0.938	0.001	0.007	0.971	0.868	0.001	-0.045	1.052	
VEG	0.945	0.001	-0.014	0.776	0.922	0.001	-0.061	1.063	
WI	0.892	0.001	-0.832	1.015	0.600	0.001	2.869	0.364	
Combination									
COM1	0.535	0.001	5.477	0.143	0.735	0.001	-9.966	2.541	
COM2	0.953	0.001	1.089	0.875	0.922	0.001	-0.507	1.056	
ExGR	0.950	0.001	-0.006	1.044	0.898	0.001	0.009	1.050	

629

630

631 Table 3 Example of the results of the supervised classification matrix of an on-ground frame.

632 Classification result	Ground-truth (%)		
	Class 1	Class 2	Class 3
Unclassified	0.00	5.69	0.00
Class 1: 'green grass'	95.85	0.05	4.51
Class 2: 'bare soil'	2.39	78.86	6.92
Class 3: 'old grass'	1.75	15.39	88.57
Overall accuracy: 85.57%			
Kappa coefficient: 0.779			

633 Table 4 Pearson correlation coefficients between the four flights over the two dates (18: 18th of March; 22: 22nd
 634 of March; 1: first flight; 2: second flight) for Excess green 2 (ExG2) (n= 90 plots).

Index	ExG2 18-1	ExG2 18-2	ExG2 22-1
ExG2 18-2	0.957		
ExG2 22-1	0.948	0.930	
ExG2 22-2	0.871	0.876	0.919

635

636 Table 5 Pearson correlation showing consistency for each breeder and agreement between breeders (n=90
637 plots)

	Breeder A	Breeder B
Breeder A	0.754	
Breeder B	0.243	0.760

638

639 Supplementary data 1 Correlations between the value of the index corresponding to the area of the frame in the
640 orthophoto (VI_UAV_Frame) and the value of the index of the area of the plot (VI_UAV_Plot) where the frame
641 was located (ExG2: Excess green 2; GLI: Green lead index; GCC: Normalised green intensity; n=30 plots).

VI_UAV_Frame	VI_UAV_Plot		
	ExG2	GLI	GCC
ExG2	0.850	0.851	0.850
GLI	0.849	0.840	0.849
GCC	0.850	0.851	0.850

642
643

FIGURE LEGENDS

Fig 1. (A) Orthophoto of the breeding trial (set images: morning 18th March); (B) subset of the 90 plots selected for the second breeder score; (C) on-ground frame imagery capture using tripod; and (D) example of one on-ground frame image.

Fig 2. Workflow for data collection, processing and analysis, with indication of the data used in each step.

Fig 3. Linear regression with 95% confidence interval between the selected indices for the first and second flight on the 18th and the 22nd of March (ExG2: Excess green 2 (A,D); GLI: Green lead index (B, E); GCC: Normalized green intensity (C, F) per ploidy (2x-4x); gray regions represent the 95% confidence interval; n=90 plots. Dashed line corresponds to the 1:1 line.

Fig 4. (A) Example of two on-ground frame images; (B) Results of the supervised pixel classification of green grass (green), old grass (yellow) and bare soil (white); frame area from (C) the orthophoto; and (D) ExG2 grid calculation.

Fig 5. Linear regression with 95% confidence interval performed with the index ExG2, GLI and GCC to estimate the area covered by green grass (A-GG).

Fig 6. Theoretical relation between ExG2, GCC and GLI. Dashed lines indicates the range of ExG2 values observed in this study.

Fig 7. Comparison of the negative selection for six cases: (A) Breeder A versus ExG2 for the flight closest to the scoring (n= 102 2x selections); (B) Breeder A versus ExG2 for the last flight (n= 102 2x selections); (C) Breeder B versus ExG2 for the flight closest to the scoring (n= 135 4x selections); (D) Breeder B versus ExG2 for the last flight (n= 135 4x selections); (E) ExG2 for the two flights showing the highest correlation (n=237 selections); and

671 (F) ExG2 for the two flights showing the lowest correlation (n=237 selections). The percentages of agreement in
672 the selections are shown, black dots represent the selections that agree and gray dots the disagreement. The
673 quadrant corresponding to the negative selection is indicated by darker lines corresponding to the median of the
674 axis values represented.

675

676 Fig 8. Comparison negative selection (50% worst selections) in validation field trial FT2: (A) Breeder A versus
677 ExG2 (n= 138 2x selections); (B) Breeder B versus ExG2 (n= 96 4x selections). The percentages of agreement in
678 the selection are shown, black dots represent the selections that agree and gray dots the disagreement. The
679 quadrant corresponding to the negative selection is indicated by darker lines.

Fig3

[Click here to download colour figure Figure3.jpg](#)