A time-integrated multi-sensor approach to characterize senescence dynamics in wheat under field conditions

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

Jonas Anderegg

MSc. ETH in Agricultural Sciences

born on 08.06.1989

citizen of Rapperswil BE

accepted on the recommendation of

PD. Dr. Andreas Hund (examiner)

Prof. Dr. Achim Walter (co-examiner)

Prof. Dr. Urs Schmidhalter (co-examiner)

2020

Tu nous entends, l'Univers, tu nous entends ?

Si tu nous entends, attend-nous, on arrive,

On voudrait tout comprendre, tout savoir, tout voir…

-   FAUVE, BLIZZARD

Table of Contents

Summary

Wheat (*Triticum aestivum*, L.) is the most extensively grown staple food crop in the world. It delivers one fifth of the human dietary calorie and protein intake, which highlights its outstanding importance for global food security. However, farm-level yields have recently stagnated in some major wheat producing regions, probably as a result of increasingly adverse environmental conditions during key developmental stages. As a consequence of global climate change, weather extremes such as heat and drought are predicted to occur with increasing frequency and severity. Wheat breeders are therefore confronted with the difficult task to improve yield and quality under rapidly changing climatic conditions.

Recent genotyping technology is now providing access to detailed genomic information for large numbers of genotypes. Furthermore, reference sequences of the bread wheat genome have recently become available. These developments are expected to boost our understanding of the molecular basis of key agronomic traits and accelerate genetic gain. Yet, under field conditions, the performance of a crop is not only determined by its genotype, but also by the environment and the interactions between the genotype and the environment (G×E). The primary breeding target traits in wheat, grain yield (GY) and grain protein concentration (GPC), are complex traits. They are determined by various processes occurring throughout the growth cycle. The G×E is particularly pronounced for such complex traits, as unpredictable interactions can critically affect crop performance at any growth stage. This hampers selection, as genotype rankings change across environments. A detailed understanding of the physiological basis of G×E enables a more targeted selection based on less complex and more predictable traits. A 'physiological breeding' strategy thus aims at maximizing the number of favorable traits enhancing growth and performance throughout the growing season and contributing to final yield.

One such potentially favorable trait is the so-called 'stay-green' trait, which consists in a prolonged maintenance of green leaf area during the grain filling phase (i.e. delayed senescence). Stay-green extends the period of CO_2 assimilation, which may favor grain filling. However, stay-green can have different genetic and physiological bases. Accordingly, its effects on GY and GPC vary depending on the genetic and environmental context in which it arises.

Recently, advances in sensors and carrier systems have paved the road to high throughput measurements of plant or canopy characteristics under field conditions. However, many of these sensor-based approaches to field phenotyping have not yet been thoroughly validated in a breeding context. This is a challenging task because subtle differences among genotypes need to be detected. In addition, genotypes display contrasting morphology, phenology and canopy structure which strongly affects sensor measurements.

The overall aim of this thesis was the development of non-invasive high throughput phenotyping techniques to quantify green leaf area dynamics during the grain filling phase of wheat. Such methods will facilitate large scale screenings of breeding populations in contrasting environments. This is expected to facilitate the identification of determinants of stay-green and contribute to an improved understanding of its effects on GY and GPC. We further aimed to achieve detection

and quantification of an important foliar disease, *Septoria tritici* blotch (STB), and to distinguish it from physiological senescence, using high throughput methods. This is expected to benefit resistance breeding and improve the understanding of the relationship between green leaf area dynamics and GY and GPC. Finally, we aimed to further characterize the stay-green phase with respect to its functionality. For this, field experiments were carried out in three consecutive years in the ETH field phenotyping platform, using a set of ~330 registered wheat cultivars.

In a first study, we scored stay-green and senescence dynamics visually and measured canopy spectral reflectance throughout grain filling. Visual scorings and spectral reflectance measurements were then compared to identify spectral features best representing visually observed senescence dynamics. We found the three-band plant senescence reflectance index (PSRI) to best approximate visual scorings across genotypes and years. Our results also suggested that cheaper sensors with a lower spectral resolution may be used without a significant loss of information. This will facilitate the transfer of the developed concepts to unmanned aerial vehicles, providing the necessary throughput for large-scale screenings of breeding nurseries.

We noted a strong environment specificity in relationships between spectral reflectance and visual scorings across three years. In particular, the wet season of 2016 differed strongly from the drier seasons of 2017 and 2018. In 2016, there was a major epidemic of STB despite intense control measures. STB causes necrotic foliar lesions, which are easily confounded with physiological senescence. We hypothesized that the environment specificity of relationships were due to the presence of STB, and that such differences should be exploitable to detect and quantify STB and delineate STB from physiological senescence. We tested this hypothesis in a separate experiment, using artificial inoculations. Our results demonstrated the feasibility of STB detection and quantification as well as delineation from physiological senescence. Our approach compared well with previous reports in terms of precision and offers increased robustness in the presence of genotypic diversity and environmental variability.

In a third experiment, we aimed to characterize stay-green more accurately, combining assessments of green leaf area dynamics with canopy temperature (CT) measurements. CT is affected by transpiration rates and water use, and may therefore be linked to photosynthetic rates and assimilation. We were able to derive repeatable temporal trends in CT during the stay-green phase. We found independent temporal trends for greenness indicators and CT and observed genotypic differences in these trends. This suggested that a combination of spectral and thermal sensors may indeed enable a more accurate characterization of stay-green. However, the relevance of such assessments for grain filling and GY remains to be investigated.

The findings presented in this thesis provide solutions to several difficulties in applying sensor-based high throughput phenotyping in a breeding context. We propose a time-integrated multi-sensor approach to characterize stay-green as dynamic traits. We use normalized values of the initial canopy reflectance and canopy temperature shortly after flowering as a baseline. This approach helps to control confounding effects of phenology, canopy structure and morphology as well as strong year-effects.

Zusammenfassung

Weizen (*Triticum aestivum*, L.) ist das weltweit am häufigsten angebaute Grundnahrungsmittel. Er liefert ein Fünftel der menschlichen Kalorien- und Proteinzufuhr, was seine herausragende Bedeutung für die globale Ernährungssicherheit verdeutlicht. In wichtigen Weizenanbaugebieten stagnierten die landwirtschaftlichen Erträge in letzter Zeit jedoch, wahrscheinlich als Folge zunehmend ungünstiger Umweltbedingungen während besonders empfindlicher Entwicklungsstadien. Als Folge des globalen Klimawandels werden Wetterextreme wie Hitze und Trockenheit voraussichtlich gehäuft und mit grösserer Intensität auftreten. Die Weizenzüchter stehen daher vor der schwierigen Aufgabe, Ertrag und Qualität unter sich rasch ändernden klimatischen Bedingungen zu verbessern.

Neueste Genotypisierungstechnologie bietet mittlerweile Zugang zu detaillierten genomischen Informationen für eine grosse Anzahl von Genotypen. Darüber hinaus sind seit kurzem Referenzsequenzen des Brotweizengenoms verfügbar. Diese Entwicklungen dürften das Verständnis der molekularen Basis der wichtigsten agronomischen Merkmale verbessern und den genetischen Fortschritt beschleunigen. Unter Feldbedingungen wird die Leistung einer Kulturpflanze jedoch nicht nur durch deren Genotyp, sondern auch durch die Umwelt und die Wechselwirkungen zwischen Genotyp und Umwelt bestimmt. Die primären züchterischen Zieleigenschaften von Weizen, Kornertrag und Kornproteinkonzentration, sind komplexe Merkmale. Sie werden durch verschiedene Prozesse während des gesamten Wachstumszyklus beeinflusst. Genotyp-Umwelt Interaktionen sind bei solchen komplexen Merkmalen besonders ausgeprägt, da unvorhersehbare Wechselwirkungen die Leistung der Pflanzen in jedem Wachstumsstadium entscheidend beeinflussen können. Dies erschwert die Selektion, da sich die Rangordnung der Genotypen über die verschiedenen Umwelten hinweg ändert. Ein detailliertes Verständnis der physiologischen Basis der Genotyp-Umwelt Wechselwirkungen ermöglicht eine gezieltere Selektion auf der Grundlage weniger komplexer und besser vorhersagbarer Merkmale. Die "physiologische Pflanzenzüchtung" zielt daher darauf ab, die Anzahl vorteilhafter Merkmale, die Wachstum und Leistung während der gesamten Wachstumssaison verbessern und zum Endertrag beitragen, zu maximieren.

Ein solches potenziell vorteilhaftes Merkmal ist das so genannte "stay-green"-Merkmal, das in einer längeren Erhaltung der grünen Blattfläche während der Kornfüllphase (d.h. einer verzögerten Seneszenz) besteht. Stay-green verlängert die Dauer der CO_2-Assimilation, was die Kornfüllung begünstigen kann. Stay-green kann jedoch unterschiedliche genetische und physiologische Grundlagen haben. Dementsprechend variieren auch seine Auswirkungen auf Kornertrag und Kornproteinkonzentration je nach genetischem Kontext und der Umwelt, in der es entsteht.

In jüngster Zeit haben Fortschritte bei Sensoren und Trägersystemen den Weg für Hochdurchsatzmessungen von Pflanzen- und Bestandesmerkmalen unter Feldbedingungen geebnet. Viele dieser sensorbasierten Ansätze zur Feldphänotypisierung sind jedoch noch nicht gründlich in einem züchterischen Kontext validiert worden. Dies ist eine anspruchsvolle Aufgabe,

da subtile Unterschiede zwischen Genotypen erkannt werden müssen. Darüber hinaus weisen Genotypen grosse Unterschiede in Morphologie, Phänologie und Bestandesstruktur auf, was Sensormessungen stark beeinflusst.

Das übergeordnete Ziel dieser Arbeit war die Entwicklung nicht-invasiver Hochdurchsatz-Phänotypisierungsmethoden zur Quantifizierung der Dynamik der grünen Blattfläche während der Kornfüllphase des Weizens. Solche Methoden werden grossangelegte Screenings von Zuchtpopulationen in kontrastierenden Umgebungen für dieses Merkmal erleichtern. Es wird erwartet, dass dies die Identifizierung von Determinanten des Stay-Green erleichtert und zu einem besseren Verständnis seiner Auswirkungen auf Kornertrag und Kornproteinkonzentration beitragen wird. Wir zielten ferner darauf ab, eine wichtige Blattkrankheit, die *Septoria tritici* Blattdürre (STB), mit Hochdurchsatzmethoden erkennen und quantifizieren zu können, und sie von der physiologischen Seneszenz abzugrenzen. Dies wird einerseits der Resistenzzüchtung zugutekommen, andererseits aber auch das Verständnis der Beziehung zwischen der Dynamik der grünen Blattfläche und Kornertrag und Kornproteinkonzentration verbessern. Schliesslich wollten wir die stay-green-Phase hinsichtlich ihrer Funktionalität weiter charakterisieren. Dazu wurden in drei aufeinanderfolgenden Jahren Feldexperimente in der ETH-Feldphänotypisierungsplattform durchgeführt, wobei ein Satz von ~330 registrierten Weizensorten verwendet wurde.

In einer ersten Studie haben wir die Stay-green- und Seneszenz-Dynamik visuell erfasst und die spektrale Reflektanz der Bestände während der gesamten Kornfüllphase gemessen. Die visuellen Bonituren und die Messungen der spektralen Reflektanz wurden dann verglichen, um spektrale Merkmale zu identifizieren, die die visuell beobachtete Seneszenz-Dynamik am besten widergeben. Der aus drei Spektralbanden berechnete *Plant Senescence Reflectance Index* approximierte die visuelle Bonitur am besten über Genotypen und Jahre hinweg. Unsere Ergebnisse legten auch nahe, dass einfachere Sensoren mit einer geringeren spektralen Auflösung ohne einen signifikanten Informationsverlust verwendet werden können. Dies wird die Übertragung der entwickelten Konzepte auf Drohnenplattformen erleichtern, was den notwendigen Durchsatz für grossangelegte Screenings von Zuchtgärten garantiert.

Wir stellten eine starke Umweltabhängigkeit der Beziehungen zwischen spektraler Reflektanz und visuellen Bonituren über die Jahre fest. Insbesondere das regnerische Jahr 2016 unterschied sich diesbezüglich stark von den trockeneren Jahren 2017 und 2018. Im Jahr 2016 kam es trotz intensiver Kontrollmassnahmen zu starkem Befall mit STB. STB verursacht nekrotische Blattläsionen, die leicht mit physiologischer Seneszenz verwechselt werden. Wir stellten daher die Hypothese auf, dass die Umweltabhängigkeit der Beziehungen auf das Vorhandensein von STB zurückzuführen ist, und dass solche Unterschiede ausgenutzt werden können, um STB zu erkennen und zu quantifizieren, sowie STB von physiologischer Seneszenz abzugrenzen. Wir testeten diese Hypothese in einem separaten Experiment mit künstlichen Inokulationen. Unsere Ergebnisse zeigten die Machbarkeit eines STB-Nachweises, einer Quantifizierung der Befallsstärken, sowie einer Abgrenzung zur physiologischen Seneszenz. Unser Ansatz war im Vergleich zu früheren Berichten hinsichtlich der Präzision ebenbürtig, hat aber Vorteile in Bezug auf die Robustheit bei genotypischer Vielfalt und variierenden Umweltbedingungen.

In einem dritten Experiment versuchten wir, stay-green genauer zu charakterisieren, indem wir Bonituren der Blattgrün-Dynamik mit Messungen der Bestandestemperatur kombinierten. Die Bestandestemperatur wird von der Transpirationsrate und dem Wasserverbrauch des Bestandes beeinflusst und kann daher mit Photosyntheseraten und CO_2-Assimilation in Verbindung gebracht werden. Wir konnten wiederholbare zeitliche Trends in der Bestandestemperatur während der Stay-green-Phase ableiten. Wir fanden unabhängige zeitliche Trends für Grünheitsindikatoren und Bestandestemperatur und beobachteten genotypische Unterschiede in diesen Trends. Dies deutete darauf hin, dass eine Kombination von Spektral- und Thermalsensoren tatsächlich eine genauere Charakterisierung von stay-green ermöglichen könnte. Die Relevanz solcher Messungen für die Kornfüllung und den Kornertrag muss jedoch noch untersucht werden.

Die in dieser Arbeit vorgestellten Ergebnisse zeigen Lösungsansätze für mehrere Probleme bei der Anwendung sensorbasierter Hochdurchsatz-Phänotypisierung in einem züchterischen Kontext auf. Wir schlagen einen zeitintegrierten Multisensor-Ansatz vor, um stay-green als eine Kombination dynamischer Merkmale zu charakterisieren. Dabei verwenden wir normalisierte Werte der anfänglichen Bestandesreflektanz oder Bestandestemperatur kurz nach der Blüte als Grundlage. Dieser Ansatz hilft, den Einfluss von Störeffekten wie Phänologie, Bestandesstruktur und -morphologie und variierender Umweltbedingungen zu minimieren.

1 General Introduction

In recent decades, growth in agricultural production has continuously outpaced the expansion of demand for agricultural commodities. This has contributed to improved food security for a large part of the world population despite rapid population growth (Southgate *et al.*, 2010). Bread wheat (*Triticum aestivum* L., 2n = 6x, AABBDD) is the most widely cultivated crop in the world, with a production area of approximately 220 million ha (FAOSTAT, 2019). Wheat production extends as far as 45°S in Argentina and 67°N in Scandinavia and Russia (Shiferaw *et al.*, 2013). The widespread cultivation of wheat illustrates its adaptability, which is a cornerstone of its success (Shewry, 2009). Unprecedented yield growth rates have been achieved in the course of the Green Revolution through a combination of the use of high-yielding, fertilizer-responsive semi-dwarf varieties and improved agricultural production techniques (Southgate *et al.*, 2010).

At present, humanity relies on wheat for roughly one fifth of its dietary calorie and protein intake, which highlights the outstanding importance of this crop for global food security (Shiferaw *et al.*, 2013). However, current gains in wheat yield are insufficient to meet projected demands (Ray *et al.*, 2013; Tilman *et al.*, 2011), and yields are even stagnating in some important wheat exporting regions such as western Europe (Brisson *et al.*, 2010; Finger, 2010). Given the cardinal importance of wheat in the global food system, accelerating yield gains is a matter of increasing urgency and will critically depend on genetic improvement of yield potential and on closing yield gaps (Reynolds *et al.*, 2012).

The objective of this general introduction is to discuss a potential avenue to achieving sustained yield growth in wheat – *physiological breeding* – as well as a tool that is essential to its implementation – *phenotyping*. Finally, a possible specific implementation of such a physiological breeding strategy, which this doctoral thesis aims to contribute to, is discussed.

1.1 Physiological wheat breeding

Traditionally, wheat breeding has had a strong focus on selection for genetically simple traits and grain yield *per se* (Araus *et al.*, 2008; Reynolds and Langridge, 2016). Yield gains have been achieved primarily through unspecified recombination of genes among elite germplasm and through introgression programs, most often focused on disease resistance and grain quality (Araus *et al.*, 2008; Jackson *et al.*, 1996; Reynolds and Langridge, 2016). Retrospective studies (*reviewed by* Araus *et al.*, 2008) reveal that breeders have achieved yield increases primarily by optimizing flowering time, reducing plant height, increasing harvest index and increasing the number of grains per area. With the exception of the number of grains per area, these traits are largely optimized. For example, harvest index (HI) appears to be close to its biological maximum (Austin, 1980) and there has been no significant progress in HI since the 1980s (Miralles and Slafer, 2007). Grain yield *per se* is a highly quantitative trait under complex genetic and environmental control. Large genotype-by-environment interactions reduce heritability and, consequently, response to selection.

An improved understanding of genotype-by-environment interactions is urgently needed to support breeding decisions (Hund *et al*., 2019; Reynolds *et al*., 2012). Grain yield is ultimately determined by a multitude of processes with cumulative effects. The ensemble of these processes occurring throughout the growing season determines grain yield by defining three essential parameters: light interception efficiency, radiation use efficiency (i.e. the efficiency to convert intercepted light to biomass) and partitioning efficiency (i.e. the efficiency to convert biomass to grain yield; Hund *et al*., 2019; Reynolds *et al*., 2009).

An improved understanding of how physiological traits (hereafter referred to as secondary traits) affect these key parameters and how they contribute to yield and quality (hereafter referred to as primary traits) in contrasting but well-defined environments will be essential to increase yield gains (Araus *et al*., 2008; Jackson *et al*., 1996; Reynolds *et al*., 2009; Reynolds and Langridge, 2016). This will allow to complement traditional approaches with strategic crossings to accumulate as many beneficial secondary traits as possible and to improve early generation selection based on the assessment of such secondary traits (Reynolds *et al*., 2009). Secondary traits are most useful if (i) the population to select from is variable, (ii) trait heritability is high, (iii) genetic correlation with the primary trait of interest is strong and (iv) assessment costs are low (Jackson *et al*., 1996; van Eeuwijk *et al*., 2018).

1.2   High Throughput Field Phenotyping

A full implementation of physiological breeding strategies relies on a detailed, quantitative description of plant performance throughout the growing season in appropriate environments representing the conditions of the targeted cropping systems. The collection of the required data is referred to as 'phenotyping' and may encompass a characterization of plant organs, entire plants and/or plant stands in terms of their structural, biochemical and physiological properties (Fiorani and Schurr, 2013; Furbank and Tester, 2011; Walter *et al*., 2015). A high temporal resolution of such assessments in combination with a detailed characterization of the environment (sometimes referred to as 'envirotyping'; Bernardo, 2016; Hund *et al*., 2019; Xu, 2016) is increasingly enabling the characterization of dynamic processes such as growth or maturation in terms of their response to short term variation in environmental conditions (*e.g.*, Grieder *et al*., 2015; Kronenberg *et al*., 2019, 2017). The larger the breeding population screened in this manner, the greater the probability of identifying superior alleles. On a more basic notion, such investigations are equally key to the identification or evaluation of new candidate secondary traits and their effects on primary traits under contrasting conditions. This enables a better conceptualization of genotypes combining as many physiological and morphological traits that optimize its performance in particular environments as possible (i.e. the definition of "ideotypes"; Martre *et al*., 2015). Such ideotypes are at the basis of selection programs (Hund *et al*., 2019; Jackson *et al*., 1996). The need to achieve a full-season, detailed characterization of as many genotypes as possible at a high temporal resolution demonstrates the need to complement classical approaches (e.g. a visual scoring by the breeder) with automated, high throughput phenotyping tools (Araus and Cairns, 2014; Furbank and Tester, 2011; van Ginkel *et al*., 2008).

Thanks to progress in sensor technology and carrier systems in recent decades, the theoretical potential of phenotyping has increased far beyond the capabilities of the human eye. For example, thermal imaging allows fast assessments of canopy temperature in large breeding experiments (Deery *et al.*, 2019, 2016; Perich *et al.*, 2020). This may facilitate the identification of drought-resistant genotypes, because leaves of genotypes avoiding dehydration are cooled by transpiration. Multi- or hyperspectral sensors may provide valuable information about biochemical or physiological properties of individual leaves or crop canopies, such as chlorophyll or nitrogen (N) concentration, water status, biomass, and even photosynthetic efficiency (*e.g.*, Becker and Schmidhalter, 2017; Bendig *et al.*, 2015; Gamon *et al.*, 1992; Yu *et al.*, 2014).

Sensor-based high throughput phenotyping and automated data analysis provide access to traits that are theoretically quantifiable by eye but only at prohibitive costs. For example, emergence rates and biophysical properties such as canopy cover, tiller number or number of ears can be assessed based on automated analysis of RGB images (Grieder *et al.*, 2015; T. Liu *et al.*, 2016; Sadeghi-Tehran *et al.*, 2019, 2017; Yu *et al.*, 2017). Similarly, canopy height and its development can be assessed using light detection and ranging (LiDAR) technology (Friedli *et al.*, 2016; Jimenez-Berni *et al.*, 2018; Kronenberg *et al.*, 2017) or unmanned aerial vehicles (UAV) - based imagery in combination with structure from motion photogrammetric methods (Hund *et al.*, 2019; Szeliski, 2011).

Despite the potential outlined above, efficient implementation of high throughput phenotyping in breeding programs has been limited so far (Araus *et al.*, 2018). For example, difficulties remain in connecting sensor measurements carried out at canopy level to biochemical properties or the physiological status of plants. This is primarily due to the complex architecture of crop stands and contrasting morphological properties of genotypes which strongly affect the reflection of the incoming radiation without being stably related to the physiological or biochemical trait of interest. In addition, genotypic variation in target traits is often relatively small within elite germplasm, particularly for biochemical traits. Finally, natural factors such as atmospheric and illumination conditions affect sensor measurements and may change drastically across sites, years, consecutive measurement time points or even during a single measurement campaign. In this respect, previous research has sometimes significantly overestimated the technological readiness of sensor-based phenotyping approaches for applications in a breeding context. Specifically, variance in target traits has often been artificially increased through varying management practices (e.g. fertilization treatments, irrigation treatments, variation in sowing dates and/or densities) and by comparing sensor and ground truth data across various growth stages. Often, this has been paired with a limitation of confounding factors such as genotypic diversity (i.e. morphology, canopy structure and phenology) and environmental variability (i.e. single-year and possibly even single-site experiments). This combination of manipulations results in unrealistically high signal-to-noise ratios, greatly simplifying the task of stable trait prediction. Such conditions do not represent the reality of breeding experiments and their relevance in this respect should therefore be evaluated with caution (*see e.g.,* Øvergaard *et al.*, 2013a, 2013b *for an illustration of such issues*). For example, with current annual yield gains, grain yield differences of about 0.05t/ha must be reliably detected to identify superior genotypes. This task

has to be performed on genotypes with contrasting morphological and canopy structural properties, grown under different environmental conditions. As a possible solution, a combination of sensors is anticipated to provide information of various canopy structural and morphological traits, allowing to statistically correct for variation in confounding factors (*e.g.*, Deery *et al*., 2019; Rebetzke *et al*., 2016).

1.3   Stay-green, green leaf area dynamics and senescence

As described above, grain yield can be understood as a function of light interception efficiency, radiation use efficiency and partitioning efficiency. A delayed senescence (i.e. the stay-green phenotype; Thomas and Smart, 1993) increases light interception by extending the growing season when compared to early-senescing genotypes. Under the basic scenario of all three efficiencies having constant additive effects on grain yield, the stay-green phenotype is therefore expected to increase yield. Positive correlations between stay-green properties and yield are indeed common in many crops (*reviewed by* Gregersen *et al*., 2013 *and* Thomas and Smart, 1993).

In wheat, the situation appears to be complex and contrasting associations between stay-green and grain yield have been reported. In addition, wheat breeders aim to improve not only GY but also GPC, which is a key quality parameter of major economic importance. Simultaneous progress in breeding for these two traits is hampered by the negative genetic relationship between them (Levi *et al*., 2017; Simmonds, 1995; Slafer *et al*., 1990), which illustrates the interconnection between carbon and N metabolism at the canopy level (Bogard *et al*., 2011). One key junction between carbon and N metabolism is senescence (Bogard *et al*., 2010, 2011), because the onset of senescence marks the basic transition of canopies from carbon assimilation to N remobilization (Thomas and Ougham, 2014). The timing of the onset is thus expected to affect both GY and GPC, and is likely to contribute to the observed negative relationship between these traits.

The term stay-green is not always used in the same way. For example, in sorghum Borrell *et al*. (2000) and Jordan *et al*. (2012) refer to genotypes as stay-green if they maintain green leaf area up until and beyond physiological maturity. In this case, stay-green does not explicitly denominate genotypes with an extended post-anthesis duration. In contrast, in wheat, the term is most often used to refer to genotypes with an increased integral of some curve describing the size of the green leaf area over chronological or thermal time. This increase can arise either from a delayed onset of senescence, from a decreased rate of senescence, from a higher greenness at maturity or a combination of any of these phenomena (Thomas and Smart, 1993; *see* Christopher *et al*., 2014; Lopes and Reynolds, 2012; Xie *et al*., 2016 *for examples*). For simplicity, the term stay-green will be used in Chapter 1 to denominate any phenotype with an increased integral under the green leaf area curve. In the description of our own work (i.e. in Chapters 2-5), we will refer to *stay-green* genotypes as genotypes with a delayed onset of senescence (i.e. with an extended period between heading and the onset of senescence in thermal time). The term *senescence dynamics* will be used to summarize the timing and the rate of senescence. The term *green leaf area dynamics* (GLAD) will be used to refer to the evolution of green leaf area after anthesis as a function of physiological senescence and other phenomena affecting green leaf area, such as foliar diseases.

1.3.1 Effects of stay-green on grain yield

Potential yield in wheat is largely determined during the vegetative and reproductive growth stages. The number of spikes per plant, the number of spikelets per spike and the number of grains per spikelet are determined before anthesis (*reviewed by* Distelfeld *et al*., 2014 *and* Slafer *et al*., 2015). In contrast, individual grain weight may be interpreted as the degree of realization of this potential during the grain filling phase, depending on post-anthesis source strength and environmental conditions. Source strength is a function of post-anthesis assimilate production and remobilization of assimilates stored prior to anthesis (Savin and Slafer, 1991). Flag leaf photosynthesis in wheat contributes about 30-50% of assimilates for grain filling (Sylvester-Bradley *et al*., 1990). Stay-green is therefore expected to increase grain yield by favoring post-anthesis assimilate production through an extended period of photosynthetic activity (*e.g.*, Kipp *et al*., 2014; Spano *et al*., 2003). This rather simple model appears to hold true for some important cereal crops such as maize and sorghum. In these crops, a prolonged post-anthesis green leaf area duration confers stable yield advantages, particularly under drought (Borrell *et al*., 2000; Jordan *et al*., 2012; Rajcan and Tollenaar, 1999a).

In wheat, there is ample evidence to suggest that source capacity is adequate to allow for complete grain filling under a wide range of environmental conditions, meaning that yields are mainly limited by sink strength (*see* Araus *et al*., 2008 *and* Borrás *et al*., 2004 *for a review*). This has been demonstrated in a number of experiments through manipulations of the sink-source relationship by seed removal or defoliation (*e.g.*, Cartelle *et al*., 2006; Richards, 1996), through post-anthesis canopy shading (*e.g.*, Savin and Slafer, 1991) or through chemical leaf desiccation (*e.g.*, Blum *et al*., 1983).

At least three distinct characteristics of wheat may explain the strong sink limitation of yields (Borrás *et al*., 2004): (i) an outstanding remobilization efficiency enables buffering of short-comings in concurrent assimilate supply, (ii) potential grain weight is determined well before final seed dry weight is reached, and (iii) radiation levels during sink size determination strongly contrast with radiation levels during the grain filling period. Whereas (i) may be related to the timing and dynamics of senescence (*see e.g.,* Gaju *et al*., 2014; Yang and Zhang, 2006), (ii) and (iii) are factors operating independently of senescence.

Though post-anthesis processes may affect individual grain weight, it is important to consider that differences in grain size and weight at physiological maturity are already evidenced at early stages in the grain filling phase (Borrás *et al*., 2004). Coat layer formation and endosperm cell division cease well before final grain weight is achieved (Millet and Pinthus, 1984). Similarly, the number of starch granules, determined around the onset of rapid grain growth, correlates well with grain weight at maturity (Borrás *et al*., 2004; Brocklehurst, 1977). Thus, it appears that maximum attainable yield is defined relatively early in the grain filling phase.

Despite the strong evidence for sink limitation of wheat yields, there are numerous reports of the stay-green phenotype correlating positively with yield. For example, Spano *et al*. (2003) found a positive association between functional stay-green, individual grain weight and final yield, when comparing four different stay-green ethylmethane sulphonate mutant lines with the control in a

glasshouse experiment. Verma *et al*. (2004) assessed a population of 48 double haploid wheat lines derived from a cross between an early senescing and a stay-green genotype in field experiments and found a positive correlation between the percentage of green flag leaf area remaining at 35 days after anthesis and final grain yield, though individual grain weight was not assessed. Kumari *et al*. (2007) found the stay-green phenotype to be positively associated with biomass, grain filling duration and yield under heat stress, but no correlation with grain weight was observed. Bogard *et al*. (2011) evaluated a double haploid population in multiple year-locations under two contrasting N fertilization regimes. Although the duration of the stay-green phase explained up to 50% of the environmental variation in grain yield, only weak phenotypic correlations were found within environments. Furthermore, anthesis date mostly explained the observed correlations, probably through an effect on the number of grains per area (Bogard *et al*., 2011). Christopher *et al*. (2008) compared a wheat line with significant yield advantage in Australian multi-environment screening trials with a locally adapted genotype and concluded that yield advantages were closely associated with its stay-green phenotype, likely a result of access to deeper soil water (Christopher *et al*., 2008, 2014). In this case, higher average individual grain weight was an important factor in determining higher yield (Christopher *et al*., 2008). This was subsequently confirmed in a double haploid population derived from a cross of the investigated genotypes (Christopher *et al*., 2016).

Other studies have reported negative correlations between stay-green duration and grain yield. For example, Kipp *et al*. (2014) found a strong negative correlation for fifty winter wheat cultivars grown under high-yielding conditions in a single year-location. Jiang *et al*. (2004) observed a negative correlation in a doubled haploid rice population, which probably resulted from the negative correlation between stay-green and seed-setting rate. Gong *et al*. (2005) reported that the strong vigor of a wheat hybrid resulted in the maintenance of green leaf area until physiological maturity, resulting in much unused carbon reserve in straws and a decreased harvest index. Naruoka *et al*. (2012) observed negative or no correlations of stay-green with GY in rain-fed and irrigated environments for a recombinant inbred line population derived from a cross between a stay-green and an intermediate spring wheat cultivar.

These contrasting findings highlight the complexity of the relationship between stay-green and grain yield in wheat. It appears likely that the genetic and physiological determinants of stay-green vary widely among studies. For example, the findings by Jiang *et al*. (2004) strongly suggest that stay-green was an emerging consequence of a reduced sink demand. In contrast, the stay-green phenotype described by Christopher *et al*. (2008) and Reynolds *et al*. (2012) is interpreted as the avoidance of drought- or heat-stress-induced premature senescence. In this context, it appears to be unclear whether stay-green affects GY by relieving temporal constraints to grain filling or whether effects are determined by modifications of the sink-source balance. Several studies suggest that an increase in post-anthesis temperature shortens grain filling duration without adequate compensations through increased grain filling rates (García *et al*., 2016; Lizana and Calderini, 2013). This negatively affects final grain weight even in the absence of any indication of source limitation (García *et al*., 2016; Yang and Zhang, 2006). Investigations on the effect of senescence dynamics on grain filling rates produced contradictory results, probably due

to contrasting environmental conditions, stress timing and yield potential (Gong *et al*., 2005; Xie *et al*., 2016; Yang and Zhang, 2006).

1.3.2 Effects of stay-green on grain protein concentration

In wheat, translocation of leaf N to developing grains accounts for 40-90% of final grain N (Kichey *et al*., 2007), highlighting the crucial importance of translocation processes for protein yield formation. Given the tight interconnection between senescence and resource remobilization from vegetative organs to the developing sink (Yang and Zhang, 2006), the importance of senescence in determining GPC is therefore obvious.

The negative relationship between grain yield and GPC is often attributed to a dilution of proteins through increased starch accumulation and/or to a competition between carbon and N for energy (Acreche and Slafer, 2009; Munier-Jolain and Salon, 2005). Thus, depending on post-anthesis N availability, stay-green may affect GPC in different ways (Acreche and Slafer, 2009; Bogard *et al*., 2010, 2011; Munier-Jolain and Salon, 2005):

(i) Under a high post-anthesis N availability, stay-green genotypes are expected to take up additional N, resulting in a better overall N supply and, *via* competition for energy in storage compound synthesis, decrease starch content in grains;

(ii) Under a low post-anthesis N availability, stay-green genotypes are expected to maintain carbon assimilation but not N uptake, which will result in a dilution of grain protein by an increased starch synthesis.

However, in this context, it may not always be clear whether the stay-green phenotype causes additional N uptake (*e.g.*, Hirel *et al*., 2007), or whether the stay-green phenotype is merely an emerging consequence of a sustained post-anthesis N uptake (Yang and Zhang, 2006).

1.3.3 Basic considerations on the regulation of senescence

Though senescence is under complex genetic and environmental control (Borrill *et al*., 2019; Lim *et al*., 2007), it is well known that the balance between N supply and demand is an important determinant of its timing and dynamics (Kichey *et al*., 2007; Rajcan and Tollenaar, 1999; Triboi and Triboi-Blondel, 2002; van Oosterom *et al*., 2010). If grain N demand exceeds soil N uptake, for example due to water or N deficiency, N translocation from stems and leaves is accelerated, resulting in an anticipation and acceleration of senescence ('self-destruction' hypothesis; Kichey *et al*., 2007; Sinclair and Wit, 1975) as well as an increased N remobilization efficiency (Gaju *et al*., 2014). Conversely, late fertilizer applications increase post-anthesis N uptake from the soil, resulting in delayed and slower senescence (Triboi and Triboi-Blondel, 2002; Yang and Zhang, 2006). Thus, post-anthesis N availability is an important factor determining the dynamics of senescence.

The relationship between demand and supply of assimilates is also known as a major factor in the regulation of senescence (Wingler *et al*., 2009). In maize, a low assimilation potential of the source accelerates senescence in order to meet sink demand through remobilization, in analogy

to N-mediated regulation of senescence (Kumar *et al.*, 2019; Lee and Tollenaar, 2007). In contrast to the effects described above for N, it appears that a low sink demand for sugars accelerates senescence. This acceleration seems to be source-mediated (as opposed to sink-mediated in the case of N), and strongly related to the accumulation of soluble carbohydrates in leaves beyond a certain threshold (Kumar *et al.*, 2019; Wingler and Roitsch, 2008). In wheat, vertical patterns of senescence have been associated with corresponding sugar levels in leaves (Shi *et al.*, 2016), suggesting a significant role of sugar source-sink relationships on the regulation of senescence.

Bogard *et al.* (2011) found genetic variation in senescence dynamics to be strongly related to genetic variation in flowering time. This could be due to direct effects of flowering, for example metabolic changes such as starch accumulation in leaves, which acts as a strong intrinsic senescence-promoting signal (Bogard *et al.*, 2011; Kumar *et al.*, 2019). However, it appears more likely that flowering time affects senescence primarily through genotype-by-environment interactions. Early flowering genotypes are likely to perceive a less stressful environment during grain filling, and benefit from increased soil N availability, partly related to soil water availability and partly due to soil N uptake partitioning between stages (Bogard *et al.*, 2010, 2011).

Although senescence is a complex process involving the up-regulation of hundreds of genes (Borrill *et al.*, 2019), single genes can still have a major direct effect on senescence and remobilization processes. A well-described example is the NAM-B1 gene, present mainly in ancestral wheat, which encodes a NAC-domain transcription factor (Uauy *et al.*, 2006). High expression levels result in a hastening of senescence, an increase in remobilization of zinc, iron and N and an increased GPC (Uauy *et al.*, 2006).

As outlined above, different internal and external factors can trigger senescence and contribute to modifications of its dynamics. However, it appears that once the senescence program is initiated, common pathways are activated, irrespective of the trigger (Guo and Gan, 2012). Also, an orderly process of senescence is largely maintained, even if senescence is prevalently stress-triggered (Guo and Gan, 2012; Lim *et al.*, 2007). In this context, it is important to note that factors influencing GLAD after anthesis but not the dynamics of physiological senescence may be related to GY and GPC in a different way than physiological senescence. Specifically, epidemics of foliar diseases reduce green leaf area similarly to premature senescence, and therefore reduce source capacity in a similar way (*see e.g.,* Bingham *et al.*, 2019). However, in contrast to other stresses, foliar diseases may more strongly interfere with whole-plant functioning and an orderly process of senescence. Physiologists generally disregard the effect of diseases. However, from a practical viewpoint, diseases and physiological senescence are likely to co-occur in temperate, high-yielding environments. In the special case of the Swiss federal wheat breeding program of Agroscope, genotypes are evaluated without fungicide applications. Thus, a distinction between different factors influencing GLAD may be important to better understand the context-specific relationships between GLAD and GY and GPC.

1.3.4 Stay-green - a secondary trait for physiological wheat breeding using high throughput phenotyping?

As described above, the relationship between stay-green and primary traits appears to be more complex in wheat than in other crops, which has been attributed to the contrasting patterns in the sink-source relationship among crops. Effect sizes appear to depend on the genetic and physiological bases of stay-green and on environmental conditions (*e.g.*, Christopher *et al*., 2018). The complexity of the regulation of stay-green often results in low to intermediate heritability across environments (*e.g.*, Crain *et al*., 2017; Lopes and Reynolds, 2012), which may discourage the use as a secondary trait (Jackson *et al*., 1996). Finally, the well-documented prevalent source-limitation of wheat yields might challenge a focus on maximizing the source capacity of wheat during grain filling.

However, there is increasing evidence for stay-green to be advantageous particularly under heat and drought or heat combined with drought environments. Among the best investigated source of stay-green is the CIMMYT line SeriM82 (Christopher *et al*., 2008, 2014, 2016, 2018; Olivares-Villegas *et al*., 2007). Here, a 1d delay in senescence onset conferred a 3.2% yield advantage (Christopher *et al*., 2016). Identified stay-green QTLs were highly heritable in certain environments (Christopher *et al*., 2018).

As a result of climate change, heat and drought events are projected to increase in frequency and severity, even in high-yielding environments mostly unaffected by such stresses so far (Holzkämper *et al*., 2015; Lehner *et al*., 2006; Trnka *et al*., 2015). Further adjustments in flowering time to avoid late season heat and drought appear to be increasingly in conflict with the need to safeguard potential yield formation during the vegetative and reproductive growth stages (Slafer *et al*., 2015). In Europe, modern cultivars require a 14-18% lower temperature sum to reach flowering than cultivars grown in the 1950s and 1960s, due in equal parts to rises in temperature and cultivar changes (Rezaei *et al*., 2018). In contrast, increases in total biomass and thus, sufficient time for plant and sink growth, are absolutely required to maintain a continued increase in grain yield (Araus *et al*., 2008; Reynolds *et al*., 2012; Slafer *et al*., 2015). It is therefore imperative to achieve genetic progress in plant performance under heat and drought stress.

1.4 Aims and structure of the thesis

As described above, it may become increasingly infeasible to escape late-season heat and drought periods by reducing time to anthesis and physiological maturity without negative effects on biomass acquisition and potential yield formation during the vegetative and reproductive stages. Consequently, developmental adaptation will increasingly have to be complemented with genetic progress in plant performance under heat and drought stress. Under such conditions, a prolonged maintenance of green leaf area after anthesis has been a major breeding aim in several crops, particularly in maize and sorghum. In wheat, a comprehensive understanding of the physiological and genetic determinants of stay-green as well as of its effects on key economic traits such as grain yield, grain protein content and N use efficiency under varying environmental conditions is

lacking. Contrasting results reported in the literature reflect varying environmental conditions, management practices and genetic material examined in different studies.

The overarching goal of this thesis was the development of remote sensing based methods for the characterization of late development in genetically diverse wheat germplasm at the canopy level. Special attention was given to the assessment of GLAD after anthesis and to the characterization of stay-green regarding its functionality. The availability of such methods is expected to facilitate (i) the characterization of genetic variability existing in breeding programs, (ii) estimation of effect sizes in contrasting environments and the subsequent definition of environment-specific ideotypes, (iii) indirect selection in early breeding generations, when yield cannot yet be accurately determined and (iv) an improved understanding of the genetic basis of the investigated traits. A central aspect of the following three chapters is the evaluation of potential ways to cope with morphological, canopy structural and phenological diversity in examined plant material as well as with contrasting environmental conditions across experimental years. Remote sensing based approaches must be robust to variation in such parameters, as they represent a fundamental element of any breeding experiment.

All experiments were carried out in a temperate, high-yielding environment, with average bread wheat yields of 6 t/ha (Hategekimana *et al*., 2012; Swissgranum, 2019). The crop was sown with a sowing density of 400 plants m^{-2}, resulting in dense canopies with a leaf area index of about 6 $m^2 m^{-2}$ at flowering (unpublished data). N fertilizer is applied according to the amount expected to be removed with the harvested product and taking the N fertilization status of the soil at sowing into account. This typically results in N fertilization of 120 - 140 kg/ha (Flisch *et al*., 2009), which is normally applied in two to four split-applications.

Chapter 2: Assessment of stay-green and the dynamics of senescence in wheat canopies
In order to capture variation in stay-green and the dynamics of senescence, experiments need to be monitored with a high temporal resolution. The potential of high throughput reflectance-based techniques to assess these dynamics has been demonstrated previously with a focus on drought- and heat-stress environments (Christopher *et al*., 2014, 2016; Kipp *et al*., 2014; Lopes and Reynolds, 2012; Montazeaud *et al*., 2016). However, an optimization of spectral indicators was not performed and investigations under high-yielding conditions are rare and limited to a single year (Kipp *et al*., 2014).

In chapter two, we tested different approaches to infer visually observed senescence dynamics from repeated hyperspectral reflectance measurements in three consecutive years. Furthermore, we aimed to estimate the potential of hyperspectral measurements to quantify the process of senescence beyond a simple representation of visually observable greenness decay over time.

Chapter 3: Disentangling the effects of foliar diseases and physiological senescence on spectral reflectance
In chapter three, we aimed to develop a method enabling the identification and quantification of foliar diseases based on canopy reflectance measurements. High throughput methods to detect major foliar diseases are expected to greatly facilitate resistance breeding and are thus important in their own right. Furthermore, given that both foliar diseases and physiological senescence affect

GLAD, a distinction between these phenomena is required to improve our understanding of how GLAD affect primary breeding target traits. We chose *Septoria tritici* blotch (STB) caused by the fungal pathogen *Zymoseptoria tritici* as a model, as it represents a major threat to wheat production in many important wheat growing areas around the world, including Switzerland (Orton *et al*., 2011; Torriani *et al*., 2015). Furthermore, STB epidemics frequently reach damaging levels and affect crop performance most during the grain filling phase (Bancal *et al*., 2007), apparently without modifying the dynamics of physiological senescence (Bancal *et al*., 2016).

Chapter 4: Combining spectral and thermal measurements to quantify performance during the stay-green phase and resistance to heat and drought

As outlined above and in chapter two, GLAD can be tracked using spectral reflectance measurements, offering the potential of upscaling to large breeding nurseries. The stay-green phenotype may result from various combinations of traits and environments. Some of these, for example a low sink demand resulting from heat stress damage in critical stages, are undesirable. Spectral indicators are limited in their potential to assess plant performance indicators such as transpiration or photosynthetic rates. However, wheat plants may respond to a high sink demand by increasing stomatal conductance (Richards, 1996). Therefore, it has been hypothesized that genotypes showing functional stay-green may be identified using a combination of sensors (Rebetzke *et al*., 2013).

In chapter four, we used repeated measurements of canopy temperature to evaluate whether temporal trends in canopy temperature could be used for the identification of genotypes maintaining high transpiration and possibly photosynthetic rates during their stay-green phase. The applicability of thermal imaging to screen for drought resistance still has to be evaluated in environments with high in-season precipitation levels. For this, we evaluated the repeatability of temporal trends in CT and their correlation with known confounding factors. Furthermore, we investigated whether genotype-specific contrasts in temporal trends between CT and spectral reflectance could be expected. Such contrasting trends may facilitate the identification of stay-green genotypes maintaining high levels of transpiration and thus potentially high photosynthetic rates. Alternatively, they may help to identify dysfunctional or partly functional stay-green genotypes as genotypes maintaining a large green leaf area but not high transpiration rates.

2   Spectral Vegetation Indices to Track Senescence Dynamics in Diverse Wheat Germplasm

Jonas Anderegg[1,*], Kang Yu[1,2], Helge Aasen[1], Achim Walter[1], Frank Liebisch[1], Andreas Hund[1]

[1]Crop Science Group, Institute of Agricultural Sciences, ETH Zurich, Zurich, Switzerland

[2]Division of Forest, Nature and Landscape, Department of Earth and Environmental Sciences, 3001 Leuven, Belgium

*Corresponding author Email: jonas.anderegg@usys.ethz.ch

Abstract

The ability of a genotype to stay-green affects the primary target traits grain yield (GY) and grain protein concentration (GPC) in wheat. High throughput methods to assess senescence dynamics in large field trials will allow for (i) indirect selection in early breeding generations, when yield cannot yet be accurately determined and (ii) mapping of the genomic regions controlling the trait. The aim of this study was to develop a robust method to assess senescence based on hyperspectral canopy reflectance. Measurements were taken in three years throughout the grain filling phase on >300 winter wheat varieties in the spectral range from 350 to 2500 nm using a spectroradiometer. We compared the potential of spectral indices (SI) and full-spectrum models to infer visually observed senescence dynamics from repeated reflectance measurements. Parameters describing the dynamics of senescence were used to predict GY and GPC and a feature selection algorithm was used to identify the most predictive features. The three-band plant senescence reflectance index (PSRI) approximated the visually observed senescence dynamics best, whereas full-spectrum models suffered from a strong year-specificity. Feature selection identified visual scorings as most predictive for GY, but also PSRI ranked among the most predictive features while adding additional spectral features had little effect. Visually scored delayed senescence was positively correlated with GY ranging from r = 0.173 in 2018 to r = 0.365 in 2016. It appears that visual scoring remains the gold standard to quantify leaf senescence in moderately large trials. However, using appropriate phenotyping platforms, the proposed index-based parameterization of the canopy reflectance dynamics offers the critical advantage of upscaling to very large breeding trials.

Keywords: high-throughput phenotyping, canopy reflectance, hyperspectral remote sensing, field-based phenotyping, feature selection

2.1 Introduction

Maximizing carbon assimilation by a prolonged green leaf area duration after anthesis is a major breeding aim in many crops. This so-called "stay green" (Thomas and Smart, 1993) has been linked to increased grain yield (GY) in several crops (*reviewed by* Gregersen *et al.*, 2013). Stay green results from a delayed onset of senescence and/or a reduction in the rate of the process (Gregersen *et al.*, 2013). The benefit of such an extended period of functional stay green, i.e. a prolonged photosynthetic activity, has been particularly well documented in maize and sorghum (Borrell *et al.*, 2000; Rajcan and Tollenaar, 1999b).

In wheat, potential GY is currently viewed as being predominately limited by sink strength, i.e. the number of grains available for grain filling, which is largely determined up until and including a short period after anthesis (*reviewed by* Borrás *et al.*, 2004; Distelfeld *et al.*, 2014; Fischer, 2008). However, several studies have reported positive correlations between delayed senescence and GY, particularly under stress conditions (Bogard *et al.*, 2011; Christopher *et al.*, 2008, 2014, 2016; Lopes and Reynolds, 2012; Montazeaud *et al.*, 2016; Verma *et al.*, 2004). Where plants are exposed to severe stress, the stay-green phenotype may be interpreted as the avoidance of premature senescence, which could result in source limitation, i.e. a lack of carbohydrates delivered to the developing grains (Borrás *et al.*, 2004). Fine-tuning senescence dynamics has therefore been proposed as a promising selection criterion in wheat breeding particularly under the scenario of an increased frequency of weather extremes, such as heat and drought.

Optimising senescence dynamics requires intense field testing for at least two reasons: (i) senescence *per se* is known to underlie complex genetic and environmental control (*reviewed by* Lim *et al.*, 2007), typically resulting in moderate to low heritability across environments (*e.g.*, Crain *et al.*, 2017; Lopes and Reynolds, 2012) and (ii) effects of altered senescence dynamics on key primary traits such as GY and grain protein concentration (GPC) often depend on the environment (Bogard *et al.*, 2011; Lopes and Reynolds, 2012a). For example, negative relationships between GY and stay-green have also been reported, especially in the absence of water- or nitrogen-limiting conditions (Derkx *et al.*, 2012; Jiang *et al.*, 2004; Kipp *et al.*, 2014; Naruoka *et al.*, 2012). A delayed or slow senescence has also been linked to a reduced efficiency of remobilization, with adverse effects on harvest index (Gong *et al.*, 2005; Yang and Zhang, 2006), nitrogen use efficiency and GPC (Gaju *et al.*, 2014; Gregersen *et al.*, 2008). GPC is a key quality parameter in bread wheat, which may be additionally lowered *via* a dilution effect if the increased post-anthesis C-compound synthesis of stay-green cultivars is not paralleled by an increased uptake and transfer of nitrogen to the developing grains (Bogard *et al.*, 2010; Cormier *et al.*, 2016). Thus, in order to exploit variation in senescence dynamics for the improvement of bread wheat, a better understanding of environmental, genetic and physiological determinants of senescence dynamics *per se* as well as of the effects of senescence dynamics on GY and GPC in contrasting environments is required. Traditional phenotyping methods, such as visual senescence inspection (*e.g.*, Bogard *et al.*, 2011) or SPAD meter measurements (*e.g.*, Xie *et al.*, 2016) do not provide the necessary throughput to assess a dynamic trait for large numbers of genotypes at high temporal resolution and in contrasting environments.

Regular ground-based normalized difference vegetation index (NDVI) measurements obtained from an active spectral GreenSeeker sensor (NTech Industries, Ukiah, CA, USA) have shown significant potential for the rapid identification of variation in senescence patterns among wheat genotypes (Christopher *et al.*, 2014; Lopes and Reynolds, 2012; Montazeaud *et al.*, 2016). However, the use of a single and relatively unspecific spectral index is likely to entail some important limitations. During senescence, wheat canopies undergo a sequence of profound biochemical and biophysical changes. These changes in part temporally overlap and their effects on the reflectance spectrum of the canopy are therefore confounded. In this context, to the best of our knowledge, the NDVI has been used primarily as a generic indicator of canopy greenness or green biomass and has not been thoroughly validated as a tool to track canopy senescence in wheat. Gitelson and Merzlyak (1994) demonstrated the insensitivity of the NDVI to physiological changes occurring during early senescence at the leaf scale. At the canopy scale, the NDVI is often saturated in dense canopies as can be observed for wheat stands under favorable conditions (Asrar *et al.*, 1984; Gu *et al.*, 2013). This is likely to limit the sensitivity and precision of the NDVI in detecting early senescence at the canopy scale. Using passive sensors with a high spectral resolution, more specific narrow-band spectral indices (SI) or full-spectrum analysis can be deployed to reduce the effect of canopy structure and other confounding factors on the assessment of biochemical or physiological traits of interest (*e.g.*, Chen *et al.*, 2010; Haboudane *et al.*, 2002; Li *et al.*, 2014). For example, the plant senescence reflectance index (PSRI) developed by Merzlyak *et al.* (1999) can be used to measure leaf and fruit senescence. It is based on the chlorophyll/carotenoid ratio which undergoes major changes as a consequence of differential breakdown rates of these pigments during early senescence, offering advantages over the NDVI (Fischer and Feller, 1994; Merzlyak *et al.*, 1999; Sanger, 1971). Similarly, Kipp *et al.* (2014) were able to estimate greenness of flag leaves and onset of flag leaf senescence in wheat using ground-based hyperspectral canopy reflectance measurements in combination with full-spectrum models, while no stable relationships were found for the NDVI.

An additional advantage of hyperspectral reflectance measurements, as compared to single SI measurements, could consist in the opportunity to track multiple processes simultaneously. For example, during late development, green leaf area, pigment composition and total content, nitrogen distribution and water content of the canopy change dramatically. Visual senescence scorings mainly capture changes in pigment composition and content, but largely disregard other canopy characteristics, potentially resulting in a loss of breeding-relevant information. For example, the dynamics of nitrogen remobilization after flowering has been identified as a key determinant of GPC in wheat (*reviewed by* Kong *et al.*, 2016). In contrast to visual scorings, all of the aforementioned traits have been shown to be amenable to assessment using hyperspectral measurements provided sufficient variability exists (Becker and Schmidhalter, 2017; Haboudane *et al.*, 2002; F. Li *et al.*, 2014; X. Li *et al.*, 2014). In a breeding context, variability for a trait of interest is typically low, and differences in morphology and canopy structure among genotypes are thus likely to mask their effects on spectral reflectance at a specific point in time. However, assessments of relative changes over time could reveal differences in trait dynamics, which can be analyzed at the level of genotypes or experimental plots. Thus, we hypothesized that capturing the dynamics of such traits using repeated reflectance measurements during late development

could complement a precise representation of canopy greenness. The objective of the present study was two-fold: First, we aimed to develop a high-throughput method based on spectral reflectance to track visually observed senescence dynamics in a large population of morphologically diverse wheat genotypes. Second, we aimed to establish whether the resulting representation of canopy greenness decay could be complemented with additional information (e.g. relating to pigment, nitrogen or water content of the canopy) derived from repeated hyperspectral reflectance measurements.

2.2 Materials and Methods

2.2.1 Plant Materials, Experimental Design and meteorological data

A field experiment was conducted in the field phenotyping platform (Kirchgessner *et al*., 2017) at the ETH Research Station for Plant Sciences Lindau-Eschikon, Switzerland (47.449N, 8.682E, 520 m a.s.l.; soil type: eutric cambisol) in the wheat growing seasons of 2016-2018. In each year 300 cultivars comprised in the GABI wheat panel (Kollers *et al*., 2013) obtained from the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) were used, which were complemented with important Swiss cultivars for a total of 335 cultivars in 2016 and a total of 352 cultivars in 2017 and 2018. The cultivars were grown in plots of 1 m × 1.4 m size. The designs were generated using the R package DiGGer (Coombes, 2009; http://nswdpibiom.org/austatgen/software). The plots were arranged in a two dimensional incomplete block design with checks. The test varieties were randomized in two complete replications (one per lot). Within each replication, these test varieties were allocated to incomplete row blocks of size one (one row per block) and incomplete range blocks of size six (six ranges per block). The check varieties were distributed as follows: In 2016, wheat cultivar CH CLARO was used as a check variety at 21 evenly distributed locations in each replicate leading to a total of 42 checks per design. In 2017 and 2018 the three Swiss cultivars CH CLARO, SURETTA and NARA (DSP, Delley, Switzerland) were allocated to nine complete blocks spanning seven rows by six ranges each, summing up to a total of 54 checks per design. In all cases, at least one check was present per row and column of the design. Crop husbandry was performed according to local agricultural practice. The experiments were sown with a sowing density of 400 plants m^{-2} on Oct 13, 2015, on Nov 1, 2016, and on Oct 18, 2017, respectively. Temperature data was retrieved from an on-site weather station. Rainfall data was obtained from a nearby weather station of the federal Swiss meteorological network Agrometeo (www.agrometeo.ch) located at ca. 250 m distance to the field trial. The temperature data was used to calculate growing degree-days (GDD) following

$$Tmean_d = \frac{\sum \frac{maxT_{d.h} + minT_{d,h}}{2} - baseT}{24}$$

$$GDD = \sum_{d=1}^{n} Tmean_d$$

where $Tmean_d$ is the mean temperature for day d after heading, $maxT_{d,h}$ and $minT_{d,h}$ are hourly maximum and minimum temperatures for day d and $baseT$ is the base temperature, set to 0°C.

2.2.2 Phenology and agronomic data

Heading date was recorded when 50% of the spikes were fully emerged from the flag leaf sheath (BBCH 59, Lancashire *et al.*, 1991). Senescence was assessed visually, separately for the flag leaf and the whole canopy, following guidelines provided by Pask *et al.* (2012). Flag leaf senescence was scored based on the portion of green leaf area on a scale from 0 (0% green leaf area) to 10 (100% green leaf area). An integer mean value was estimated for plants located in a central region of about 0.5 m × 0.5 m of each plot. Whole plot senescence was scored on the same scale by estimating the overall greenness of the plot when inspected at a view angle of approximately 45° considering the entire plot area. Where necessary, the canopy was opened by hand to enable inspection of lower canopy layers. All scorings were done in 2-4 day intervals. Senescence scorings were done from approximately 20 days after flowering to complete canopy senescence. All heading and senescence scorings were done by the same person. The progression of leaf and whole plot senescence as assessed by visual scorings was then fitted against thermal time after heading (BBCH 59) for each individual plot using linear interpolation as well as a Gompertz model with asymptotes constrained to 0 and 10 (eq. 1; Gooding *et al.*, 2000),

$$S = 10e^{-e^{-b*(t-M)}} \text{ (eq. 1)}$$

where *S* represents the scaled senescence scoring, *t* is the accumulated thermal time after heading for a given plot, *b* is the rate of senescence at time *M* and *M* is the accumulated thermal time after heading when senescence rate is at its maximum. Eq. (1) was fit for each experimental plot using the R package '*nls.multstart*' (Padfield and Matheson, 2018). Senescence dynamics parameters were then extracted as follows (Figure 2.1): Onset of senescence (On_sen) was defined as the time point when values fell below 80% of the initial maximum, midpoint of senescence (Mid_sen) when values fell below 50%, end of senescence (End_sen) when values fell below 20%, and duration (T_sen) was defined as the time between onset and end of senescence, similar to the procedure applied to NDVI data by Christopher *et al.* (2014). We will refer to the duration between heading and the onset of senescence as the duration of stay-green.

GY was determined by manually harvesting the sowing rows 7 and 8 (out of 9). Grain moisture content was measured on a subset of 290 plots in 2016, 108 plots in 2017 and 84 plots in 2018, using a Wile 55 moisture meter (Farmcomp Oy, FIN-04360 Tuusula, Finland). Where available, grain weight was normalized to 14% water content using the plot-specific moisture content. The mean value of the measured plots was used otherwise. GPC was determined using near-infrared transmission spectroscopy (Infratec™ 1241 Grain Analyzer; Foss, DK-3400 Hilleroed, Denmark).
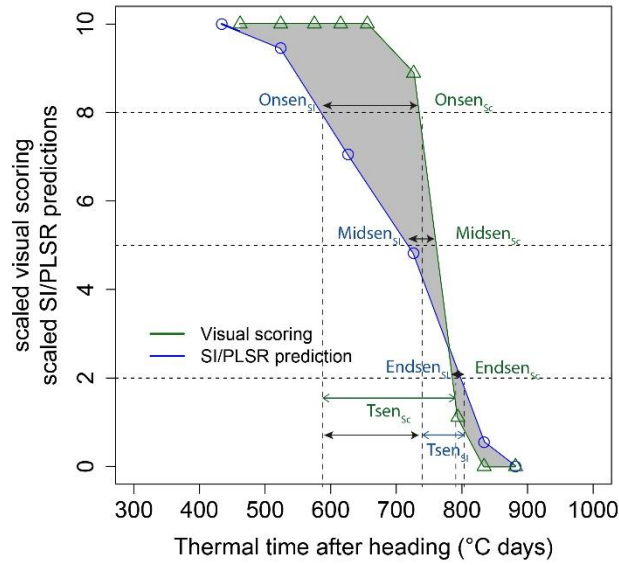
Figure 2.1 Scaled visual scorings of canopy greenness (Sc) and a scaled spectral index (SI) as a function of thermal time after heading for one experimental plot. Linear interpolation was used to derive the onset (On$_{sen}$), mid (Mid$_{sen}$) and end (End$_{sen}$) of the rapid senescence phase, its duration (T$_{sen}$) and the deviation of the SI curve from the Sc curve (error; shaded area). Black arrows represent the difference between SI- and Sc-derived parameters. The mean of these differences across all plots represents a measure of bias.

2.2.3   Statistical Analysis

The derived senescence dynamics parameters and agronomic traits were spatially corrected using two-dimensional P-splines as implemented in the R-package SpATS (Xose Rodriguez-Alvarez *et al*., 2018). To fit an independent smoothed surface to each replicate, the replicates were allocated diagonally in a grid of 49 rows by 41 ranges with replicate one ranging from row 1 to 22 and range 1 to 18 and replicate two ranging from row 27 to 49 and range 23 to 41. The spatial model was:

$$Y_{ijkl} = f(r_i, c_j) + K_l + G_k + R_i + C_i + \varepsilon_{ijkl}$$
$$(1)$$

where $f(r_i, c_j)$ is a smoothed bivariate surface defined by row r (i=1,…,49) and range c (j=1,…,41) as covariates (*for details see* Xose Rodriguez-Alvarez *et al*., 2018), K is the fixed effect of the check or the mean of all test genotypes (l = 1, 2, 3, $\bar{\mu}$ test), G is the random effect of the test genotypes (k = 1, …, 351), with check genotypes coded as missing. R_i and C_j are random factors of the rows and ranges, respectively, and ε is the random error vector. Twenty spline points were used each for rows and ranges.

To obtain best linear unbiased estimators (BLUEs) for all genotypes, the factor genotype was considered as a fixed effect in model (1) (k = 1, …, 354) and K was omitted from the model. The sum of the genotypic BLUE and the plot-specific residual error was extracted as a spatially corrected plot value.

Within-season repeatability (w^2) of the spatially corrected traits was calculated according to Xose Rodriguez-Alvarez *et al.* (2018) based on the genetic effective dimensions provided by SpATS as:

$$w^2 = \frac{ED_g}{m_g - 1}$$
(2)

where ED_g is the effective dimension for the genotypes and m_g is the total number of genotypes evaluated.

Spatially corrected plot values derived from (1) were used for the multi-year model using the R package '*asreml-4*' (Butler *et al.*, 2018):

$$Y_{ihkl} = \mu + K_l + G_i + Y_h + B_{k(h)} + GY_{ih} + \varepsilon_{ihkl}$$
(3)

where Y_{ihkl} is the spatially corrected senescence dynamics parameter or single plot measurement estimated in (1), μ is the overall mean, Y fixed effect of the year (h = 2016, …, 2018), B is fixed effect of the replication within year h (k = 1, 2), GY_{ih} the random genotype-by-year interaction and εihkl is the random normally distributed error with a year-specific variance. The effect of the replicate was specified only for years where more than one replicate was measured (i.e. for reflectance-based traits, where both replicates were measured only in 2016).

Across-year heritability was derived according to the method proposed by Cullis *et al.* (2006) as:

$$H_C^2 = 1 - \frac{avsed^2}{2\hat{\sigma}_G}$$

where H_C^2 is the heritability that is appropriate for complex residual structures (though not needed here) and avsed is the average standard error of prediction differences provided by the predict.asreml function. In the original equation provided by Cullis *et al.* (2006), the avsed is expressed as the mean variance of a difference between a pair of genotype $\bar{v}BLUP_{diff}$, the square of avsed (Isik *et al.*, 2017).

2.2.4 Hyperspectral assessment of senescence dynamics

Hyperspectral reflectance measurements
Canopy hyperspectral reflectance in the optical domain from 350 to 2500 nm was measured using a passive spectroradiometer (ASD FieldSpec® 4 spectroradiometer, ASD Inc., USA) equipped with an optic fiber with a field of view of 25˚. Whenever possible, measurements were carried out between 10:00 and 14:00 local time under clear and cloudless conditions. However, given the need for frequent measurements and the geographic location of the experiment, this was not always possible. Reflectance spectra were recorded as the average of 15 – 25 separate spectral records. Measurements were taken from nadir view holding the sensor at a height of approximately 0.4 m above the canopy. In 2016, reflectance spectra were recorded for 1-2 locations per plot holding the sensor in a nadir position above a crop row. In 2017 and 2018, 5
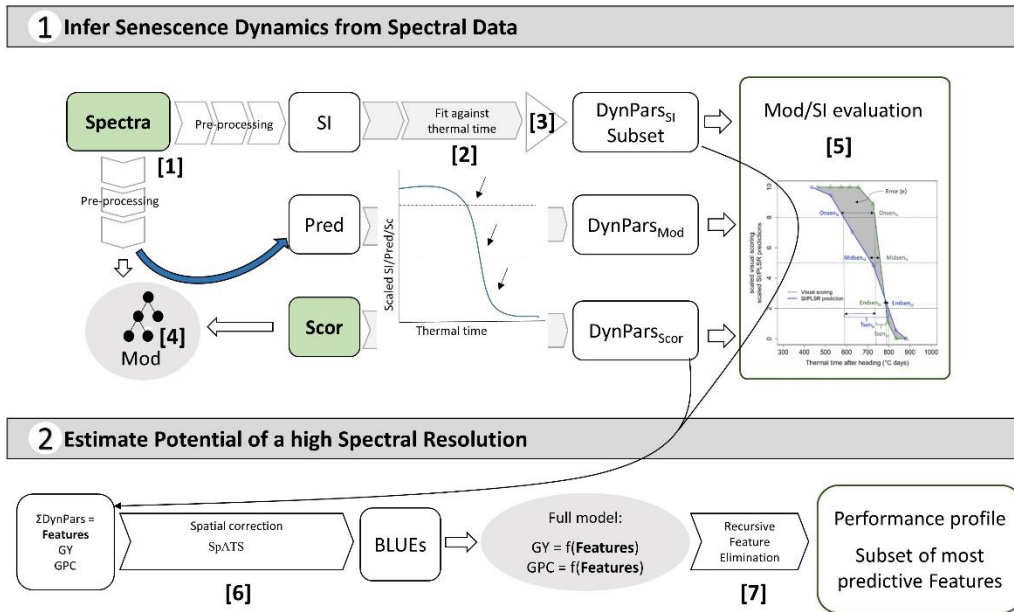
Figure 2.2 Overview of the objectives of this study and the implemented workflow: pre-processing of reflectance spectra and conversion to spectral indices (SI) [1]; full-spectrum models (Mod) to obtain predictions (Pred) of visual senescence scorings (Sc) based on reflectance spectra [4]; fitting of SI, Sc and Pred against thermal time and extraction of corresponding dynamics parameters (DynPars) [2]; Unsupervised DynPars$_{SI}$ subset selection [3]; Model and SI evaluation based on DynPars [5]; Spatial correction and calculation of best linear unbiased estimators (BLUEs) [6]; Modelling of primary traits (i.e. grain yield (GY) and grain protein concentration (GPC)) and supervised feature selection by recursive feature elimination [7] to determine the most predictive features and estimate the potential benefits of a high spectral resolution.

spectra were recorded while moving the fiber optic along the diagonal of each plot. This change in the measurement procedure was decided to reduce the variance of reflectance measurements due to plot heterogeneity in senescence observed in the first year. A Spectralon® white reference panel was used for calibration before measuring canopy reflectance, and the calibration was repeated approximately every 10 min. Under more variable conditions, the device was re-calibrated more frequently. When light conditions changed perceivably, measurements were interrupted immediately and the device was recalibrated before continuing the measurements under stable light conditions. In 2016, both replicates were measured, requiring about 3 hours on average, whereas in 2017 and 2018 measurements were limited to one replicate, requiring about 2 hours on average. The experiments were measured between heading and physiological maturity on 7 dates in 2016, on 8 dates in 2017 and on 12 dates in 2018. Thus, the frequency of spectral measurements was slightly lower than the frequency of visual scorings. The resulting hyperspectral dataset was then analyzed from two different perspectives relating to the main objectives of this study (Figure 2.2, upper and lower panel, respectively). The two approaches are described in more detail in the following sections and in supplementary methods (see supplementary material chapter 2). For ease of notation, reflectance at specific wavelengths will be abbreviated as R followed by the wavelength (e.g. R750).

Spectral Indices and full-spectrum models to infer senescence dynamics

An assessment of the performance and robustness of SI and full-spectrum models to track canopy senescence across environments was performed. A detailed description of the methodology is provided in supplementary methods (see supplementary material chapter 2). In brief, a large number of published spectral indices were computed and full-spectrum models to infer visually observed senescence scorings were calibrated from pre-processed reflectance spectra (Figure 2.2, [1], [4]). Models were used to generate predictions of senescence scorings for unseen data of the same environment as used in model calibration and of environments not included in model calibration. The resulting SI values and model predictions were scaled to range from 0 to 10, representing the minimum and maximum value recorded or predicted for the assessment period, respectively. Scaled values were fitted against thermal time after heading, and parameters describing the observed dynamics were extracted from time courses as was done for visual scorings (Figure 2.2, [2]). A subset of spectral indices was then selected using several filtering criteria to reduce multi-collinearity of the dataset (Figure 2.2, [3]). For full-spectrum models, waveband selection was performed in each experiment using recursive feature elimination. Performance and robustness of selected SI and full-spectrum models was assessed by comparing the dynamics parameters obtained from selected SI and model predictions to those obtained from visual scorings as shown in Figure 2.1 (step [5] in Figure 2.2). Pearson product moment correlation coefficients were calculated for the dynamics parameters obtained from linear interpolation of visual scorings, SI and model predictions. The mean difference in GDD over all experimental plots between the SI-derived and the scoring-derived parameters was also calculated to reveal potential general bias. Finally, the area between the resulting lines was calculated as a measure of precision in tracking the entire process.

Multiple spectral indices during senescence to predict primary traits

Finally, all senescence dynamics parameters obtained from scorings and from the selected SI (hereafter referred to as features) were analyzed directly for their association with GY and GPC. BLUEs or spatially corrected values were used for the analysis (Figure 2.2, [6]). We aimed to answer three separate questions in a step-wise procedure: Firstly, whether a phenotypic correlation between senescence dynamics and GY and GPC existed for any given trait in any given year; we used simple linear regression models for this purpose. Secondly, if the results suggested the presence of such a linear correlation, we investigated the potential of additional information contained in multiple SI time courses as opposed to the time course of a single SI or visual scoring. Such single SI or scoring values are likely to capture only part of the changes occurring during senescence (e.g. the dynamics of chlorophyll breakdown) while other processes might hold complementary information. Thirdly, we aimed at identifying the most important features to predict the trait. The rationale behind this was the following: Given a significant correlation between senescence dynamics and GY and GPC and a number of features describing aspects of senescence, the feature identified by the model to be the most relevant feature to predict GY or GPC should also be the one feature that most precisely captures the relevant aspects. For this purpose, we conducted supervised feature selection by recursive feature elimination (Figure 2.2, [7]; *see* Ambroise and McLachlan, 2002; Granitto *et al*., 2006; Guyon *et al*., 2002 *for a*

detailed description and discussion of the methodology). A detailed description is provided in supplementary methods (see supplementary material chapter 2).

2.3 Results

2.3.1 Experiments represented contrasting environments

Weather conditions during the main growing phase of the three experimental years strongly contrasted (Figure 2.3). The year 2016 was characterized by a wet summer with high precipitation causing severe lodging and high levels of foliar diseases. In particular, high levels of *Septoria tritici* blotch (STB) caused by the fungal pathogen *Zymoseptoria tritici* were observed. A total of 88 plots had to be excluded from further analyses due to heavy lodging. An additional 24 plots were excluded due to extended patches affected by take-all disease (*Gaeumannomyces graminis* var. *tritici*), which made objective senescence scorings and reflectance measurements impossible. Contrarily, the years 2017 and 2018 were characterized by dry summers and in 2017 additionally

Table 2.1 Descriptive statistics, within-year repeatability and across-year heritability for heading date (HD), grain yield (GY), grain protein concentration (GPC) and senescence dynamics parameters derived from linearly interpolated visual canopy senescence scorings (Sc,Lin) and from linearly interpolated values of the PSRI (PSRI,Lin). Data referring to individual years is reported sequentially for consecutive years (2016/2017/2018).

	Mean (Min; Max)	w^2	h^2 (2016-2017-2018)
HD (Julian Day)	152 (146; 158)/ 152 (145; 157)/ 144 (137; 150)	0.96/ -/ -	0.97
GY (t ha^{-1})	4.66 (2.81; 6.86)/ 5.39 (3.40; 7.55)/ 5.63 (2.92; 8.16)	0.54/ 0.40/ 0.41	0.55
GPC (%)	12.7 (10.2; 15.8)/ 12.9 (10.7; 15.8)/ -	0.83/ 0.70/ -	0.84*
Onsen$_{Sc,Lin}$ (°Cdays)	647 (435; 854)/ 603 (353; 748)/ 598 (447; 748)	0.72/ 0.44/ 0.64	0.60
Midsen$_{Sc,Lin}$ (°Cdays)	701 (544; 894)/ 675 (456; 815)/ 646 (487; 789)	0.78/ 0.65/ 0.68	0.75
Endsen$_{Sc,Lin}$ (°Cdays)	748 (592; 926)/ 730 (516; 854)/ 706 (529; 873)	0.79/ 0.62/ 0.64	0.76
Tsen$_{Sc,Lin}$ (°Cdays)	101 (28; 258)/ 128 (49; 322)/ 108 (26; 301)	0.22/ 0.24/ 0.24	-0.80
Onsen$_{PSRI,Lin}$ (°Cdays)	-	0.75/ -/ -	0.57
Midsen$_{PSRI,Lin}$ (°Cdays)	-	0.79/ -/ -	0.77
Endsen$_{PSRI,Lin}$ (°Cdays)	-	0.84/ -/ -	0.76
Tsen$_{PSRI,Lin}$ (°Cdays)	-	0.74/ -/ -	0.06

by high temperatures with daily maximum temperatures exceeding 30°C on several days, particularly during grain-filling. While both biotic and abiotic stresses can affect senescence dynamics in wheat, the underlying responses are stress-specific and may be controlled by very different genes or gene networks (Guo and Gan, 2012). Consequently, the three experimental years can be considered contrasting environments for the assessment of senescence dynamics and effects on GY and GPC.

2.3.2 Large variability and moderate to high heritability for senescence dynamics and agronomic traits

Large variability was observed for heading date, GY and GPC among the >330 genotypes in all years (Table 2.1). Heading occurred 8 days earlier in 2018, likely due to the comparably dry conditions in spring (Figure 2.3). Large variability was also observed for senescence dynamics, with a difference of > 300 °C days in the onset between the earliest and the latest genotype in all years. Similarly, the duration of senescence varied strongly across genotypes. The rate of senescence was somewhat lower in 2017 as expressed by an increased duration of the process.
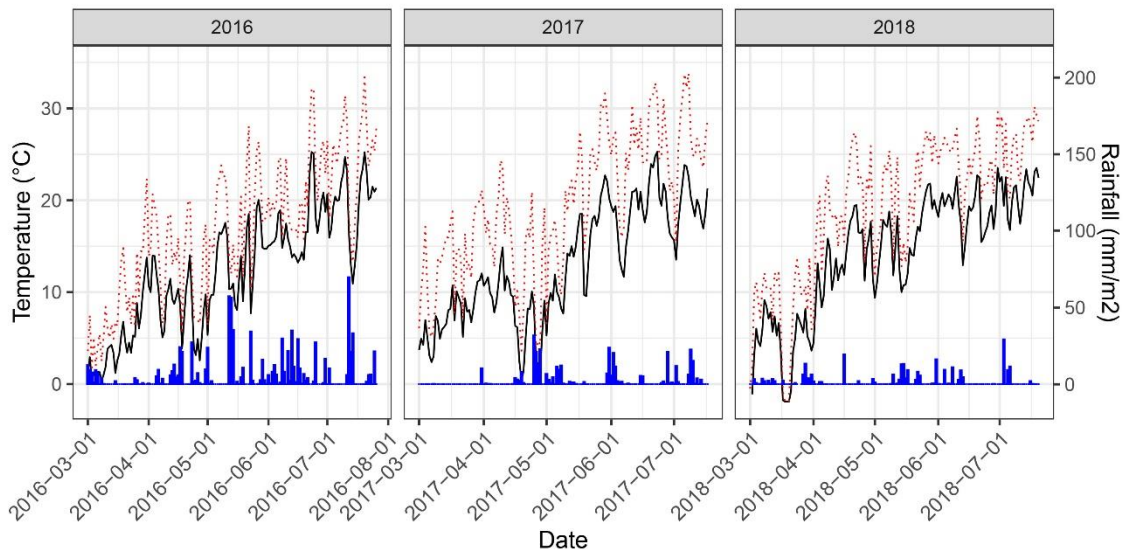

Figure 2.3 Daily mean temperatures (black solid line), daily maximum temperatures (red dotted line) and rainfall measured at 2 m above the ground for the main growing period of the experiment at the field phenotyping platform of ETH Zurich. Temperature data was retrieved from an on-site weather station. Rainfall data was obtained from a nearby weather station of the federal Swiss meteorological network Agrometeo (www.agrometeo.ch).

Across all genotypes, flag leaf senescence was somewhat delayed with respect to canopy senescence in 2016 and 2017, especially in early senescing genotypes. In 2018, this sequential vertical pattern of senescence was much less pronounced (data not shown). The stay-green phase was shorter in the dry seasons of 2017 and 2018 and longer in the wet season of 2016. Correlations between the senescence dynamics parameters extracted from the non-linear model fit and linear interpolation of visual senescence scorings were high for On_{sen} (r = 0.94), Mid_{sen} (r = 0.99) and End_{sen} (r = 0.96), suggesting a good approximation of the dynamic patterns through linear interpolation. Therefore, linear interpolation was used for further analyses. Repeatability for

senescence dynamics parameters was higher in 2016 than in 2017 and intermediate in 2018 (Table 2.1). Repeatability for On_{sen}, Mid_{sen} and End_{sen} was moderate to high, ranging from 0.72 to 0.79, from 0.44 to 0.65 and from 0.64 to 0.68 in 2016, 2017 and 2018, respectively. For T_{sen}, repeatability was distinctly lower in all years. Across-year heritabilities were intermediate to high for On_{sen}, Mid_{sen} and End_{sen} (Table 2.1).
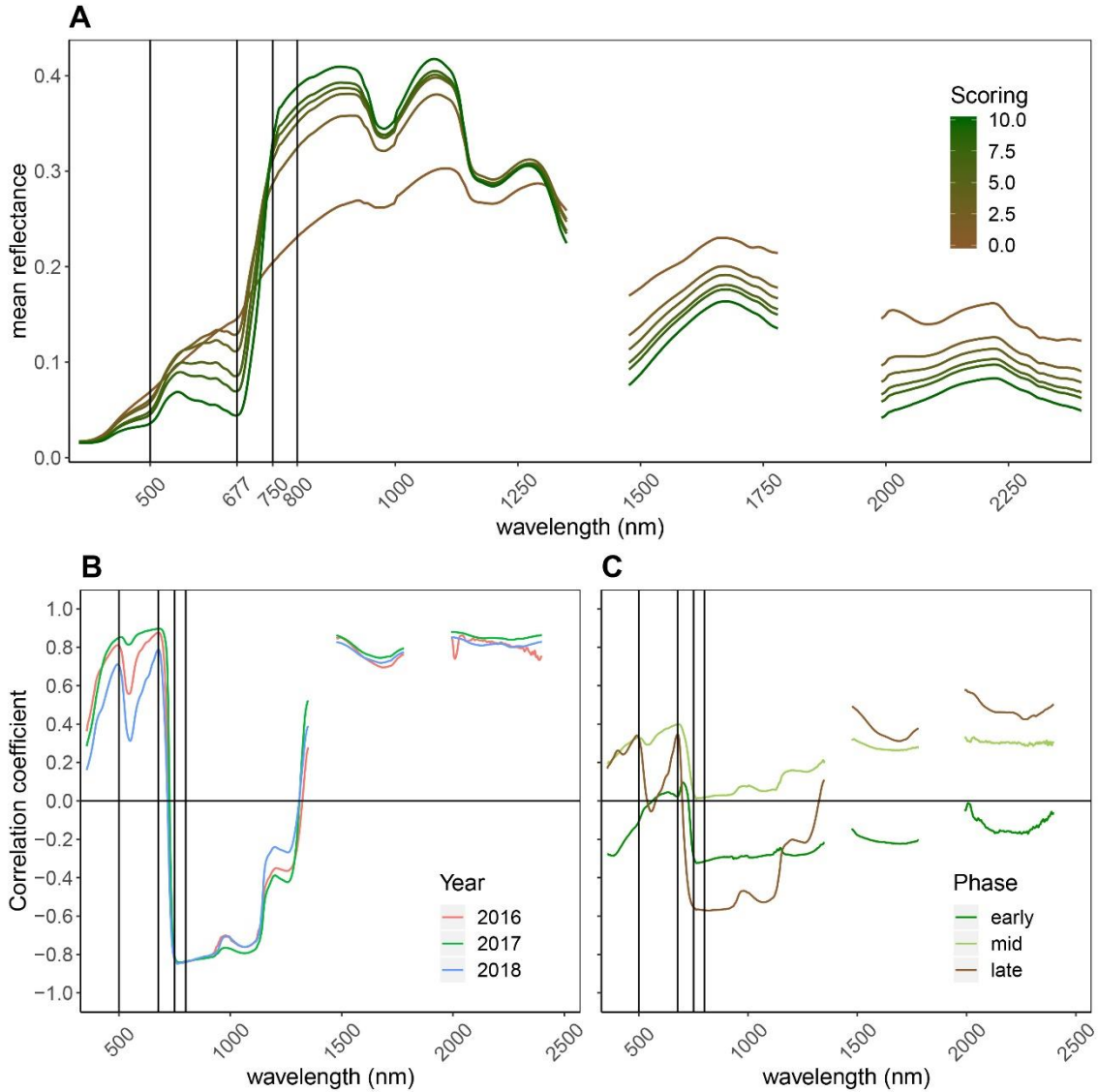

Figure 2.4: General reflectance patterns of senescing wheat canopies. **(A)** Mean reflectance spectrum of wheat genotypes through the process of senescence (10 denotes completely green canopies, 0 denotes complete senescence, based on visual scorings). Data from all time points and all years was used to calculate the mean reflectance spectrum per scoring. Vertical lines mark the wavebands constituting the NDVI and the PSRI. **(B)** Pearson correlation between reflectance at each wavelength and visual senescence scorings. Positive correlations indicate increasing reflectance as senescence progresses, negative correlations indicate decreasing reflectance as senescence progresses; Year-specific correlation coefficients. **(C)** Separate analyses for early senescence (scorings = [0:3]), intermediate senescence (scorings = [4:7] and late senescence (scorings = [8:10]). This part of the graph is based on data of the 2018 experiment.
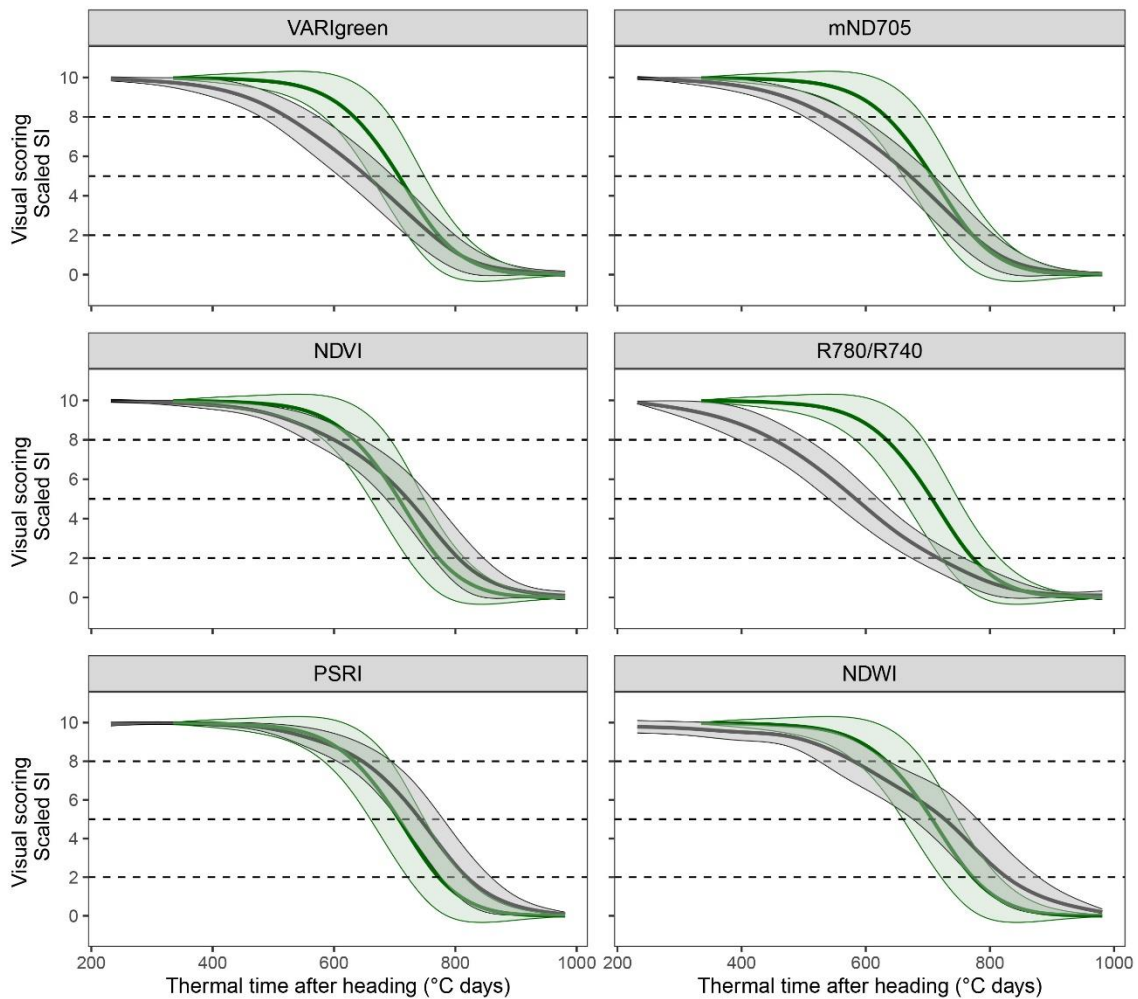
Figure 2.5: Dynamic pattern of scaled spectral indices (in grey) and visual canopy senescence scores (in green; identical in all subplots) over thermal time after heading. Mean linearly interpolated values over all experimental plots of the 2016 experiment (thick lines) and their standard deviations (thin lines) are shown. Dashed lines mark the thresholds defined as onset, midpoint and end of senescence.

2.3.3 Spectral reflectance is associated with visual senescence scorings in a non-linear manner

Senescence led to major changes in canopy reflectance throughout the recorded spectrum (Figure 2.4A). Reflectance in the visible range (VIS; 400-700 nm) increased strongly, whereas reflectance in the near infrared (NIR; 750-1300 nm) portion of the spectrum decreased. In the short-wave infrared (SWIR; 1475-1781 nm and 1991-2400 nm) portion of the spectrum, reflectance increased. Pearson correlation coefficients between the reflectance at each wavelength and the visual canopy senescence scores were calculated for each year (Figure 2.4B) and separately for different phases of the senescence process (Figure 2.4C). High positive correlations were found between the reflection in the VIS, with peaks at around 500 nm and 680 nm, as well as in the SWIR, indicating a decrease of light absorption (resulting in an increase in reflection) in these parts of the spectrum as senescence progresses. Strong negative correlations were found in the NIR with a peak near 750 nm, indicating a strong decrease of reflectance in this portion of the

spectrum as senescence progresses. These patterns were consistent across years. When different phases of the senescence process were analyzed separately, major differences in the correlations over large parts of the spectrum were found, indicating that reflectance throughout the spectrum is associated in a non-linear manner with visual senescence scorings.

2.3.4 Spectral indices track visually observed senescence dynamics across all years

A subset of 83 SI-derived senescence dynamics parameters was retained for further analyses. These included 21 T_{sen} parameters, suggesting that this parameter could be measured with a satisfactory repeatability ($w^2 > 0.5$) using certain SI. Several SI could be identified for which the mean value across all experimental plots followed clearly contrasting dynamic patterns (*see* Figure 2.5 *for examples*). Generally, the NDVI-derived senescence dynamics parameters correlated well with the scoring-derived parameters. However, for some SI the senescence parameters consistently correlated better with the scoring-derived parameters and were less biased (i.e. deviated less from scorings) than the NDVI-derived parameters (Table 2.2, Figure 2.5). PSRI-derived onset of senescence correlated best with scoring-derived onset and was unbiased ($r = 0.72$, $d_{Onsen} = 6°C$ days, $r = 0.78$ and $d_{Onsen} = -11°C$ days and $r = 0.75$, $d_{Onsen} = -7°C$ days for 2016-2018, respectively) as opposed to the parameter derived from NDVI ($r = 0.64$, $d_{Onsen} = -43°C$ days, $r = 0.63$ and $d_{Onsen} = -61°C$ days and $r = 0.51$, $d_{Onsen} = -57°C$ days in 2016-2018, respectively). PSRI also predicted midpoint of senescence with a high accuracy ($r = 0.76$, $d_{Midsen} = 43°C$ days, $r = 0.91$, $d_{Midsen} = 25°C$ days and $r = 0.86$, $d_{Midsen} = 26°C$ days in 2016-2018, respectively). Endpoint of senescence was predicted quite accurately ($r \approx 0.7$ across all years) by several SI, whereas the NDVI was clearly less stable across different years. Across all three years, End_{sen} derived from VARIgreen was correlated best with scoring-derived End_{sen} ($r = 0.79$, $d_{Endsen} = 10°C$ days, $r = 0.83$, $d_{Endsen} = -18°C$ days and $r = 0.82$, $d_{Endsen} = -17°C$ days in 2016-2018, respectively), while NDVI was clearly less precise and less stable across years ($r = 0.59$, $d_{Endsen} = 53°C$ days, $r = 0.78$, $d_{Endsen} = 65°C$ days, and $r = 0.54$, $d_{Endsen} = 52°C$ days in 2016-2018, respectively). The PSRI performed best in tracking the senescence process as observed visually from the onset to the end of the process, as expressed by comparably small error reprinted by the area between the curves (Table 2.2). Since repeatability for T_{sen} assessed visually was low in all years, results of the correlation analysis should be interpreted with caution, but the strongest correlations were found

Table 2.2 Pearson correlation (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$) between the senescence dynamics parameters derived from visual scorings and spectral indices, mean deviation in GDD between the derived parameters, and the total error throughout the entire process. Only spectral indices outperforming the NDVI in all three years for at least one parameter are listed, and the respective cells are highlighted in green. Several additional Spectral indices gave a better representation of End_{sen}, but only the VARIgreen, which performed best, is listed here.

Index	r_{Onsen}	$Bias_{Onsen}$ (°Cd)	r_{midsen}	$Bias_{Mnsen}$ (°Cd)	r_{Endsen}	$Bias_{Endsen}$ (°Cd)	r_{tsen}	$Bias_{tsen}$ (°Cd)	Mean error
mND705	0.50***/0.61***/ 0.44***	-101/-128/ -122	0.81***/0.83***/ 0.67***	-28/-45/ -50	0.77***/0.84***/ 0.69***	22/5/ -11	0.10*/0.04/ 0.10*	124/132/ 111	569/703/ 664
PRInorm	0.64***/0.76***/ 0.63***	41/23/ 19	0.53***/0.83***/ 0.67***	71/68/ 79	0.47***/0.56***/ 0.30***	82/119/ 104	0.16***/0.47*** /0.03	41/96/ 85	686/787/ 767
PSND1	0.64***/0.63***/ 0.51***	-41/-59/ -55	0.76***/0.84***/ 0.66***	25/8/ 5	0.59***/0.78***/ 0.55***	54/65 /53	0.22***/0.17**/ 0.04	95/125/1 07	497/578/ 539
PSRI	0.72***/0.78***/0. 75***	6/-11/ -7	0.76***/0.91***/ 0.86***	43/25/ 26	0.63***/0.85***/ 0.81***	64/62/ 43	0.27***/0.25*** /0.07	58/73/ 49	498/473/ 393
VARIgreen	0.44***/0.49***/0. 25***	-121/-175/ -135	0.71***/0.71***/ 0.56***	-45/-72/- 49	0.79***/0.83***/ 0.82***	10/-18/ -17	0.05/-0.05/ 0.03	131/157/ 117	648/898/ 716
NDVI	0.64***/0.63***/0. 51***	-43/-61/ -57	0.76***/0.83***/ 0.66***	24/7/3	0.59***/0.78***/ 0.54***	53/65/ 52	0.22***/0.16**/ 0.05	96/126/ 109	498/581/ 543

again for the PSRI. Thus, in summary, the PSRI outperformed the NDVI and all other tested SI in assessing most of the senescence dynamics parameters investigated here. Importantly, the observed correlations were stable across the three years. A comparison of experimental plots for which NDVI-derived On_{sen} strongly differed from the scoring-derived On_{sen} with RGB images suggested that this might be largely due to canopy structural effects such as leaf angles, spike geometry and spike orientation. This is illustrated in Figure 2.6 for two contrasting example plots, sown respectively with a genotype with changing spike orientation during grain filling (Figure 2.6, left) and a genotype with relatively stable spike orientation (Figure 2.6, right).

2.3.5 Full-spectrum models are environment-specific

We aimed to develop a further optimized spectral model to track wheat canopy senescence exploiting the full spectrum. Both tested algorithms resulted in significantly improved predictions of senescence scorings compared to the best SI for held out samples of the same year. This resulted in smaller errors in tracking the entire process (Table 2.2, Table 2.3).

Cubist regression models performed better and reduced the RMSE by an average of 0.2 with respect to the PLSR models (Table 2.3). Overall, PLSR-derived senescence dynamics parameters were not higher correlated with scoring-derived parameters than the SI-derived parameters (Table 2.3). In contrast, cubist produced better estimates of On_{sen}, Mid_{sen} and End_{sen} and outperformed the SI in most cases. The difference between algorithms was particularly ample in 2016. However, when models were validated across years, correlations were drastically reduced in many cases. Major differences were found for accuracy in predicting all senescence dynamics parameters,
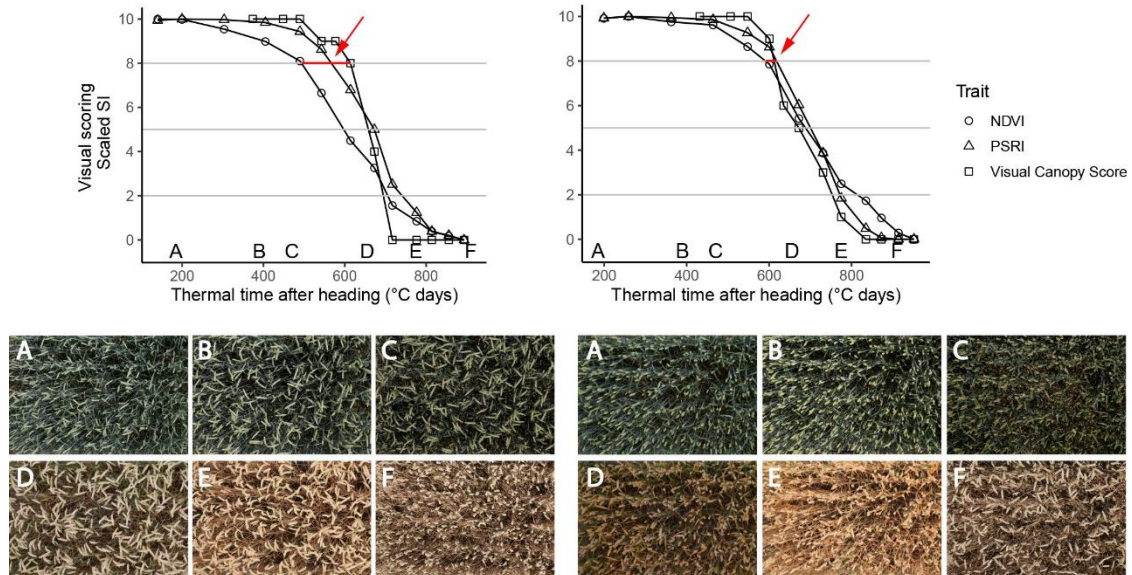

Figure 2.6: Time courses of PSRI, NDVI (nadir view) and visual scorings (whole plot, 45° viewing angle) for two experimental plots of the 2018 experiment. Left: Genotype with changing spike orientation during grain filling. With time, spikes make up an increasingly dominant part of the image. Concomitantly, NDVI values decrease early in the grain filling phase, while visual scorings indicate no change in canopy greenness (evidenced by red arrow). Right: Genotype characterized by relatively stable spike orientation during grain filling and comparable NDVI, PSRI and scoring time courses (evidenced by red arrow); Letters A-F in the upper part of the Figure represent time points when corresponding images were taken. Images were taken by the field phenotyping platform (FIP, Kirchgessner *et al.*, 2017).

Table 2.3 Within-year and across-year validation results for partial least squares regression (PLSR) and cubist regression models. Results are shown for smoothed reflectance spectra as input data. Data was mean-centered and scaled to unit variance prior to modelling. Root mean square error (RMSE) of the scoring predictions, Pearson correlation coefficients between the senescence dynamics parameters derived from visual scorings and full-spectrum models, and the total error in assessing the entire process (area between the curves) are shown. Cases where the full-spectrum models outperformed the PSRI are highlighted in green, other cases are highlighted in red. As the main interest lies on the capability of full-spectrum models to represent the entire process of senescence, an average RMSE for 10 different random upsamples of the test data are reported in brackets, where each upsample contains all possible scoring values an equal number of times, i.e. exactly the number of times of the most frequent observations.

Algorithm	Type of validation	Train Experiment	Validation Experiment	RMSE	r_{onsen}	r_{midsen}	r_{endsen}	r_{tsen}	e
PLSR	within year	2016	2016	1.01 (1.15)	0.50	0.81	0.71	0.12	412
PLSR	within year	2017	2017	0.82 (1.00)	0.75	0.88	0.76	0.09	425
PLSR	within year	2018	2018	0.81 (1.06)	0.81	0.88	0.83	0.22	231
PLSR	across year	2017	2016	3.81 (3.65)	0.21	0.30	0.25	0.05	712
PLSR	across year	2018	2016	3.19 (3.12)	0.21	0.43	0.59	-0.03	518
PLSR	across year	2017, 2018	2016	4.77 (4.58)	0.15	0.30	0.37	-0.09	658
PLSR	across year	2016	2017	2.42 (3.23)	0.63	0.85	0.81	-0.05	522
PLSR	across year	2018	2017	2.29 (2.21)	0.53	0.85	0.81	-0.27	511
PLSR	across year	2016, 2018	2017	1.76 (1.92)	0.53	0.84	0.83	-0.10	559
PLSR	across year	2016	2018	2.08 (2.42)	0.79	0.87	0.83	0.14	361
PLSR	across year	2017	2018	1.36 (1.57)	0.76	0.81	0.73	0.10	412
PLSR	across year	2016, 2017	2018	1.42 (1.66)	0.79	0.85	0.83	0.19	363
cubist	within year	2016	2016	0.78 (0.87)	0.76	0.85	0.85	0.30	290
cubist	within year	2017	2017	0.66 (0.82)	0.81	0.89	0.79	0.06	371
cubist	within year	2018	2018	0.66 (0.74)	0.83	0.89	0.86	0.29	171
cubist	across year	2017	2016	0.98 (1.44)	0.30	0.53	0.52	-0.02	464
cubist	across year	2018	2016	1.80 (2.31)	0.29	0.56	0.66	-0.01	456
cubist	across year	2017, 2018	2016	1.80 (2.24)	0.31	0.46	0.54	0.04	458
cubist	across year	2016	2017	1.03 (1.25)	0.76	0.85	0.80	-0.02	441
cubist	across year	2018	2017	1.46 (1.94)	0.76	0.86	0.79	-0.06	454
cubist	across year	2016, 2018	2017	1.25 (1.74)	0.77	0.86	0.81	-0.10	455
cubist	across year	2016	2018	1.17 (1.54)	0.82	0.86	0.87	0.25	200
cubist	across year	2017	2018	0.98 (1.44)	0.82	0.85	0.84	0.15	243
cubist	across year	2016, 2017	2018	0.97 (1.42)	0.82	0.86	0.86	0.23	224

depending on which year(s) were used for training and validation, respectively. Generally, adding a second year to the training data did not substantially improve model performance on samples of the held out year. In some cases, the correlations were even negatively affected by adding additional training data, especially when PLSR was used. Commonly observed problems were (i) remaining non-linearity in the predicted vs. observed regressions, particularly for PLSR (Figure 2.7A), and (ii) year-specific bias in the predicted vs. observed regression (Figure 2.7A, Figure 2.7B). Neither data type or pre-processing procedure was clearly and consistently superior to
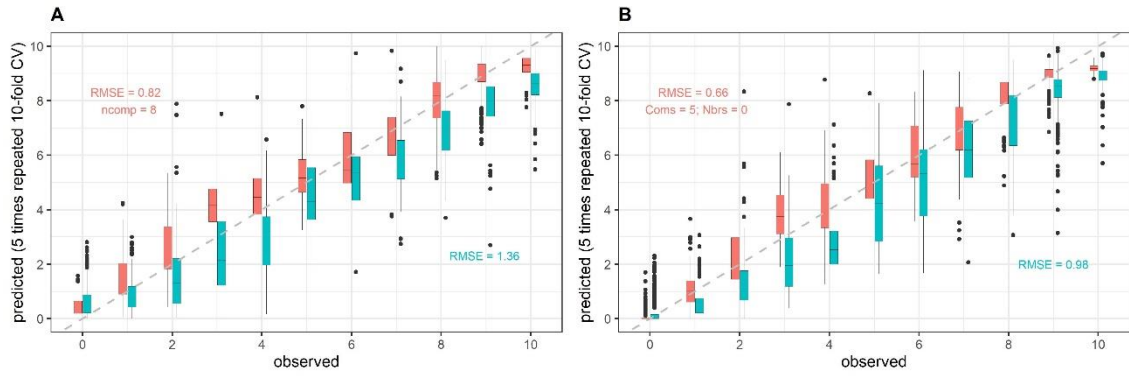
Figure 2.7 Example of model within-year (red) and across-year (cyan) validation results. Predictions of senescence scorings obtained from full-spectrum models are plotted against the visual scorings (observed). Here, averaged reflectance spectra were used, and the data was mean-centered and scaled to unit variance prior to modelling. Data from the 2017 experiment was used for model training. Models were validated on held-out samples of the same year (within-year validation) as well as on samples from the 2018 experiment (across-year validation). The full dataset was used for model training, i.e. no down-sampling was performed, whereas validation datasets were randomly down-sampled. **(A)** Results for partial least squares regression; **(B)** Results for cubist regression.
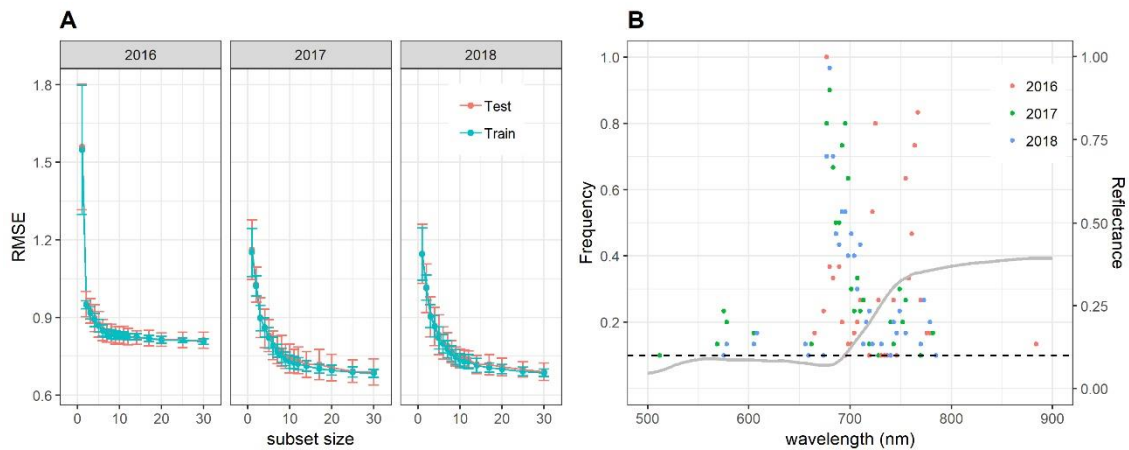

Figure 2.8 (A) Performance of the cubist regression models to predict visual senescence scorings depending on the number of wavelengths used as predictors. Mean performance as measured by the RMSE of predictions and standard deviations are shown based on 30 resamples of the data. **(B)** Frequency of wavelengths resulting among the most informative to predict visual scorings of canopy senescence. Frequencies denote the number of times out of 30 resampling iterations in which a given wavelength was retained in the cubist regression model down to a subset size of 12 wavelengths during recursive feature elimination. Only wavelengths which were among the top 12 predictors in at least 10% of the resamples (i.e. in at least 3 resamples, marked by the dashed horizontal line) are shown. The grey line represents the mean reflectance spectrum of canopies with a visual scoring of 8 (early senescence).

another. Notably, the year-specific bias could not be removed by using first derivatives or continuum-removed spectra.

Near-optimal models could be created using 6-8 wavelengths for 2016 and 12-14 wavelengths for 2017 and 2018 (Figure 2.8A). In all years, most of the commonly selected wavelengths were contained in the 650 nm to 800 nm range (Figure 2.8B). However, there were some obvious differences between 2016 on the one hand and 2017 and 2018 on the other hand. Models for 2016 frequently used several wavelengths between 720 nm and 770 nm, whereas models for 2017 and

2018 relied more heavily on the region from 670 nm to 720 nm, i.e. the chlorophyll absorption maximum and the red edge. Models for 2016 used a combination of R677 and one wavelength in the NIR (most often 764 nm or 767 nm) as the top two predictors in all 30 resamples and this combination contained most of the spectral information (Figure 2.8A). Contrarily, models for 2017 and 2018 used several wavelengths (typically 3-6) in the range from 677 nm to 695 nm before including a wavelength in the NIR or around 575 nm. Given the limited potential of full-spectrum models to infer senescence dynamics across years, we aimed to optimize the PSRI to the case of wheat canopy senescence and identify the factors driving its temporal dynamics. For this purpose, we simplified the PSRI to a simple ratio index and searched the spectrum for optimal waveband compositions for these simple ratio indices as well as for the original 3-band PSRI formula (Figure 2.9). The 750 nm waveband in the denominator of the PSRI is at the upper limit of the red edge. Moving R750 towards R800 did not significantly affect the accuracy of the index, whereas moving it into the red edge affected it negatively (Figure 2.9, upper left panel). Thus, similarly to the NDVI, the PSRI appears to be driven largely by chlorophyll absorption and canopy structure. However, omitting R500 from the PSRI (i.e. reducing the PSRI to a simple ratio index R678/R750) resulted in a decrease of its accuracy (Figure 2.9, lower left panel). Substituting R500 by neighboring wavelengths had little effect, although a small improvement was observed when replacing R500 by R525 (Figure 2.9, upper right panel).
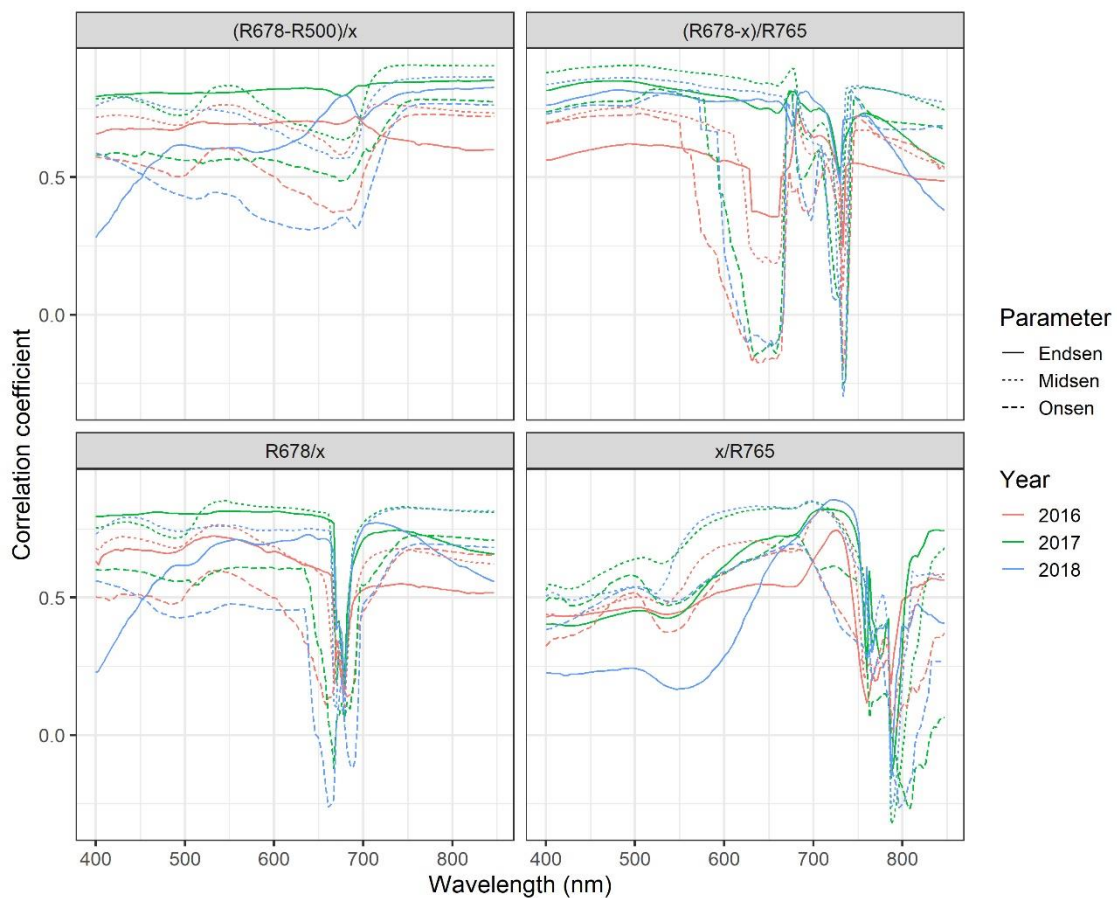

Figure 2.9 Correlation-based sensitivity analysis of the spectral bands (500, 678 and 756) constituting the PSRI. The "x" in the SI formula denotes the reflectance at the waveband that was varied in the depicted range.

2.3.6 Grain yield and grain protein concentration correlate with senescence dynamics

Simple linear regression models suggested the presence of significant, albeit rather weak, linear phenotypic correlations between senescence dynamics and GY and GPC in all years (Table 2.4). The strongest linear correlation was found between the PSRI-derived onset of senescence and GY in 2016 (r = 0.369, p < 0.001) which was slightly higher than the linear correlation between scoring-derived midpoint of senescence and GY (r = 0.365, p < 0.001) and significantly higher than the correlation between NDVI-derived onset of senescence and GY (r = 0.311, p < 0.001). A significant linear correlation was also found between T_{sen} derived from several SI and GPC in 2016 (r = -0.297, p < 0.001 for NPCI). In 2017 and 2018, there was only a weak (r < 0.19) linear correlation between senescence dynamics parameters and GY and GPC. In these years, scoring-derived parameters were always among the three most highly linearly correlated senescence dynamics parameters for both traits.

Heading date correlated negatively with the duration of the stay-green phase. The strongest correlation was observed in 2016, when the correlation between stay-green and GY was also strongest. However, multiple linear regression suggested a significant effect of stay-green duration on GY even when accounting for heading date, whereas heading date did not have a significant effect on GY (Table S 2). Both heading date and stay-green correlated negatively with GPC in both years (2016 and 2017, Table S 2). Thus, it seems that senescence dynamics had a direct effect on GY and GPC in our experiments.

2.3.7 Visual senescence scorings accurately track senescence-related processes affecting final grain yield

Given the phenotypic correlations between senescence dynamics parameters and GY and GPC in all three years, recursive feature elimination was performed for each trait × year combination. Performance of the models with a given subset size differed across years (Figure 2.10A). Multiple SI improved the prediction accuracy for GY and GPC as compared to single SI (Figure 2.10A). However, after inclusion of 2-3 features, mean model performance levelled off rapidly. In addition, there was significant variance in model performance estimates and feature ranks across resamples (Figure 2.10A, Figure 2.10B). Feature ranks showed lower variance in 2016 for both GY and GPC models, whereas in 2017 and 2018, there was considerable variance across resamples (Figure 2.10B). The most important feature in the 2016 GY model (i.e. midsen derived from the Gompertz model fitted to visual canopy senescence scorings) had an average rank of 1.40 (±0.97), indicating that it was consistently retained as the most predictive feature. For the 2017 GY model, the most important feature had an average rank of 4.10 (±5.57) and for the 2018 GY model, it had an average rank of 5.83 (±4.77), indicating much lower consistency across resamples (data not shown). In the GY model for 2016, features derived from the visual scorings were clearly the most predictive (Figure 2.10B). The features derived from the non-linear fit of visual canopy senescence scorings had lower ranks than the corresponding feature derived from linear interpolations, except for End_{sen}, for which the features had almost identical ranks. SI-derived features had much higher mean ranks than scoring derived features. The lowest ranked

31

SI-derived features were derived from the mND705, the PSRI, the NDRE, the REIP and the VARIgreen. No T_{sen} parameters were among the top 15 features of the GY models for any year, suggesting that the duration of senescence as assessed here did not affect GY in any experiment. Feature ranks were quite unstable across years, particularly for the GY models. In contrast to the GY models, no scoring-derived features were among the most important features in the GPC models. Instead, features derived from the PSND4 and the PRInorm had relatively low mean ranks in both years. This was in strong contrast to the results of the simple regressions, which suggested mainly a negative correlation between the duration of senescence and GPC in 2016 and a negative correlation between visually assessed stay-green and GPC in 2017 (Table 2.4).
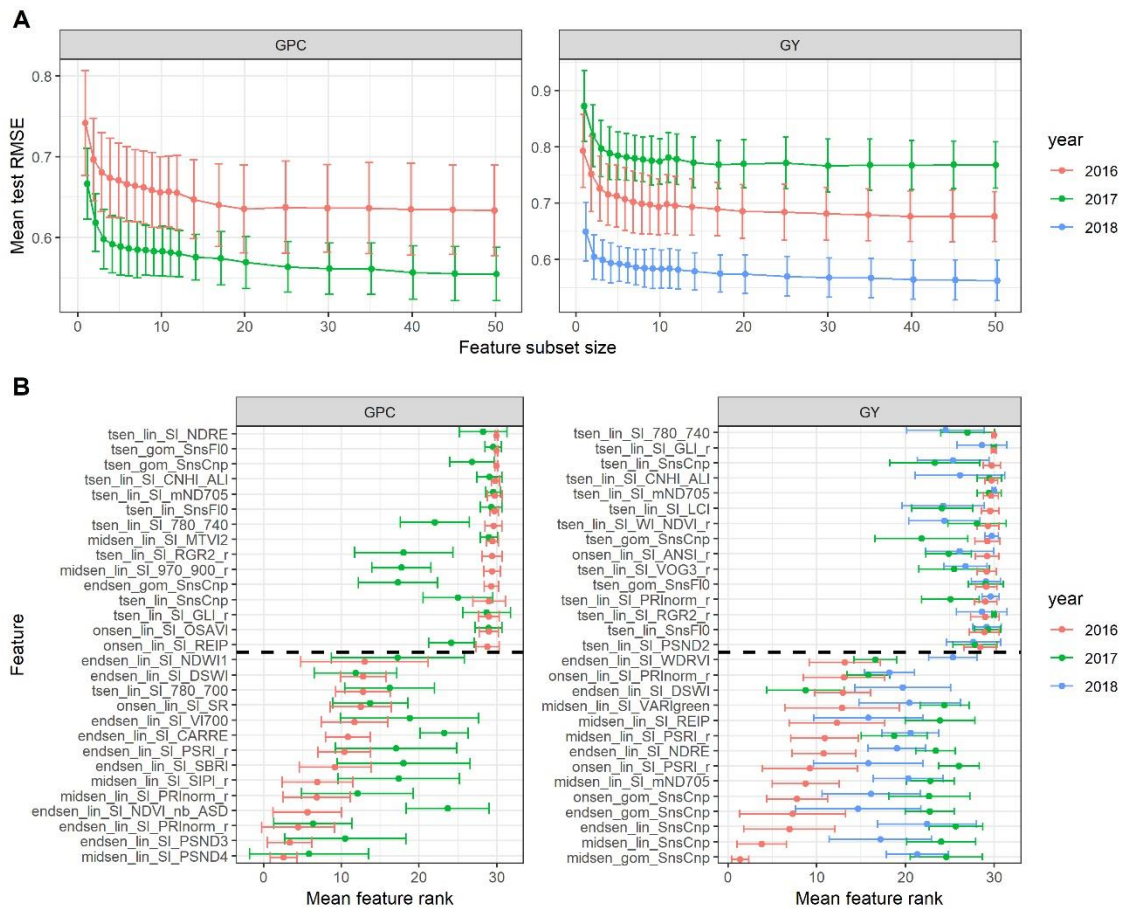

Figure 2.10 (A) Performance of the random forest regression models to predict grain protein concentration (GPC) and grain yield (GY). Mean performance and standard deviation are shown based on 30 resamples of the data for models containing a decreasing number of features selected by recursive feature elimination. **(B)** Feature ranks as determined by recursive feature elimination. Mean feature rank and standard deviation are shown based on 30 resamples of the data for the top and lowest 15 features, separated by the broken line. Features are plotted according to their descending mean rank in the 2016 models.

Table 2.4 Correlation (**, p < 0.01; ***, p < 0.001) between senescence dynamics parameters and grain yield (GY) and grain protein concentration (GPC) in different years. Pearson correlation coefficients are reported for the three most highly correlated features for each trait × year combination.

Year	Trait	Senescence dynamics parameter	Pearson r
2016	GY	onsen_PSRI	0.369***
		midsen_NDWI1	0.365***
		midsen_gom_SnsCnp	0.365***
	GPC	tsen_NPCI	-0.297***
		tsen_HI	-0.282***
		tsen_R780/R700	-0.259***
2017	GY	onsen_NDRE	0.187***
		onsen_gom_SnsCnp	0.182***
		onsen_MSR_rev	0.172**
	GPC	midsen_SnsCnp	-0.283***
		midsen_gom_SnsCnp	-0.269***
		onsen_gom_SnsCnp	-0.245***
2018	GY	tsen_RGR	0.191***
		midsen_SnsCnp	0.177***
		onsen_SnsCnp	0.173**

2.4 Discussion

2.4.1 Large genetic variability in senescence dynamics and minor effects on grain yield and grain protein concentration

Within-year repeatability of On_{sen}, Mid_{sen} and End_{sen} was moderate to high for scoring and SI-derived parameters, which is in line with previous reports (Blake *et al.*, 2007; Crain *et al.*, 2017; Lopes and Reynolds, 2012). Within-year repeatability of T_{sen} derived from some SI was similar, but was nearly zero for visual scorings (Table 2.1). This suggests that the duration and the rate of the senescence process is more accurately estimated using specific SI.

In this study, a positive correlation between the duration of the stay-green phase and GY was observed in all years. However, a strong correlation was found only in the wet season of 2016, whereas in the relatively dry and hot seasons of 2017 and 2018, correlations were weaker (Table 2.4). This is somewhat unexpected, as drought and heat stress are likely to anticipate and accelerate senescence (Gregersen *et al.*, 2013). This could result in source-limited GY and therefore enhance differences in GY between stay-green and early senescing genotypes (Borrás *et al.*, 2004). In 2016, visual senescence scorings were affected by foliar diseases, mainly STB. It seems likely that disease symptoms affected senescence scorings particularly during the late stay-green phase. High levels of STB can reduce GY significantly (*reviewed by* Fones and Gurr, 2015). Thus, differences in STB severity likely contributed to the observed correlation between visually assessed senescence dynamics and GY. Another possibility is that the phenotypic correlation between senescence dynamics and GY in 2016 arose at least in part from pleiotropic effects. Bogard *et al.* (2011) demonstrated that phenotypic correlations between senescence dynamics and

GY were mainly related to differences in flowering date in a doubled haploid mapping population. In our experiments, heading date was significantly correlated with the duration of the stay-green phase, but effects on yield were not statistically significant. This highlights that, in addition to facilitating the investigation of direct effects of secondary trait dynamics on primary traits, the implementation of high throughput phenotyping protocols may equally benefit the elucidation of such pleiotropic effects. A detailed understanding of such interdependencies is paramount to improve genetic crop models and fine tune dynamic traits in breeding (Chenu *et al.*, 2017, 2009). It should also be understood that such aspects will have to be taken into account when investigating the genetic determinants of senescence dynamics. In the subsequent sections, we discuss the results of different approaches to phenotype senescence as a dynamic trait.

2.4.2 Spectral indices emphasizing reflectance in the visible to near-infrared range accurately track canopy senescence dynamics

Regular NDVI measurements have been used by several authors to evaluate stay-green, mainly under drought conditions (Christopher *et al.*, 2014, 2016, 2018; Lopes and Reynolds, 2012; Montazeaud *et al.*, 2016). In this study, the PSRI gave a better representation of visually recorded canopy senescence dynamics than the NDVI. The dynamics of the SI suggest that the accuracy of the NDVI is not primarily hampered by saturation effects, as it tends to decrease earlier than the PSRI (Figure 2.5, Figure 2.6).

At the leaf scale, the PSRI specifically measures changes in pigment composition by comparing the reflectance at 500 nm, which is controlled by the combined absorption of chlorophyll *a*, chlorophyll *b* and carotenoids with absorption at 678 nm, which is controlled by chlorophyll *a* only (Merzlyak *et al.*, 1999). Major changes in pigment composition have been observed for flag leaves of field-grown wheat plants after about 20 days post-anthesis (Lu *et al.*, 2001). These changes in pigment composition coincided with the onset of a steep decrease in total chlorophyll content and thus probably with the onset of chloroplast dismantling, which marks the beginning of senescence (Havé *et al.*, 2017; Lu *et al.*, 2001). Therefore, it appears plausible that the PSRI is indicative of wheat canopy senescence. However, unlike at the leaf scale, R750 changes drastically during senescence at the canopy scale (Figure 2.4A). Thus, PSRI values at the canopy scale are strongly driven by R750. Reflectance in the NIR is dominated by leaf area index among other canopy structure parameters (Jacquemoud *et al.*, 2009). Based on a comparison with RGB images, we hypothesized the PSRI to be less sensitive to variation in canopy structure than the NDVI. The NDVI is highly sensitive to canopy structure, as R800 is one of two constituting wavebands. Canopy structure may change drastically prior to and during senescence. For example, leaf-roll can be induced by water shortage resulting in major canopy structural changes and an increased contribution of soil reflectance that is not necessarily related to senescence. Furthermore, changes in spike geometry are likely to interfere with the retrieval of biochemical information. Both factors strongly affect reflectance in the NIR, while reflectance in the VIS is less affected (Gutierrez *et al.*, 2015). It appears that the inclusion of a second waveband in the VIS stabilized the PSRI against canopy structural effects during early senescence (Figure 2.6).

The relatively low sensitivity of the observed correlations between the PSRI and visual canopy senescence dynamics to shifts in the constituting wavebands suggests that multispectral information is sufficient to obtain accurate estimates of canopy senescence dynamics. This makes the trait amenable to phenotyping using multispectral cameras which can be mounted on unmanned aerial vehicles (Aasen *et al.*, 2018; Aasen and Bolten, 2018). This would greatly facilitate large-scale screenings and frequent measurements. Such large-scale screenings and a high temporal resolution of measurements are likely to be the primary benefits of digital phenotyping of senescence dynamics in the near future.

2.4.3 Non-linear models outperform PLSR in tracking senescence dynamics, but are similarly environment-specific

Full-spectrum models improved the inference of visual senescence scorings from spectral data as compared to the best SI, but their power to track senescence dynamics was limited by the extraction of year-specific relationships between reflectance and scorings, and, in the case of PLSR, by their inflexibility to capture non-linear relationships between spectral reflectance and visual scorings. Such non-linearities likely arise from the fact that senescence is a complex process, during which major physiological and structural changes at the leaf and canopy scales occur sequentially or simultaneously with most of them having strong but contrasting effects on the reflectance characteristics of plant canopies. Such changes include chlorophyll degradation and changes in pigment composition, loss of cellular structure, mesophyll breakdown and water loss at the leaf level (Gitelson and Merzlyak, 1994) as well as a reduction in leaf area index and ground cover, changes in leaf and spike geometry, nutrient redistribution to the spikes and water loss at the canopy level.

PLSR failed to accurately track visually observed senescence dynamics in our experiment, and was outperformed by several SI, even when validated on held out samples of the same experiment. Kipp *et al.* (2014) found no stable relationships between various types of SI and flag leaf color, but reported a good predictive performance of PLSR models. This is not necessarily in contradiction to our observations, since we also found improved prediction of visual senescence scorings when exploiting the full spectrum. However, our objective was not to predict absolute values of greenness, but to track temporal changes throughout the process of senescence and extract parameters that describe these dynamics. Therefore, we scaled both scorings and spectra-derived predictions to a uniform range and only exploited the relative temporal changes (Figure 2.2, upper panel). With this intermediate step, we eliminated initial and terminal differences across genotypes or experimental plots, which can have multiple origins and interfere with the retrieval of dynamics parameters and measures of overall accuracy. The increase in accuracy of cubist compared to PLSR models was paralleled by an increased across-year applicability of the models on average, indicating that the problem of year-specific modelling was not exacerbated by using a more flexible algorithm.

In general, the RMSE of the cubist models was low (<0.7 in 2017 and 2018). We speculate that this is close to the performance ceiling set by the precision of visual scorings. Achieving substantial improvements by further optimizing the models seems therefore unlikely. Rather,

more precise ground truth data would be required. Visual scorings are subjective and limited in tracking small changes between assessment time points. SPAD meter or color measurements have been used by other authors (*e.g.*, Kipp *et al.*, 2014; Xie *et al.*, 2016). These tend to be more objective, more sensitive to subtle changes and relate more directly to a physiological trait. On the other hand, they sample only a small part of the leaf and are laborious to obtain. Also, senescence typically does not progress uniformly along the leaf, resulting in difficulties to obtain a good average value per plot. Thus, in our opinion, these measurements do not produce better average values per plot than a visual scoring. Furthermore, small gains in precision need to be weighed against the necessity of sampling a sufficiently large genotypic diversity at a high temporal resolution in several years/environments to achieve robust models, as illustrated above.

2.4.4 Model transferability is strongly related to differences in environmental conditions

We found major differences in the applicability of models across years. In particular, the dynamics of visual scorings in the 2016 experiments were very poorly predicted by models trained on 2017 and/or 2018 data (Table 2.3). Furthermore, models trained and validated within the 2016 experiment performed poorly compared to the other two years. This could be due to the different measurement protocol applied in 2016. Interestingly, however, models trained using data from 2016 performed well in 2017 and 2018. Therefore, it seems more likely that limited model applicability in 2016 is at least in part a consequence of a larger variability in how progression of senescence affected hyperspectral reflectance across genotypes in this year. In the same experiment and during the same period, major differences were found for STB severity among genotypes and STB was the dominant disease throughout the stay-green phase (*see* Karisto *et al.*, 2017 *for details*). In contrast, in 2017 and 2018 foliar diseases were at very low levels due to dry weather conditions. Several STB severity metrics were found to affect spectral reflectance in 2016, with strong effects particularly in the NIR (*see* Yu *et al.*, 2018 *for details*). We therefore hypothesize that STB altered the temporal evolution of the hyperspectral reflectance signal during the late stay-green and early senescence phases with respect to disease-free plots. Assuming that STB also affected the visual canopy senescence scorings at least during early senescence, this would explain the strong contribution of wavebands in the NIR to models in 2016 (Figure 2.8B). The results of the SI dynamics seem to offer some additional support for this hypothesis. Indeed, the difference in accuracy between the PSRI and the more generic NDVI in tracking visually assessed senescence is relatively small in 2016 as compared to 2017 and 2018 (Table 2.2). This suggests that changes in pigment composition were not much better indicators of senescence in 2016 than was a generic indicator of greenness such as the NDVI. Leaves affected by STB develop necrotic lesions, but do not undergo controlled dismantling of the photosynthetic apparatus resulting in the typical changes in color and in pigment composition probably contributing to the increased performance of the PSRI. Finally, the difference between performance of PLSR and cubist was particularly large for 2016 (Table 2.3). Under the scenario that STB affected overall greenness in the late stay-green and early senescence phase (see above) this pattern is to be expected, since STB should affect the spectral reflectance in a different manner than physiological senescence, which will dominate in later phases, increasing the non-linearity

between spectral reflectance and visual senescence scorings through the entire process. We hypothesize that repeated hyperspectral reflectance measurements during late stay-green and throughout senescence might allow to distinguish purely physiological senescence from partly disease-driven loss of green leaf area, and facilitate an indirect assessment of disease resistance in field-grown wheat at high throughput.

2.4.5 Digital senescence phenotyping may benefit crop breeding primarily through increased temporal resolution and throughput of measurements

Relatively strong linear correlations were observed between senescence dynamics parameters and GY and GPC in 2016 and results from feature selection are most conclusive for this year. Increases in model performance could be observed both for GY and GPC in 2016 when using multiple features and the obtained feature ranks were relatively stable across resamples. For GY, the most important features are either directly derived from visual senescence scorings of the canopy or from SI that were found to predict these scorings well (Table 2.2). Specifically, End_{sen} derived from NDRE is highly correlated to End_{sen} derived from visual scorings (r = 0.73), $mND705$ was found to be most accurate to predict Mid_{sen} (r = 0.81), followed by the PSRI (r = 0.76). NDRE and mND705 have been developed to improve sensitivity to chlorophyll content with respect to the NDVI (Barnes *et al.*, 2000; Sims and Gamon, 2002). This is achieved primarily by replacing the reflectance in the red by reflectance in the red-edge, which is less prone to saturation at high chlorophyll contents of leaves and vegetation and more robust in presence of leaf or canopy structural effects (Demetriades-Shah *et al.*, 1990; Gitelson and Merzlyak, 1994; Sims and Gamon, 2002).

The low average feature ranks of visual canopy senescence scorings and SI that accurately track these scorings suggest that the dynamics of chlorophyll breakdown was most predictive of GY, and that this trait could be assessed with a high precision using visual scorings or the proposed SI. This can be well explained, as the onset of chlorophyll breakdown marks the onset of remobilization and the end of photo-assimilation, thereby directly affecting source capacity. However, several additional conclusions can be drawn from these findings.

Firstly, it can be concluded that feature selection on time courses of multiple SI resulted in the identification of features most strongly associated with GY and describing a dynamic trait interpretable in terms of plant physiology.

Secondly, given that no SI-derived feature was more predictive of GY than scoring-derived features, we conclude that potential precision gains in estimating the switch from stay-green to remobilization using hyperspectral high throughput phenotyping techniques rather than visual scorings may be limited. It should be noted, however, that most of the SI used in this study were not developed for use in wheat canopies during senescence, and only few of them have been tested for their applicability during this growth stage (Barmeier and Schmidhalter, 2017; Erdle *et al.*, 2013; Hassan *et al.*, 2018). Significant relationships seem to be maintained during later growth stages, but tend to be unstable across stages (Erdle *et al.*, 2013). Nonetheless, we assume that the selected features summarize a considerable part of the total information contained in

hyperspectral measurements during this phase. We further conclude that visual scorings apparently allow assessing a key trait during senescence in a reliable manner. Further research should therefore aim at understanding the factors hampering across-year applicability of otherwise successful full-spectrum models to infer senescence scorings and how these factors can be accounted or corrected for. A method to obtain highly accurate training data of canopy greenness will also be required to achieve good predictive models. Additionally, the lower mean ranks of PSRI-derived features and higher linear correlation coefficients between PSRI-derived features and GY provide additional evidence for the superior precision of the PSRI compared to the NDVI.

Thirdly, in a first step, improvements in precision may be achieved mainly by increasing the temporal resolution of measurements. The higher ranks of features derived from the parametric models are likely the result of the smoothing properties of non-linear model fits, better approximating the gradual nature of the senescence process and reducing the impact of measurement or scoring errors associated with a particular time point on the estimation of dynamics parameters. In addition, parametric models would also allow for the derivation of measures that better separate distinct characteristics of the senescence process. In particular, the derivation of a parameter describing specifically the rate of senescence or any process occurring during senescence, could be highly beneficial to elucidate effects of senescence dynamics on primary traits, particularly GPC and nitrogen use efficiency, but also GY (Gregersen *et al*., 2008; Kong *et al*., 2016; Wu *et al*., 2012; Xie *et al*., 2016). In contrast, the T_{sen} parameter used here is partly reflected by the other parameters since it was derived by substracting On_{sen} from End_{sen}. It also integrates over the whole process, which may be overly simplistic and may not adequately represent senescence dynamics observed at the leaf or canopy scale (Bogard *et al*., 2011; Gaju *et al*., 2014).

Several of the selected features had relatively low linear correlation coefficients whereas some other highly ranked features also had high linear correlation with GY. Thus, it seems that rf extracted some non-linear relationships between features and GY and these seemed to be more predictive of GY than the linear correlations found for some features. Unfortunately, the final rf model is not interpretable due to its ensemble nature. We chose rf as a base learner for feature selection (1) because it is affected much less by the presence of non-informative predictors and multi-collinearity among predictors than parametrically structured models, (2) for its capability to capture non-linear relationships between predictors and the response and interactions between predictors which could not be excluded in our case and, most importantly, (3) precisely because of its ensemble nature that allowed it to produce stable variable importance rankings even in the presence of highly collinear predictors and consequently facilitated the removal of the less important one during subsequent feature elimination steps. We recognize that this may have come at the cost of less-than-optimal performance in the presence of strictly linear relationships between predictors and the response and might, in some cases, have resulted in the extraction of relationships that are difficult to interpret in terms of plant physiology or phenology. However, the fact that the scoring and PSRI-derived features were among the most highly ranked features,

while we also found high linear correlations suggests that these weaknesses of the rf algorithm should have impacted the result only marginally.

The above observations could not be confirmed in 2017 and 2018 in spite of the fact that simple linear regressions suggested that visual scorings and corresponding SI were again among the most predictive features (Table 2.4). It seems likely that the overall effect of senescence dynamics on GY and GPC was too weak in 2017 and 2018, which would also explain the increased variability of feature ranks across resamples. In the presence of small effects and under the hypothesis that the correlations between features and responses are close to linear, the results of linear regressions may be more reliable.

Finally, it should be noted that our analysis was based on the observation of a phenotypic correlation between senescence dynamics parameters and primary traits. We did not observe a significant effect of heading date on GY. Nevertheless, it cannot be excluded with certainty that this phenotypic correlation arose primarily as a result of pleiotropic effects, and this might have affected our conclusions. Subjecting genotypes to very harsh conditions post-anthesis is likely to accentuate direct effects of senescence dynamics on primary traits, enabling a more precise evaluation of the potential benefits of a high spectral resolution during late development.

2.5 Conclusions

Using existing variability in senescence dynamics for wheat improvement requires intensive field-testing of large populations in contrasting environments. We hypothesized that repeated spectral reflectance measurements may facilitate an accurate assessment of this developmental phase at high throughput. Our results show that time series of the PSRI accurately track visually observed canopy senescence dynamics across a large number of genotypes and under varying environmental conditions. When a substantial effect of senescence dynamics on GY was present, correlations between scoring-derived and PSRI-derived senescence dynamics parameters and GY were very similar. We therefore conclude that visual scorings could be replaced by PSRI measurements without a significant loss in precision. On the other hand, the high spectral resolution of measurements did not confer significant advantages over visual scorings or measurements of a single spectral index in our experiment. This is encouraging for the breeding and plant-phenotyping community, since it implies that senescence dynamics may be accurately tracked using less sophisticated and potentially cheaper spectral sensors. Thus, we conclude that digital senescence phenotyping will benefit wheat breeding through an increased temporal resolution and high throughput of measurements.

3 In-Field Detection and Quantification of Septoria Tritici Blotch in Diverse Wheat Germplasm Using Spectral–Temporal Features

Jonas Anderegg[1,*], Andreas Hund[1], Petteri Karisto[2], Alexey Mikaberidze[2]

[1]Crop Science Group, Institute of Agricultural Sciences, ETH Zurich, Zurich, Switzerland

[2]Plant Pathology Group, Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland

*Corresponding author Email: jonas.anderegg@usys.ethz.ch

Abstract

Hyperspectral remote sensing holds the potential to detect and quantify crop diseases in a rapid and non-invasive manner. Such tools could greatly benefit resistance breeding, but their adoption is hampered by (i) a lack of specificity to disease-related effects and (ii) insufficient robustness to variation in reflectance caused by genotypic diversity and varying environmental conditions, which are fundamental elements of resistance breeding. We hypothesized that relying exclusively on temporal changes in canopy reflectance during pathogenesis may allow to specifically detect and quantify crop diseases whilst minimizing the confounding effects of genotype and environment. To test this hypothesis, we collected time-resolved canopy hyperspectral reflectance data for 18 diverse genotypes on infected and disease-free plots and engineered spectral-temporal features representing this hypothesis. Our results confirm the lack of specificity and robustness of disease assessments based on reflectance spectra at individual time points. We show that changes in spectral reflectance over time are indicative of the presence and severity of *Septoria tritici* blotch (STB) infections. Furthermore, the proposed time-integrated approach facilitated the delineation of disease from physiological senescence, which is pivotal for efficient selection of STB-resistant material under field conditions. A validation of models based on spectral-temporal features on a diverse panel of 330 wheat genotypes offered evidence for the robustness of the proposed method. This study demonstrates the potential of time-resolved canopy reflectance measurements for robust assessments of foliar diseases in the context of resistance breeding.

Keywords: high throughput phenotyping, field-based phenotyping, feature engineering, feature selection, spectral vegetation index

3.1 Introduction

Hyperspectral remote sensing has shown significant potential for the rapid, non-invasive assessment of crop diseases at different scales, ranging from single leaves (*e.g.*, Ashourloo *et al.*, 2014; Mahlein *et al.*, 2010) to the canopy (*e.g.*, Cao *et al.*, 2013; Yu *et al.*, 2018) to fields and regions (Wakie *et al.*, 2016). Applications have been proposed primarily in the context of precision agriculture, but resistance breeding may equally benefit (Mahlein, 2016). The identification of novel sources of durable, quantitative disease resistance requires screening large and diverse germplasm collections under field conditions. Reflectance-based techniques hold the potential to reduce associated costs and allow for the screening of more genetic variation, if deployed on adequate phenotyping platforms (Aasen *et al.*, 2018; Aasen and Bolten, 2018; Kirchgessner *et al.*, 2017). This may enable indirect selection in early breeding generations and facilitate the identification of novel sources of resistance.

However, to benefit crop breeding, new methods must accurately estimate phenotypes for large numbers of diverse genotypes under field conditions (Araus *et al.*, 2018; Araus and Cairns, 2014; Furbank and Tester, 2011). This represents a significant challenge because genotypic diversity and contrasting environmental conditions are major sources of variation in spectral reflectance. This variation arises mostly from (i) genotype morphology, canopy cover and canopy 3-D structure (Gutierrez *et al.*, 2015; Haboudane *et al.*, 2002; Jacquemoud *et al.*, 2009; Zarco-Tejada *et al.*, 2005), (ii) differences in genotype phenology and the timing of developmental transitions such as heading, flowering and senescence (Kipp *et al.*, 2014; Pimstein *et al.*, 2009; Stuckens *et al.*, 2011) and (iii) reactions to other biotic or abiotic stresses, which may result in similar spectral responses as the disease of interest (Zhang *et al.*, 2012). At present, effects of diseases on canopy reflectance are often investigated at specific points in time (*see e.g.*, Cao *et al.*, 2013; Yang, 2010; Yu *et al.*, 2018). Such investigations have provided valuable but highly context-specific insights (i.e. specific to the genotype, growth stage and/or site and environment under study; *see e.g.*, Delalieux *et al.*, 2007; Zhang *et al.*, 2012; Zheng *et al.*, 2019). Accordingly, identified spectral features and corresponding thresholds or calibration curves are not sufficiently robust (i.e. universally applicable) for use in resistance breeding. Largely due to such difficulties, high throughput phenotyping of disease resistance under field conditions using hyperspectral reflectance is still elusive (Araus *et al.*, 2018).

Septoria tritici blotch (STB) caused by the fungal pathogen *Zymoseptoria tritici* is a serious threat to wheat production in major wheat growing areas around the world (Orton *et al.*, 2011; Torriani *et al.*, 2015). The development of cultivars with improved resistance to this disease has become a significant objective in wheat breeding and constitutes a key component of STB management strategies (Brown *et al.*, 2015; McDonald and Mundt, 2016; O'Driscoll *et al.*, 2014). Several major resistance genes conferring near-complete resistance to certain *Z. tritici* isolates have been identified and used in commercial cultivars (*reviewed by* Brown *et al.*, 2015). However, these genes are frequently overcome within a few years of their introduction due to the high evolutionary potential of *Z. tritici* populations (McDonald and Mundt, 2016). Genetic loci conferring broadly effective partial resistance are thought to be more durable than major

resistance genes (McDonald and Linde, 2002; McDonald and Mundt, 2016). However, sources of partial resistance are much more difficult to identify, as subtle differences in disease severity must be accurately quantified under field conditions, ideally over time.

Automated image analysis has shown great potential to accurately quantify STB resistance and characterize different components of resistance in genetically diverse breeding material (Karisto *et al.*, 2018; Stewart *et al.*, 2016). However, such measurements are more labor-intensive than visual scorings and do not provide the necessary throughput to routinely screen large breeding trials over time. Some recent work has investigated the potential of reflectance-based techniques to detect and quantify STB non-destructively at the leaf and canopy level (Odilbekov *et al.*, 2018; Yu *et al.*, 2018). At the canopy level, the above-mentioned challenges are particularly pronounced in the case of STB, because epidemics frequently reach damaging levels and affect crop performance most during the grain filling phase (Bancal *et al.*, 2007). Consequently, detection and quantification of STB must be achieved in fully developed canopies with a complex architecture and a clear delineation of STB and physiological senescence is essential for efficient selection.

Recently, efforts have been made to increase the specificity of reflectance-based methods. For example, new spectral vegetation indices (SVIs) with improved specificity to diseases have been developed by several authors for various patho-systems (*e.g.*, Ashourloo *et al.*, 2014; Mahlein *et al.*, 2013). Other work has demonstrated that SVI combinations may allow to differentiate between diseases (Mahlein *et al.*, 2010) and to delineate disease symptoms and nitrogen deficiency in wheat (Devadas *et al.*, 2015). Yu *et al.* (2018) investigated the potential of different spectral features to robustly estimate STB severity at the canopy level in a large population of genetically diverse wheat genotypes. Other work has demonstrated that the sequence of temporal changes in hyperspectral reflectance signatures at the leaf level may be disease-specific, allowing to differentiate between sources of biological stress (Mahlein *et al.*, 2012, 2010; Wahabzada *et al.*, 2015).

Here, we aimed to achieve robust reflectance-based detection and quantification of STB under field conditions by exploiting changes in hyperspectral canopy reflectance over time. The basic rationale of the proposed approach is that pathogenesis consists in a specific and fixed sequence of events producing a constant outcome (i.e. disease symptoms). Accordingly, these events and outcomes should result in a specific and fixed sequence of changes in canopy spectral reflectance over time, irrespective of the genotype or environment under study. It seems highly likely that relying exclusively on this type of information increases the robustness of resulting estimations.

To test the feasibility of this approach, we engineered spectral-temporal features based on hyperspectral time series measurements. These features are designed to capture relevant changes in reflectance over time whilst minimizing the effect of the known confounding factors discussed above. We put forward the following hypotheses: (H_1) Confounding effects of contrasting morphology, canopy cover and canopy 3-D structure are strongly reduced, if only relative changes in reflectance over time at the level of individual plots are analyzed. (H_2) Confounding effects of phenology can be reduced by using combinations of STB-sensitive and STB-insensitive spectral

features. Specifically, we hypothesize that several plant traits are relatively unaffected in their temporal dynamics by STB. Thus, related spectral features can be used as a baseline of changes in spectral reflectance over time, arising primarily from advancing crop phenology. This baseline can then be used to correct temporal patterns observed in STB-sensitive features for variation in phenology. Finally, we hypothesized (H_3) that the sequence and the dynamics of STB-sensitive features is to a certain extent specific to this disease and not related to other biotic or abiotic stresses.

Thus, the objective of this study was (i) to evaluate the potential of time-resolved hyperspectral reflectance measurements to detect and quantify STB infections, (ii) to delineate STB and physiological senescence and (iii) to estimate the robustness of the proposed method and hence its potential for breeding applications.

3.2 Materials and Methods

3.2.1 Plant Materials, Experimental Design, Phenology and meteorological data

A field experiment was carried out in the field phenotyping platform (FIP, Kirchgessner *et al*., 2017) at the ETH Research Station for Plant Sciences Lindau-Eschikon, Switzerland (47.449N, 8.682E, 520 m a.s.l.; soil type: eutric cambisol) in the wheat growing season of 2017-2018. A subset of 18 bread wheat (*Triticum aestivum*) genotypes was selected from the GABI wheat panel (Kollers *et al*., 2013; complemented with Swiss cultivars) for contrasting levels of resistance to STB and for contrasting stay-green properties based on previous experiments at the same location (Anderegg *et al*., 2020; Karisto *et al*., 2018). The set comprised morphologically diverse genotypes (e.g. awned and unawned) and there were obvious differences in canopy structural parameters (e.g. flag leaf angle) among the selected genotypes (Figure S 7.1). The study was conducted as a two-factorial experiment in a split-plot design with the presence/absence of artificial pathogen inoculation as a whole-plot factor and genotype as a split-plot factor.

Artificial inoculation with *Z. tritici* spore suspension was done on May 21, 2018. *Z. tritici* strains ST99CH_1A5, ST99CH_1E4, ST99CH_3D1, ST99CH_3D7 were used (Zhan *et al*., 2002; see also http://www.septoria-tritici-blotch.net/isolate-collections.html). Spores were grown for six days in 200ml of YSB liquid media (10g yeast extract and 10g sucrose in 1l water) in several flasks for each strain. The spore suspension was filtered and pooled together for each strain. Spore concentration was adjusted and spores suspensions of each strain were mixed to achieve 150ml of inoculum for each plot containing in total 10^6 spores/ml (2.5×10^5 sp/ml of each strain). Inoculum was sprayed in the evening into wet canopy of each plot.

There were two replications for the whole-plot factor. On the same site, the entire GABI panel was also grown in two replicates (two spatially separated lots). One replication of the inoculated plots each was located in a row adjacent to a lot of the GABI panel, separated by one row sown with the resistant cultivar CH NARA (DSP, Delley, Switzerland). One replication of the non-inoculated control plots each was spatially randomized within a lot of the GABI panel. The experiments were sown with a sowing density of 400 plants m^{-2} on Oct 18, 2017. The plots sown

with the GABI panel (and thus control plots within it) were treated with fungicides on three occasions: (i) Input, Bayer (a mixture of sprioxamin at 300 g/liter and prothioconazole at 150 g/liter) was applied with a dose of 1.25 liter/ha on 23 April, 2018 (BBCH 31), (ii) Aviator Xpro, Bayer (a mixture of bixafen at 75 g/liter and metconazole at 41.25 g/liter) was applied with a dose of 1.25 liter/ha on 14 May, 2018 (BBCH 51), and (iii) Osiris, BASF (a mixture of epoxiconazole at 56.25 g/liter and metconazole at 41.25 g/liter) was applied with a dose of 2.5 liter/ha on 28 May, 2018 (BBCH 65). The inoculated control plots did not receive fungicide treatment. Temperature data was obtained from an on-site weather station. Rainfall data was obtained from a nearby weather station of the federal Swiss meteorological network Agrometeo (www.agrometeo.ch) located at ca. 250 m distance to the field trial. The temperature data was used to calculate growing degree-days (GDD) following

$$Tmean_d = \frac{\sum \frac{maxT_{d.h} + minT_{d,h}}{2} - baseT}{24}$$

$$GDD = \sum_{d=1}^{n} Tmean_d$$

where $Tmean_d$ is the mean temperature for day d after sowing, $maxT_{d,h}$ and $minT_{d,h}$ are hourly maximum and minimum temperatures for day d after sowing and $baseT$ is the base temperature, set to 0°C. Heading date was recorded when 50% of the spikes were fully emerged from the flag leaf sheath (BBCH 59, Lancashire *et al.*, 1991). BBCH scores within the main growth stages were linearly interpolated between assessment dates. Stay-green was assessed visually as described

Table 3.1 Overview of wheat phenology, canopy reflectance measurements, visual scorings and samplings. Visual incidence scorings of *Septoria tritici* blotch (STB), Leaf samplings (Fl1 = Second leaf, Fl0 = Flag leaf), average growth stage (GS), canopy stay-green (Stg Cnp) and Flag leaf stay-green (Stg Fl0) are reported. Days after inoculation (dai) and days after heading (dah) are indicated for each date.

Date	dai	dah	Reflectance	STB Scoring	Leaf samplings	GS	Stg Cnp	Stg Fl0
25.05.2018	4	0				59		
30.05.2018	9	5	x			65		
02.06.2018	11	7	x			73		
05.06.2018	14	10	x					
06.06.2018	15	11		t1				
10.06.2018	19	15	x					
14.06.2018	23	19		t2	t2 (Fl1)		10	10
16.06.2018	25	21	x				10	10
19.06.2018	28	24	x	t3	t3 (Fl0)	77	10	10
20.06.2018	29	25	x				9	10
22.06.2018	31	27	x				8	9
26.06.2018	35	31	x	t4	t4 (Fl0)	79	8	8
30.06.2018	39	35	x			83	3	3
02.07.2018	41	37	x	t5	t5 (Fl0)		1	1
07.07.2018	46	42	x			87	0	0
09.07.2018	48	44	x				0	0
12.07.2018	51	47	x			89	0	0

previously (Anderegg *et al.*, 2020), separately for the flag leaf and the whole canopy, following guidelines provided by Pask *et al.* (2012). Flag leaf stay-green was scored based on the portion of green leaf area on a scale from 0 (0% green leaf area) to 10 (100% green leaf area). An integer mean value was estimated for plants located in a central region of about 0.5 m × 0.5 m of each plot. Canopy stay-green was scored on the same scale by estimating the overall greenness of the plot when inspected at a view angle of approximately 45° considering the entire plot area. All scorings were done by the same person in 2-3 day intervals. An overview of measurements, scorings and samplings is given in Table 3.1. Growth stages were recorded until physiological maturity following the BBCH scale (Lancashire *et al.*, 1991).

3.2.2   Hyperspectral reflectance measurements

Canopy hyperspectral reflectance in the optical domain from 350 to 2500 nm was measured using a passive non-imaging spectroradiometer (ASD FieldSpec® 4 spectroradiometer, ASD Inc., USA) equipped with an optic fiber with a field of view of 25˚. Five spectra were recorded as the average of 15 – 25 separate spectral records while moving the fiber optic once along the diagonal of each plot at a height of approximately 0.4 m above the canopy. A Spectralon® white reference panel was used for calibration before measuring canopy reflectance. The calibration was repeated after measuring one-half of a replicate (i.e. after 9 plots, approximately every 3-5 min). Measurements were carried out on 14 dates between heading and physiological maturity (i.e. between May 30 and July 12, 2018) resulting in an average of one measurement every three days. The maximum distance between two consecutive measurement dates was six days. In parallel, one lot of the GABI panel was measured on 13 dates. Here, the sensor calibration was repeated approximately every 10 min after completion of measurements on two rows.

3.2.3   STB disease assessment

The amount of STB in each plot was assessed on five dates (t1 – t5) between 16 days after inoculation (dai; June 6, 2018) and 42 dai (July 2, 2018). STB was quantified by combined assessments of disease incidence (i.e. the proportion of leaves showing visible symptoms of STB) and conditional disease severity (i.e. the amount of disease on symptomatic plants). Disease incidence was assessed visually for 30 plants per plot by inspecting the leaves of one tiller per plant. Incidence scorings were obtained per leaf layer. Conditional disease severity was then measured using automated analysis of scanned leaves exhibiting obvious disease symptoms. To this end, eight infected leaves were sampled per plot, transported to the laboratory and imaged on flatbed scanners following the method described by Stewart *et al.* (2016) and Karisto *et al.* (2018). However, to avoid interfering excessively with the development of the disease epidemic, leaf samples from inoculated plots were taken only if disease incidence was at least 1/3 (i.e. if at least 10 out of 30 examined leaves exhibited symptoms of STB infection). Thus, no leaf samples were taken at t1, while at t2, second leaves from the top were sampled from a subset of plots. Starting at t3, all plots were sampled at the flag leaf layer. In contrast, from non-inoculated control plots, eight leaves were sampled without reference to their disease status due to very low disease incidence. Automated image analysis was then used to extract the percent of leaf area covered by lesions (PLACL) from the generated leaf scans using thresholds in the HSV color space and

functions of the python API of openCV V3.0.0 (https://opencv.org/). The precision of the automated image analysis method used here to assess STB has been demonstrated repeatedly in greenhouse and field-studies (Krishnan *et al*., 2018; Lendenmann *et al*., 2014; Meile *et al*., 2018; Stewart *et al*., 2018; Zhong *et al*., 2017). The procedure was optimized to minimize the effect of insect damage, powdery mildew infections and physiological senescence, particularly leaf-tip necrosis, on the derivation of PLACL. PLACL was extracted only from t2 and t3 scans, as leaves increasingly displayed physiological senescence at later time-points. The developed python script with detailed annotations can be retrieved from github (https://github.com/and-jonas/stb_placl). Finally, overall disease severity was calculated by multiplying disease incidence with conditional disease severity for inoculated plots, whereas it was directly extracted from the leaf scans for control plots.

3.2.4   Data analysis

All data analyses were done in the R environment for statistical computing (R version 3.5.2; R Core Team, 2018). Raw spectra were smoothed using the Savitzky-Golay smoothing filter (Savitzky and Golay, 1964) with a window size of 11 spectral bands and a third order polynomial, using the R package '*prospectr*' V0.1.3. (Ramirez-Lopez and Stevens, 2014). Spectral regions comprising the wavelengths from 1350 nm to 1475 nm, from 1781 nm to 1990 nm and from 2400 nm to 2500 nm were removed because of the very low signal-to-noise ratio resulting from high atmospheric absorption. Spectra were averaged for each experimental plot. Pre-processed spectra, consisting of reflectance values at 1709 wavelengths, were then used for time-point specific analysis as well as for time-integrated analyses, as described in the next sections. For ease of notation, the reflectance at a specific wavelength will be abbreviated by R followed by the wavelength in nm (e.g. R750).

Benchmark time-point specific analysis
The relationship between spectral reflectance and STB was studied on a diverse panel of wheat genotypes by Yu *et al*. (2018), but the analysis was limited to single time-points. We performed a comparable analysis for each measurement time-point as a benchmark and to estimate model transferability across time. Yu *et al*. (2018) reported improved prediction of STB severity metrics and classification accuracy when using the full spectrum rather than single SVIs. Therefore, our analysis focuses on these approaches. We tested two parametrically structured linear models (Partial Least Squares (PLS) regression and ridge regression) and two tree-based ensemble models (random forest regression and cubist regression) for their capability to predict STB severity metrics. For classification, Partial Least Squares Discriminant Analysis (PLSDA) was used (*for details on these methods we refer to* Kuhn and Johnson, 2013 *and citations therein*). Prior to modelling, spectral resolution was reduced to 6 nm by binning (i.e. by computing average values for six adjacent wavelengths) due to very high correlation of reflectance values at neighboring wavelengths. Following a standard procedure (Kuhn and Johnson, 2013), model hyperparameters were tuned using 10-times repeated 10-fold cross-validation. Thus, training and test datasets comprised 90% and 10% of the original dataset, respectively. The root mean square error of predictions (RMSE) and overall classification accuracy as performance metrics for regression and classification, respectively. The overall accuracy reflects the agreement between

the predicted and the observed classes. This agreement can also be expressed in terms of sensitivity and specificity of the model, with

$$Sensitivity = \frac{\#plots\ correctly\ predicted\ as\ diseased}{\#diseased\ plots}\ and$$

$$Specificity = \frac{\#plots\ correctly\ predicted\ as\ healthy}{\#healthy\ plots}.$$

The simplest model with a performance within one standard error of the absolute best model was chosen as the final model. Variable importance for the projection (VIP) was computed for PLSDA models to estimate the importance of individual wavebands to predict the class (i.e. 'healthy' or 'diseased'). In the regression setting, two different training datasets were used for model fitting: the full dataset, including all control plots, and a dataset consisting of the inoculated plots only. When all control plots were used for model fitting, the RMSE and R^2 of the resulting models was adjusted by removing the predicted and observed values for the control plots again, in order to avoid overly optimistic performance estimates resulting from a good prediction of disease severity in control plots. The R packages '*caret*' V6.0.80 (Kuhn, 2008), '*mixOmics*' V6.3.2 (Rohart *et al.*, 2017), '*pls*' V2.7.0 (Mevik *et al.*, 2018), '*Cubist*' V0.2.2 (Kuhn *et al.*, 2018), '*ranger*' V0.10.1 (Wright and Ziegler, 2017) and '*elasticnet*' V1.1.1 (Hastie, 2018) were used for the analysis.

Time-integrated analysis

Summarizing H_1-H_3, we hypothesized that the analysis of temporal dynamics in hyperspectral reflectance signatures may facilitate a robust detection and quantification of STB across diverse wheat genotypes under field conditions. To evaluate these hypotheses, we condensed the hyperspectral time series into time series of SVIs, similar to the procedure described previously
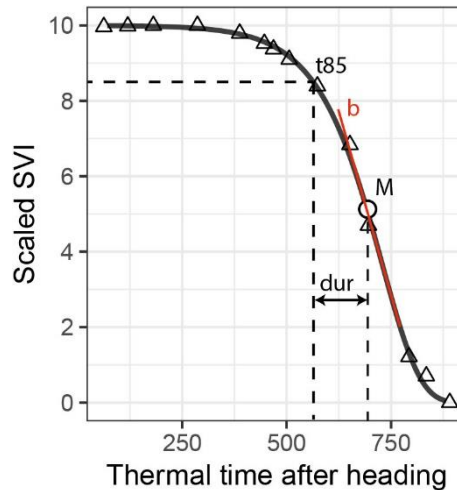

Figure 3.1 Extraction of dynamics parameters for one spectral vegetation index (SVI; here scaled values of the Plant Senescence Reflectance Index) and one experimental plot (here a non-inoculated control plot). The t_{85} parameter is the time point when fitted scaled SVI values decrease to 8.5; M is a parameter of the Gompertz model, representing the time point when the rate of decrease is at its maximum; the dur parameter represents the duration in thermal time between t_{85} and M; b is a parameter of the Gompertz model, representing the maximum rate of decrease. M and t_{85} are labelled 'key points', dur and b are labelled 'change parameters'.

(Anderegg *et al*., 2020). Thereby, we obtained a comprehensive summary representation of the hyperspectral dataset collected over time, interpretable in terms of plant physiology and canopy characteristics. The smoothness of SVI values over time was evaluated graphically and only SVIs showing a clear and interpretable temporal trend were maintained for further analyses. Values of the selected SVIs were scaled to range from 0 to 10, representing the minimum and maximum values recorded during the assessment period for the corresponding experimental plot, respectively. To simplify subsequent steps in the analysis, the scale for SVIs with increasing values over time was inverted. Measurement dates were converted to thermal time after heading by subtracting the plot-specific accumulated thermal time at heading from the accumulated thermal time at each measurement date. The scaled SVI values were then fitted against thermal time after heading for each experimental plot and SVI using a Gompertz model with asymptotes constrained to 0 and 10 (eq. 1).

$$S = 10\exp[-\exp[-b * (t - M)]] \text{ (eq. 1)}$$

where S represents the scaled SVI value, t is the accumulated thermal time after heading, b is the rate of change at time M and M is the accumulated thermal time after heading when the rate is at its maximum (Gooding *et al*., 2000). Eq. (1) was fitted using the R package '*nls.multstart*' V1.0.0 (Padfield and Matheson, 2018). Two types of dynamics parameters for each experimental plot and SVI were extracted from the resulting model fits: (1) 'key time-points', which are specific points in time when a certain criterion (e.g. a threshold) is met; and (2) 'change parameters', which represent the rate or duration of a process (Figure 3.1). We extracted two key time-points: the M parameter of the Gompertz model, and the time when fitted values decreased to 8.5 (t_{85}). As change parameters we used the rate parameter b of the Gompertz model, and the duration between t_{85} and M (dur). While the b and M parameters of the Gompertz model fully describe the fitted curve, the t_{85} and dur parameters are affected by both Gompertz model parameters, thus representing a mix of both. The threshold was set to 8.5 because (i) this level efficiently captured
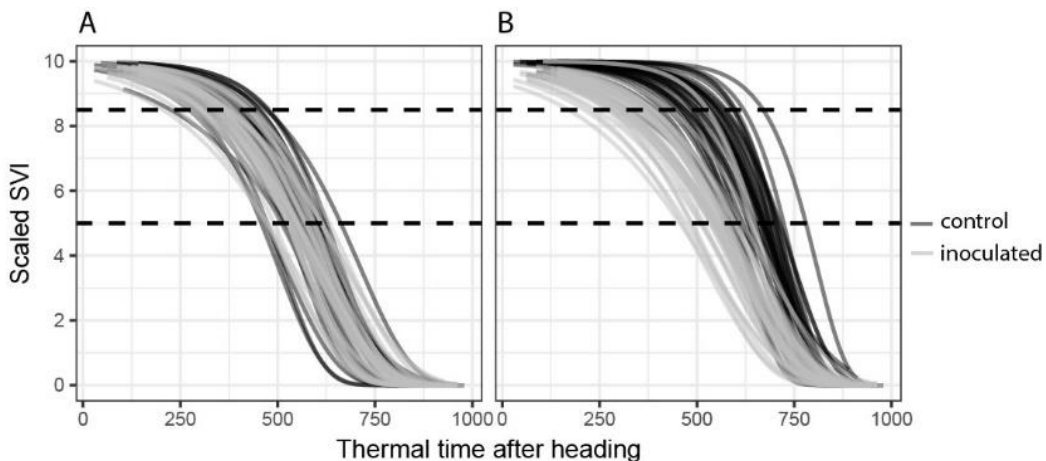

Figure 3.2 Temporal dynamics of spectral vegetation indices (SVIs). Gompertz model fits for all 72 experimental plots are shown. **(A)** A disease-insensitive SVI (here the Flowering Intensity Index, FII) displays the same temporal patterns for both treatments (inoculated and control). **(B)** A disease-sensitive SVI (here the Modified Chlorophyll Absorption Ratio Index, MCARI2) displays contrasting temporal patterns for control and inoculated plots.

observed variation during the late stay-green phase (Figure 3.2), (ii) it was little affected by somewhat unstable values during the stay-green phase observed for some SVIs, and (iii) for some SVIs, the initial highest values were not optimally represented by the Gompertz model.

Next, SVIs least affected in their temporal dynamics by the presence or absence of STB infections were selected separately for each dynamics parameter as follows: for the key time-points (t_{85}, M) by selecting SVIs with the smallest average difference between the parameter values of the inoculated and non-inoculated control plots; for change parameters (b, dur) by selecting SVIs with the smallest average deviance from 1 of the ratio of the change parameter values. For each dynamics parameter, a subset of eight SVIs with the smallest difference or ratio was selected. All other SVIs were considered to be significantly affected by STB infection. Figure 3.2 shows an example of an STB-sensitive and an STB-insensitive SVI.

We then performed unsupervised subset selection (i.e. without considering the response) on both sets of SVIs (i.e. the STB-sensitive and STB-insensitive SVIs) with the aim of removing redundant SVIs. For each dynamics parameter (t_{85}, M, dur, b), pairwise Pearson correlation coefficients between the parameter values derived from all used SVIs were computed. For change parameters, the maximum linear correlation allowed was set to $r = 0.9$, whereas for the key time-points it was set to $r = 0.95$, as these were generally highly collinear. In cases where pair-wise correlations were higher than these threshold values, only one of the two SVIs was retained, preferring narrow-band SVIs over broad-band SVIs, SVIs with a specific physiological interpretation and SVIs developed specifically for use in wheat or barley canopies over more generic SVIs. Additionally, the goodness of the Gompertz model fit was evaluated qualitatively (i.e. graphically) and used as an additional selection criterion. The parameters were then combined by calculating differences between the key time-points derived from selected STB-sensitive and STB-insensitive SVIs and the ratios of the change parameters derived from STB-sensitive and STB-insensitive SVIs (Figure 3.3). These differences and ratios were calculated for all possible
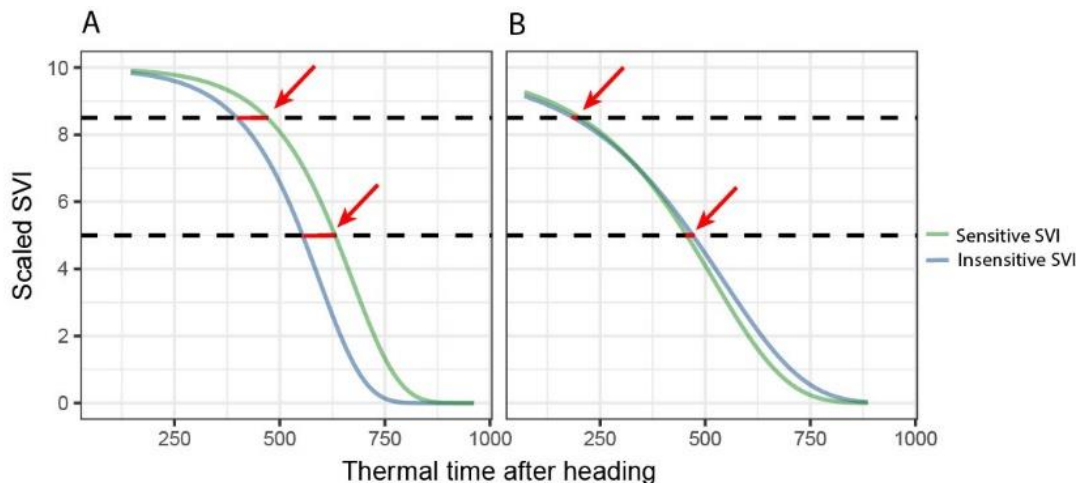

Figure 3.3 Derivation of the final key time points based predictors for disease classification and quantification. Key time points extracted from disease-sensitive and disease-insensitive spectral vegetation indices (SVI) are combined to isolate the effect of the disease from other effects (e.g. contrasting stay-green) by calculating the differences (highlighted by red arrows). **(A)** Control, 'healthy' plot, **(B)** Inoculated, 'diseased' plot. For change parameters, the ratio, rather than the difference, was calculated.

pairs of STB-sensitive and STB-insensitive SVIs and were then used as features for (1) the classification of plots into non-inoculated healthy control plots and inoculated, diseased plots and (2) the prediction of STB severity in each plot. This final step was performed primarily to estimate whether a combination of features outperform single features in predicting STB severity (relevant with respect to H_3), and to identify the most predictive features. Models and model fitting procedures were identical to the time-point specific analysis.

Selection of spectral-temporal features

While tree-based models are considered naturally resistant to non-informative predictors, and some perform feature selection intrinsically, the presence of highly correlated predictors makes the interpretation of resulting variable importance measures challenging (Strobl *et al.*, 2007). Hence, we performed supervised feature selection by recursive feature elimination with cubist and random forest regression as base-learners, using a nested cross-validation approach. The dataset was resampled 30 times with an 80:20 split using stratified sampling. Samples were binned into eight classes based on percentiles of STB severity to ensure balanced evaluation datasets. Thus, for each resample, feature elimination was carried out on 80% of the data, and model performance was evaluated on the remaining 20% in 28 decreasing steps. Eliminated features were assigned a rank corresponding to the iteration after which they were excluded (i.e. those eliminated first had rank 28, whereas the feature retained as the last had rank 1). In each iteration, the base-learner hyperparameters were tuned using 10-fold cross-validation (*see* Ambroise and McLachlan, 2002; Granitto *et al.*, 2006; Guyon *et al.*, 2002 *for a detailed discussion of the methodology*).

Independent model validation

Due to the relatively small size of the experiment (n = 72 experimental plots) we did not rely exclusively on the cross-validated training performance estimates for model evaluation. The performance and robustness of the developed models was further evaluated using data of 360 wheat plots sown with 330 registered varieties obtained from one replication of the GABI panel. Low to intermediate disease incidence and very low conditional disease severity as well as late appearance of symptoms in all 36 control plots spatially randomized within the two replications of the panel suggested that STB disease should not have reached damaging levels in the vast majority of these plots. Therefore, these plots were considered as essentially disease-free. For all of these plots, spectral-temporal features were extracted from the 13 measurement time-points as described previously and were then used to generate a class label and class probabilities from the classification models as well as a prediction of disease severity from the regression models. To distinguish the performance measures obtained for held-out samples of the main experiment (i.e. the cross-validated training performance) from those obtained for the independent plots, these are referred to as the internal accuracy (accint) and the generalized accuracy (accgen), respectively. It is important to note that accuracy represents only model specificity in this case, as no independent plots with significant levels of STB were available. In a final validation step, the spatial distribution of class labels and severity predictions were examined by creating plots of the experimental design. Thus, we aimed to test the robustness of the models to heterogeneous field conditions. Field heterogeneity may affect plant physiology and thus hyperspectral reflectance

over time (e.g. through the development of local drought stress). The presence or absence of spatial patterns in model predictions can therefore be interpreted as an additional measure of model robustness.

Validation of the most predictive feature in a contrasting environment
Spectral-temporal features were engineered specifically to minimize effects of genotype and environment. However, it is still conceivable that relative changes in spectral reflectance over time are also affected by environmental conditions. This may result in the extraction of environment-specific relationships between spectral-temporal features and disease severity. We therefore evaluated the relationship between the most predictive spectral-temporal feature and STB severity using data from a separate year. This dataset enables a rigorous validation of the spectral-temporal features as predictors of STB for several reasons: (i) it originates from a strongly contrasting environment with wet and cool weather conditions, (ii) sowing parameters were different, likely affecting canopy structural parameters, (iii) only natural variance in disease resistance was observed, as artificial inoculations were not performed, (iv) a large number of morphologically, phenologically and structurally diverse genotypes not contained in the training population were assessed, (v) reflectance measurements were not carried out with a sufficient frequency to allow fitting of parametric models, and we had to use linear interpolations of individual measurement time points instead, losing the advantage of the smoothing effect, finally (vi) sampling procedures to quantify STB severity were not optimal for our purpose. As parametric models could not be fitted, the time point when interpolated values decreased below 50% of their initial value (t_{50}) was extracted as an equivalent to the *M* parameter of the Gompertz model. For more details on datasets and experiments we refer to Karisto *et al.* (2018) and Anderegg *et al.* (2020).

3.3 Results

3.3.1 Development of STB disease

Towards the end of the vegetation period, all inoculated plots had substantial levels of STB. In contrast, non-inoculated control plots were essentially disease-free until late in the vegetation period. Thus, artificial inoculations were effective in all plots and the dataset was suitable for testing the feasibility of the classification of plots into healthy and diseased canopies based on reflectance spectra or spectral-temporal features. Furthermore, large variation in the levels of STB could be observed among the inoculated plots, probably attributable to different levels of resistance, with the largest variation observed during late stay-green (i.e. at t3, June 19, 2018, compare with Figure 3.4B). Thus, the dataset was also suitable for testing the feasibility of disease quantification using reflectance spectra or spectral-temporal features.
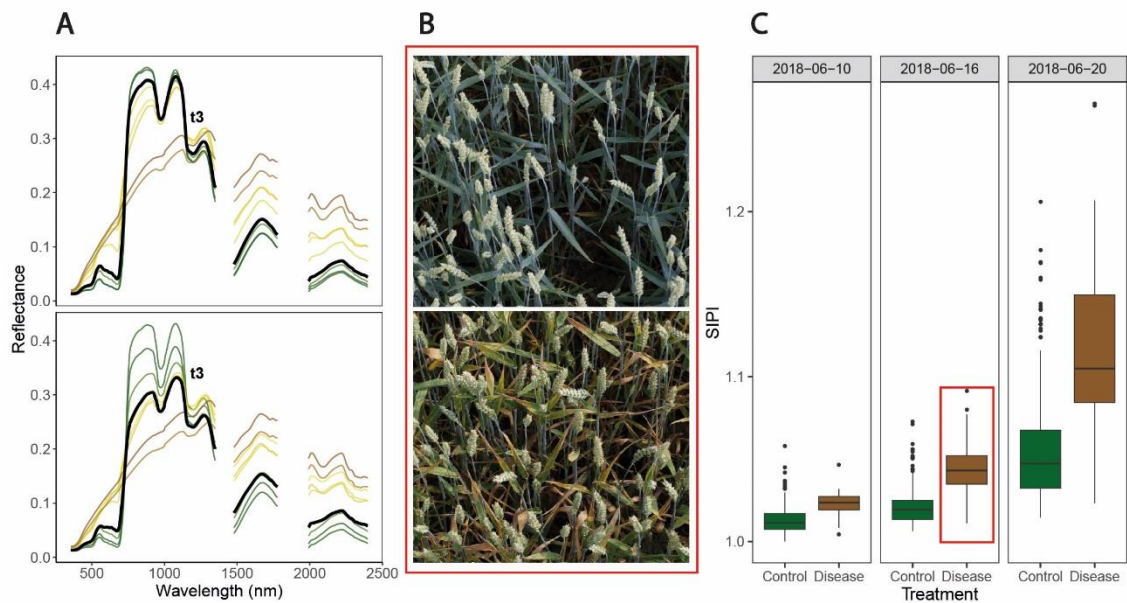
Figure 3.4 Symptoms of *Septoria tritici* blotch (STB) and associated spectral reflectance characteristics over time. **(A)** Date-wise averaged reflectance spectra of healthy canopies (upper panel) and of inoculated, diseased canopies (lower panel). Colors approximate the average color of the vegetation on the corresponding measurement date (estimated based on average visual canopy senescence scorings). The thick black spectra mark the average reflectance spectra measured at t3 (i.e. June 19, 2018). **(B)** Images of two inoculated plots, taken on June 16, 2018. The genotype in the upper panel was highly resistant to STB, developing visible symptoms only later, whereas the genotype in the lower panel was highly susceptible and displays severe symptoms of STB. Images were captured using the Field Phenotyping Platform (FIP, Kirchgessner *et al.*, 2017). **(C)** Values of the Structure Insensitive Pigment Index (SIPI) for both treatments on three measurement dates. The plots shown in panel B are contained in this boxplot (indicated by the red box). No obvious signs of apical senescence were visible by June 20, 2018 in any of the healthy control plots, but senescence started shortly after.

High levels of STB in inoculated plots were the result of both high incidence and high conditional severity, particularly at t4 and t5 (Figure 3.5A, Table 3.2). Visual assessments of scanned leaves suggested a high conditional severity in all inoculated plots at these late stages. In contrast, the non-inoculated control plots displayed very low levels of STB even at late stages. STB incidence increased in some plots at t4 and t5, but visual assessments of the sampled leaves demonstrated very low conditional disease severity even at t5. Since the subset of genotypes used for the experiment also included some highly susceptible genotypes, this suggests that natural infections did not cause agronomically significant levels of STB in this experiment. This was likely the result of fungicide applications and the very low rainfall in the period from May to July. Rainfall in this period totaled 178 mm, which represents 52% of the average of 343 mm in the reference period 1981–2010 (MeteoSwiss, 2019).
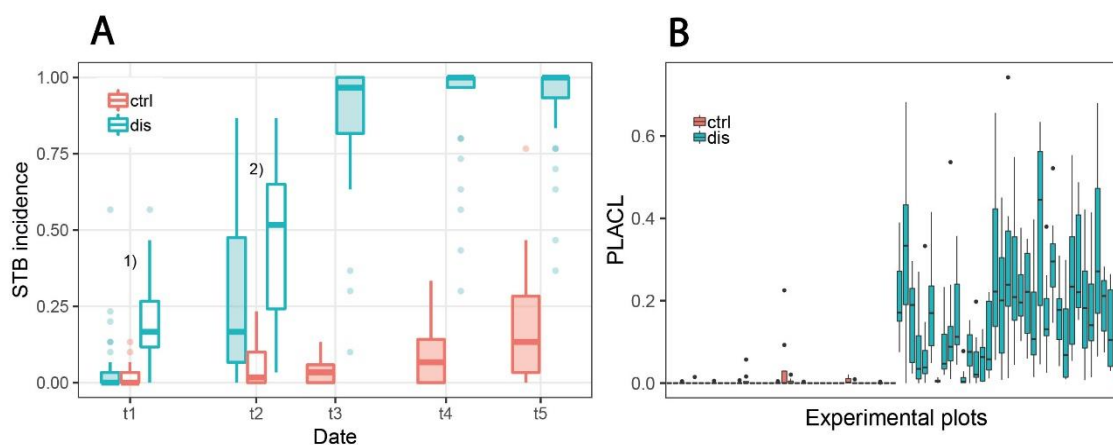
Figure 3.5 Development of *Septoria tritici* blotch (STB) disease. **(A)** STB incidence at five different assessment dates for non-inoculated 'healthy' control plots (ctrl) and for artificially inoculated 'diseased' plots (dis). In diseased plots, STB incidence on flag leaves was assessed at all time points, whereas for the control plots, it was assessed only from t3 onwards. Filled boxes represent STB incidence on flag leaves; open boxes represent STB incidence on lower leaf layers. [1]Open boxes represent STB incidence on third leaves from the top, [2]open boxes represent STB incidence on second leaves from the top. **(B)** Conditional disease severity measured as percent leaf area covered by lesions (PLACL) at t3 (June 19, 2018) for eight flag leaves per plot for all 72 experimental plots.

STB incidence was low at t1, both in inoculated and in control plots. Symptoms were apparent at significant levels only on lower leaf layers of inoculated plots, whereas flag leaves were essentially disease-free in both treatments. At t2, there was approximately a five-fold increase in STB incidence at the flag leaf and subtending leaf layer in many inoculated plots (Figure 3.5A, Table 3.2). At t3, STB incidence on flag leaves reached very high levels in most inoculated plots, and PLACL reached an average of 17%, indicating a moderate conditional STB severity on average. Thus, the observed differences in STB severity among inoculated plots is primarily a result of differences in conditional severity, with PLACL ranging from 0% to 38% (Figure 3.5B). In control plots, almost no lesions were detected. There were some signs of physiological

Table 3.2 Summary of *Septoria tritici* blotch (STB) assessments. STB incidence and conditional severity was assessed at the level of individual leaf layers, namely the flag leaf (Fl0), the penultimate leaf (Fl1) and the ante-penultimate leaf (Fl2). Conditional severity was measured as percent leaf area covered by lesions (PLACL). Severity was calculated as the product of STB incidence and conditional severity. Values are reported separately for non-inoculated control plots and inoculated plots, separated by a slash. Mean values across all plots are reported, with minima and maxima in brackets. Disease assessments were carried out on five dates (t1 – t5) covering the growth phases of 15 days after inoculation (dai) to 41 dai.

Time point	dai	STB Incidence			PLACL		STB Severity	
		Fl0	Fl1	Fl2	Fl0	Fl1	Fl0	Fl1
t1	15	- / 0.05 (0, 0.57)	- / 0.10 (0, 0.53)	0.02 (0, 0.13) / 0.20 (0, 0.57)	- / -	- / -	- / -	- / -
t2	23	- / 0.26 (0, 0.87)	0.05 (0, 0.23) / 0.45 (0.03, 0.87)	- / -	- / -	- / 0.05 (0, 0.25)	- / -	- / 0.03 (0, 0.20)
t3	28	0.03 (0, 0.13) / 0.87 (0.10, 1)	0.09 (0, 0.50) / 0.91 (0.10, 1)	- / -	- / 0.17 (0, 0.38)	- / -	**0 (0, 0.04) / 0.16 (0, 0.38)**	- / -
t4	35	0.09 (0, 0.33) / 0.92 (0.30, 1)	- / -	- / -	- / -	- / -	- / -	- / -
t5	41	0.17 (0, 0.77) / 0.92 (0.37, 1)	- / -	- / -	- / -	- / -	- / -	- / -

senescence on sampled flag leaves at t3, but these were mostly limited to yellowing of the entire leaves and/or leaf tip necrosis. As there was ample variation for disease severity at t3 among the inoculated plots, and physiological senescence did not significantly affect extraction of PLACL from leaf scans, this time-point was chosen as a measure of overall disease severity and used as response variable in the time-integrated analysis.

3.3.2 Effects of STB and phenology on canopy spectral reflectance

Over the assessment period, observed changes in spectral reflectance were similar for diseased and healthy canopies (Figure 3.4A), showing the typical pattern of senescing canopies. However, an obvious effect of STB infections consisted in an early marked decrease in reflectance in the NIR not observable in healthy canopies. This decrease preceded the appearance of physiological senescence and was observable in the pre-symptomatic phase of STB infections. Furthermore, an early increase in reflectance in the VIS, especially for wavelengths greater than 535 nm, was observed. An early increase in SWIR reflectance in diseased canopies compared to healthy canopies was also discernable. However, these differences were small compared to changes in reflectance over time.

Canopy spectral reflectance seemed to remain relatively constant throughout the stay-green phase in healthy canopies (Figure 3.4A, upper panel). However, the examination of SVI values over time revealed significant changes in canopy reflectance during this period (Figure 3.4C). Importantly, variation caused by advancing phenology (i.e. within-treatment variation in Figure 4C) was prominent with respect to STB-induced variation (i.e. between-treatment variation in Figure 3.4C). This is true even for the structure insensitive pigment index (SIPI), which has been proposed as a potential surrogate for crop disease under field conditions (Yu *et al*., 2018; Figure 3.4C). For several other spectral indices, initial variation as well as variation over time was even larger (data not shown).

3.3.3 Time-point specific full-spectrum analysis

Binary classification into healthy and diseased canopies using reflectance spectra
PLSDA models correctly classified all held-out samples in most resampling iterations, resulting in classification accuracies $\text{acc}_{\text{int}} \geq 0.96$ for all time-points (Figure 3.6). The optimal number of components used by the PLSDA models (determined *via* repeated CV) was between 5 and 17, depending on the time-point. Correct class labels were obtained for all held-out samples even for the first time-point at 9 dai, when no visual symptoms of STB were present in most experimental plots. However, prediction accuracies for the independent GABI plots were distinctly lower, particularly for models calibrated with data from early and late measurement time-points (data not shown). A satisfactory performance on independent plots was observed for models calibrated with data from the late stay-green phase (i.e. between 2018-06-10 and 2018-06-22), which correctly predicted the independent plots as disease-free in most cases ($\text{acc}_{\text{gen}} \geq 0.88$ in all cases).

VIP scores quantify the importance of wavebands to predict the response, i.e. to generate the class label ('healthy' or 'diseased') or to predict STB disease severity here (Yu *et al*., 2018). VIP scores for the first three components showed some general patterns across time-points (Figure 3.6). The

near-infrared region (NIR, 750-1300 nm) and the short-wave infrared region (SWIR; 1475-1781 nm and 1991-2400 nm) had a relatively high importance (Figure 3.6). However, the relative importance of the SWIR compared to the NIR drastically changes over time. The importance of the SWIR is comparably low during early stay-green, but its importance greatly increases and exceeds the importance of the NIR during late stay-green. Furthermore, the red-edge (RE, 680-750nm) had a low importance at the beginning, but is increasingly important at later stages, as indicated by a gradual left-shift of the peak in VIP at the NIR for later time-points. Finally, at early time-points, there is a significant contribution of wavebands in the visible range (VIS, 400-700nm). This feature is somewhat transformed over time, resulting in a narrow peak in VIP at wavelengths around 535nm at intermediate time-points. Towards later time-points, this feature broadens again.
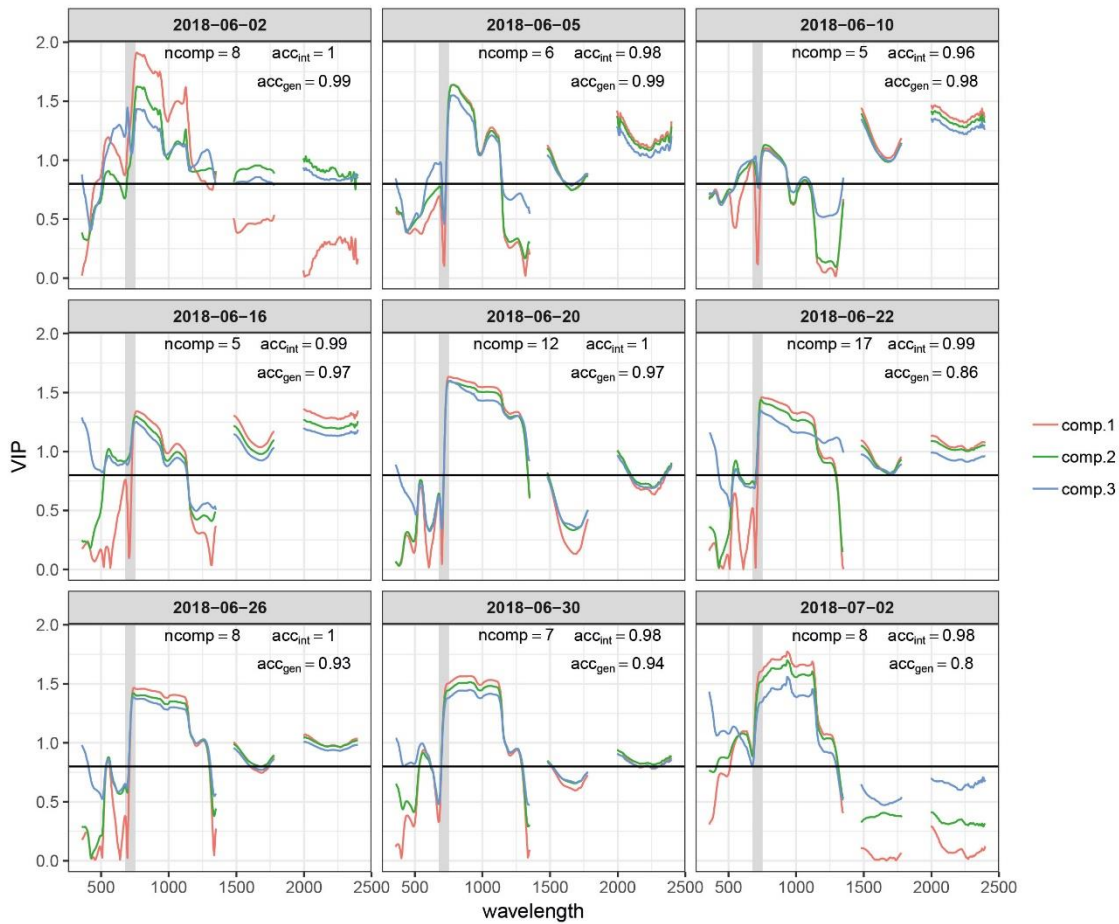
Figure 3.6 Variable importance for the projection (VIP) of the time point specific partial least squares discriminant analysis (PLSDA) models for the first three components (comp.1 – comp.3). The total number of components used by the model (ncomp), the prediction accuracy for plots included in the experiment (acc_{int}) and prediction accuracy for the independent test set comprising >300 plots of an adjacent experiment measured on the same day or with a maximum delay of one day (acc_{gen}) are also reported for each model. The earliest and latest time points are not represented. The grey shaded area represents the spectral range between 680 nm and 750 nm, i.e. the red edge. The horizontal black line marks a commonly used threshold value for an important contribution (VIP = 0.8).
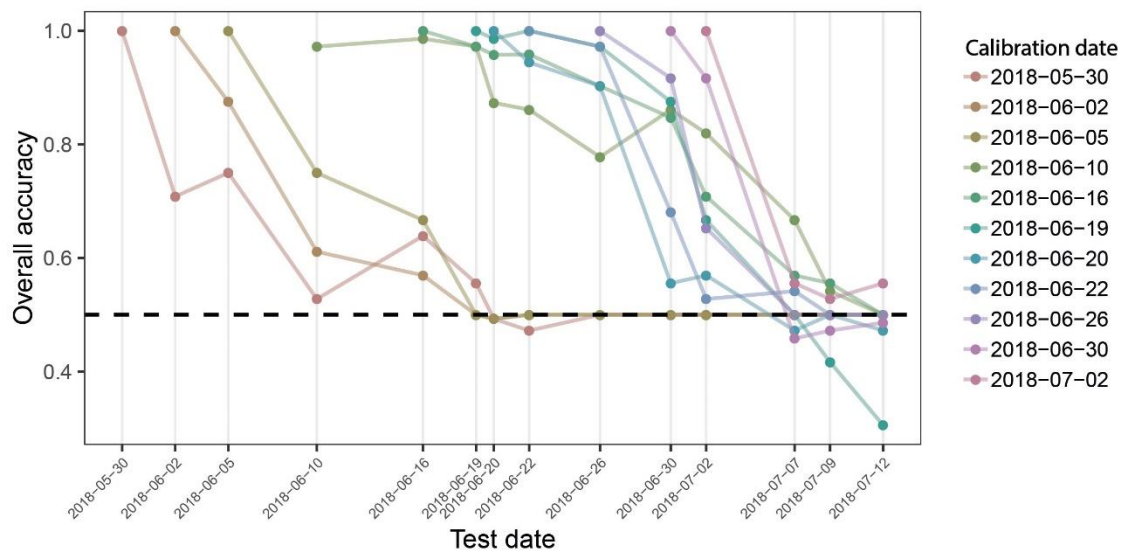
Figure 3.7 Overall prediction accuracy of PLSDA models across time. Binary classification models were calibrated for each measurement time point, using reflectance spectra collected on this date as predictors and the known class label (i.e. "diseased" or "healthy") as response. The performance of these models was evaluated on held out samples of the same date as well as on all plots of the subsequent measurement time points as the overall accuracy of classification. Colored lines track the performance of each date-specific model across time (e.g. the left-most red line represents the performance of the model calibrated with spectra obtained on May 30, 2018, when tested on the same date, and for all subsequent measurement dates). The broken black line indicates a performance of 0.5, i.e. the performance of a random guess of the class label.

Given the common patterns but also significant differences across time-points, we aimed to evaluate the robustness of the developed models to temporal changes in reflectance induced by advancing crop phenology, as differences in crop phenology are typically present among genotypes in breeding programs. Model performance across time is shown in Figure 3.7. The performance of models calibrated with data from early and late time-points quickly deteriorates. In contrast, models calibrated with data of intermediate time-points show a higher stability over time, although the performance of some models still decreases rather fast. The models created using data from 2018-06-10 and 2018-06-19 were most robust over time, and produce accurate class predictions over a period of about 10 days. It is essential to note that these performance estimates are derived from the same plots used to calibrate the models, although at different time-points. Given the lower acc$_{gen}$ compared to the acc$_{int}$ (see above), significantly decreased performance should be expected on entirely independent plots (different genotypes).

Regression models to quantify disease severity using reflectance spectra

Cubist regression models performed best in predicting disease severity. The smallest RMSE was obtained for models trained on data from inoculated plots only (RMSE = 0.061, R^2 = 0.67; Figure 3.8). The underlying model was simple, building on only four variables (R748, R766, R892 and R1084) in a single model tree. Model performance was slightly decreased when all available data was used for model fitting (RMSE$_{adj}$ = 0.066, R^2_{adj} = 0.55). Here, significant improvements were achieved by increasing model complexity. Validation on the largely disease-free plots of the GABI panel suggested a high specificity of the model (i.e. disease levels on healthy plots were predicted to be virtually zero for almost all plots). This was true irrespective of whether the control
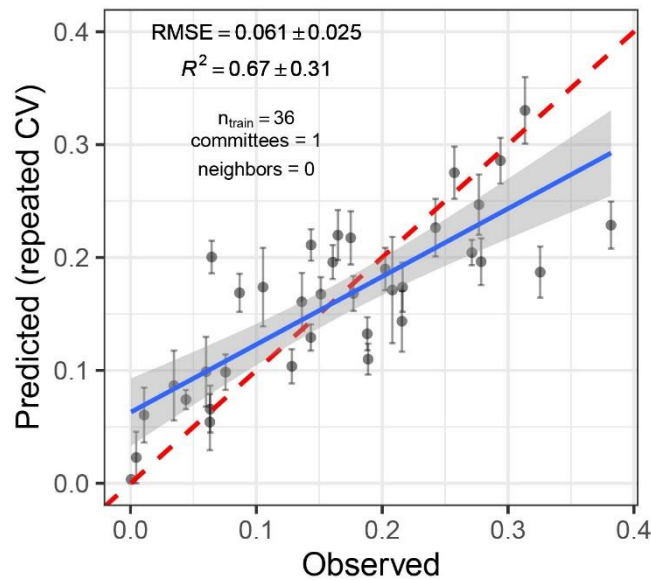
Figure 3.8 Predicted vs. observed *Septoria tritici* blotch (STB) severity levels of 36 artificially inoculated wheat plots. Mean and standard error of predictions are shown. STB severity was measured on flag leaves using a combination of visual incidence scorings and scans of flag leaves exhibiting disease symptoms. Predictions were obtained from a cubist regression model based on the reflectance spectrum of the canopies measured on June 19, 2018. The broken red line represents the 1:1 line, the blue line represents the least squares line of the linear regression of predicted vs. observed values, the gray area represents the 95% confidence interval of the least squares line.

plots were included in the training dataset or not. PLS and ridge regression performed similarly (RMSE = 0.063, R^2 = 0.64), whereas random forest regression performed comparably poorly (RMSE = 0.087, R^2 = 0.46). A strong systematic bias was observed in predictions of the random forest, with low values of disease severity overestimated and high values underestimated.

3.3.4 Time-integrated analysis using combinations of spectral vegetation indices

Engineering of spectral-temporal features as new predictors

Fifty-seven of the tested SVIs were deemed amenable for analysis of their temporal dynamics in the proposed framework (i.e. they displayed a clear and interpretable temporal trend, which could be represented using a Gompertz-type model). For seven of these, the last three measurements were excluded prior to modelling their temporal dynamics, as values increased again in later stages. After subset selection, seven and 13 SVIs were retained as insensitive and sensitive SVIs, respectively, for the key time-points. For change parameters, four and twelve SVIs were retained as insensitive and sensitive SVIs, respectively. In total, 24 distinct SVIs were retained, of which ten sensitive and 14 insensitive SVIs. From their fitted dynamics, a total of 278 (i. e. 4 insensitive SVIs * 12 sensitive SVIs * 2 parameters + 7 insensitive SVIs * 13 sensitive SVIs * 2 parameters) spectral-temporal features were generated as pairwise combinations of dynamics parameters obtained from sensitive and insensitive SVI. These features were then used for classification and regression, as described below.

Binary classification into healthy and diseased canopies using spectral-temporal features

A PLSDA model using 4 components achieved a classification accuracy $acc_{int} = 1.00$, suggesting correct classification of each experimental plot as healthy or diseased canopy based on spectral-temporal features. In the external validation, the model achieved $acc_{gen} = 0.86$, thus correctly classifying 304/353 plots as healthy. This is slightly less accurate than the best time-point specific models (Figure 3.6).

Regression models to quantify disease severity using spectral-temporal features

Overall, disease severity predictions from spectral-temporal features were similarly accurate as those obtained from time-point specific models based on reflectance spectra. The lowest RMSE was achieved when no control plots were used as training data using the PLS algorithm (RMSE = 0.068, $R^2 = 0.71$). Differences in performance among algorithms were smaller than in time-point specific analyses, but the random forest performed relatively poorly also in this case (RMSE = 0.076, $R^2 = 0.62$). Both tree-based models, and particularly the random forest, produced systematically biased predictions, which was not observable for PLS and ridge regression. In contrast to the time-point specific analyses, validation on the largely disease-free plots of the GABI panel suggested the necessity to include the control plots in the training data in order to obtain accurate predictions for low levels of disease, except for ridge regression. When these were included, tree-based models produced good estimates of the low disease-levels, whereas PLS and
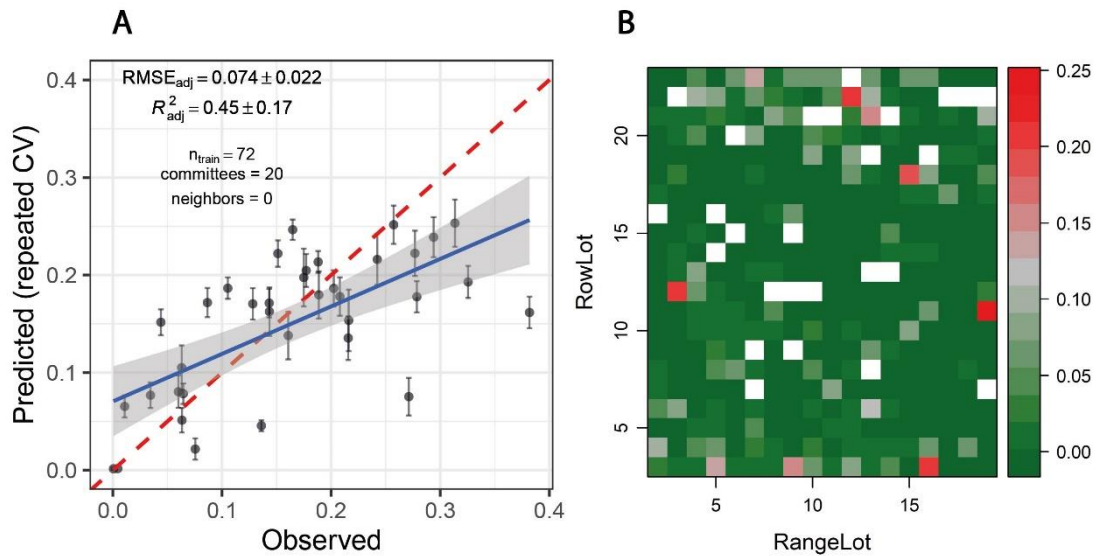

Figure 3.9 (A) Predicted vs. observed *Septoria tritici* blotch (STB) severity levels of 36 artificially inoculated wheat plots. Mean and standard error of predictions are shown. Data from all experimental plots (n = 72) was used to tune/train the model, but reported performance estimates are based only on artificially inoculated plots (n = 36) in order to avoid overly optimistic performance estimates resulting from a good prediction of disease severity in control plots. STB severity was measured on flag leaves using a combination of visual incidence scorings and scans of flag leaves exhibiting disease symptoms. Predictions were obtained from a cubist regression model based on spectral-temporal features for the same 36 plots and 36 non-inoculated control plots sown with the same genotypes. The broken red line represents the 1:1 line, the blue line represents the least squares line of the linear regression of predicted vs. observed values, the gray area represents the 95% confidence interval of the least squares line. **(B)** Spatial distribution of predicted STB severity levels of ~360 largely disease-free plots of the GABI wheat panel, grown next to the plots used as training dataset. White fields correspond to the plots contained in the training dataset.

Table 3.3 Spectral vegetation indices (SVI) identified to be insensitive in their temporal dynamics to the presence or absence of *Septoria tritici* blotch (STB) disease. For each dynamics parameter, a subset of eight SVI with the smallest difference (diff, GDD) or ratio (dimensionless) of the parameters between treatments was selected. The reported SVI were retained after subset selection based on pairwise correlations. The mean pairwise correlation (corr) is reported per dynamics parameter. Values in brackets report the minimum and maximum pairwise correlation.

SVI	DynPar	diff/ratio	Pairwise corr
R780/R740	t85	75.3	
CHLRE	t85	55.6	
DCNI	t85	45.6	
FII	t85	33.2	0.66 (0.26, 0.95)
GNDVI	t85	81.3	
PRInorm	t85	58.2	
SIPI	t85	45.3	
R780/R740	M	37	
CHLRE	M	55.6	
DCNI	M	25.7	
FII	M	4.4	0.82 (0.60, 0.95)
GNDVI	M	40.5	
Prinorm	M	21.6	
SIPI	M	22.7	
R780/R750	b	1.03	
CHLRE	b	0.98	
MTCI	b	0.95	0.73 (0.48, 0.88)
SR	b	0.99	
R780/R750	dur	1.03	
CHLRE	dur	1	
MTCI	dur	1.02	0.75 (0.53, 0.88)
SR	dur	0.99	

ridge regression still predicted disease severity of > 0.05 in a significant number of plots. The inclusion of the control plots resulted in a lower systematic bias of the tree-based predictions, while only marginally decreasing model performance ($RMSE_{adj} = 0.074$ and $RMSE_{adj} = 0.076$ for cubist and random forest, respectively; Figure 3.9A). Importantly, the inclusion of the control plots also strongly reduced or eliminated spatial patterns in predictions of the GABI panel, except for ridge regression (Figure 3.9B). Thus, cubist regression seemed to perform best when taking all evaluated aspects of model performance into account.

Feature selection and validation

Feature selection was performed to identify the most important spectral-temporal features and to estimate the benefit of adding additional features. The difference between *M* derived from the modified chlorophyll absorption ratio index (MCARI2) and the structure insensitive pigment index (SIPI) was consistently the most informative spectral-temporal feature (Table S 3). The MCARI2 was designed to estimate green leaf area index in crop canopies, whereas the SIPI measures pigment concentrations and ratios in leaves. This feature was retained as the last in all 30 resamples by the random forest and in 29 resamples by cubist. Following features had much increased ranks. The most influential features were all based on the *M* parameter of the Gompertz model, whilst other parameters were clearly less important. In particular, differences in change parameters did not seem to be informative of disease severity. Most selected features used the SIPI, R780/R740 and PRInorm indices as STB-insensitive index, even though they seemed to be somewhat more affected by the presence of disease than the FII on average (Table 3.3). There was little evidence for the existence of complementary information among the spectral-temporal features, as model performance was affected little by the sequential removal of features (Figure 3.10). However, the small sample size resulted in very high variance of the performance estimates obtained from the test set and contrasting patterns between the cross-validated training and the test performance estimates (Figure 3.10). The top-selected spectral-temporal feature was found to be informative of STB severity in the separate experiment carried out under contrasting environmental conditions (Pearson r = 0.53, p < 0.001, Figure 3.11).
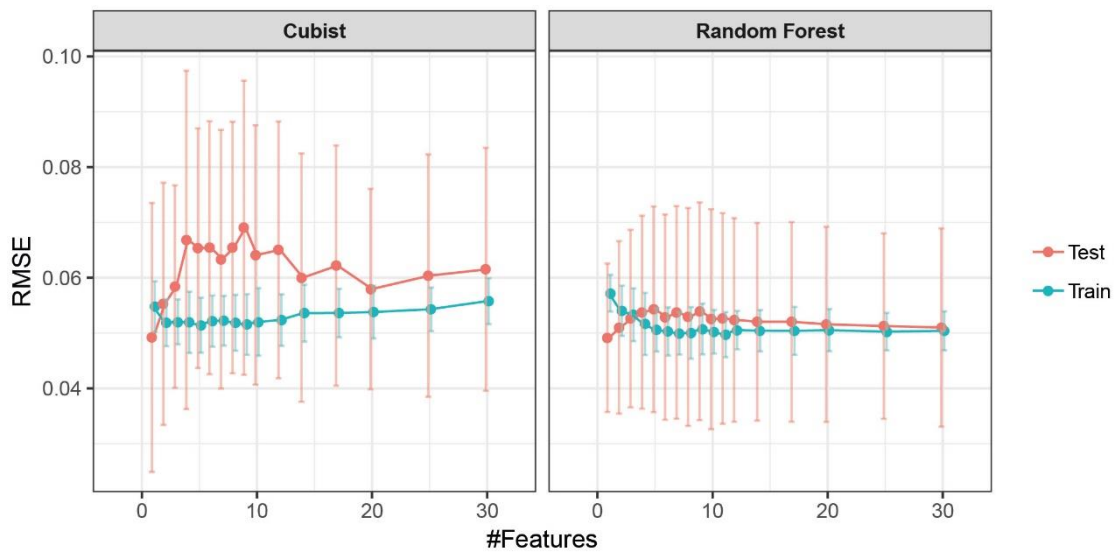

Figure 3.10 Performance profile of models based on spectral-temporal features to predict STB severity as a function of the number of spectral-temporal features used. Mean performance and standard deviation are shown based on 30 resamples of the data
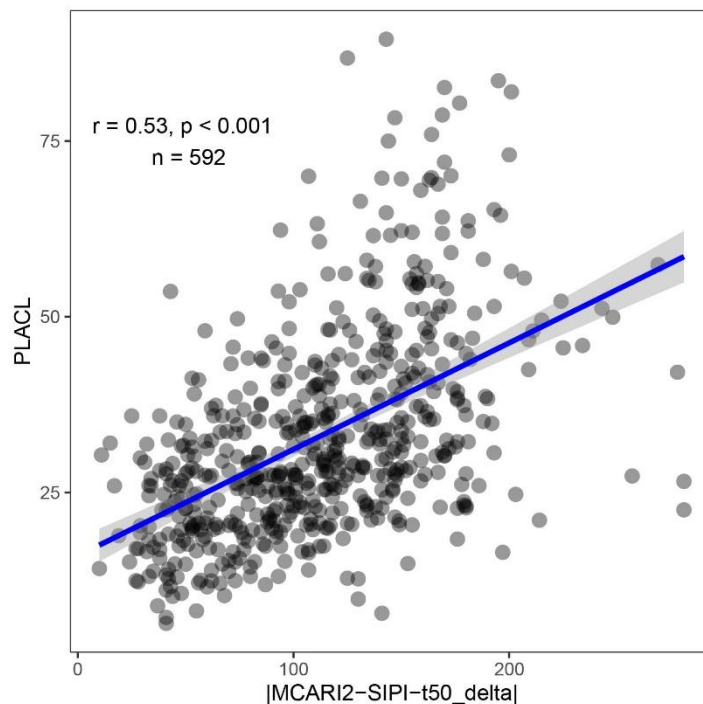
Figure 3.11 *Septoria tritici* blotch (STB) severity as measured by the percent leaf area covered by lesions (PLACL) plotted against the spectral-temporal feature identified as most predictive of STB severity (the difference between t_{50} extracted from the MCARI2 and SIPI spectral vegetation indices, expressed in growing degree days). The Pearson product moment correlation coefficient and the p-value of the linear correlation are based on 592 experimental plots. Data was collected in a separate experiment conducted in 2016 without artificial inoculation. In this experiment, 330 genotypes were grown in two replicates, but some plots were excluded from the analysis due to heavy lodging.

3.4   Discussion

3.4.1   Limitations of time-point specific analyses

Current reflectance-based approaches to high throughput phenotyping of crop diseases under field conditions suffer from a lack of specificity and from insufficient robustness to genotypic diversity and environmental variability (i.e. context specificity). This problem has previously been described in detail with regard to growth stages of the crop, different phases in the pathogenesis and the presence of other stresses (Devadas *et al*., 2015; Zhang *et al*., 2012; Zheng *et al*., 2019).

Our results prominently illustrate context-specificity of the relationship between spectral reflectance and disease. Firstly, variation on a specific date in potentially disease-sensitive spectral features, such as the SIPI (*see* Bajwa *et al*., 2017; Yu *et al*., 2018), is quickly overridden by variation caused by advancing phenology (Figure 3.4C), illustrating the difficulty in defining thresholds or calibration curves. Secondly, unstable VIP values of single wavelengths in PLSDA models, systematic shifts in VIP patterns (Figure 3.6) and limited model applicability over time, even for the plots contained in the training dataset (Figure 3.7), illustrate marked short-term changes in the relationship between STB and spectral reflectance. Thirdly, the decreased classification accuracy on independent test plots (Figure 3.6) indicates context-specificity related to the effect of genotypes and, possibly, field heterogeneity.

There was a short period during late stay-green when classification models were transferable between time-points to some extent (Figure 3.6). This can be explained by the relatively synchronized appearance of moderate to high levels of STB in front of the relatively stable background of a stay-green canopy. In this intermittent phase, the signal caused by STB is strong compared to the noise caused by genotypic diversity and in-field measurements (see also Figure 3.4A). Nonetheless, the regression model for STB severity based on reflectance spectra is still context-specific, as reflectance in the NIR (used as predictors) gradually decreases during the stay-green and senescence phase irrespective of the presence of STB (Figure 3.4A). NIR reflectance is also strongly affected by genotype morphology, canopy 3-D structure and canopy cover (Gutierrez *et al*., 2015; Jacquemoud *et al*., 2009) and is therefore not specific to STB if analyzed on a particular point in time. In addition, time-point specific models highlight the potential of detecting STB in different phases, using different spectral features. This potential would be left unused if only a short period would be targeted.

3.4.2   Potential of temporal changes in reflectance to detect and quantify STB

Due to the strong limitations of models based on reflectance spectra, we evaluated the potential of exploiting temporal changes in reflectance for disease detection and quantification instead. Models based on spectral-temporal features were characterized by a somewhat lower performance compared with models trained on reflectance spectra of a specific time-point (Figure 3.7, Figure 3.8). Nevertheless, classification accuracies were similar to the time-point specific PLSDA models, and regression models suggested that spectral-temporal features were also informative of disease severity. This is encouraging, particularly given the strongly contrasting morphological, canopy structural and stay-green properties of the genotypes comprised in the experiment.

3.4.3   Selected spectral indices and resulting spectral-temporal features

Even though high levels of STB developed during the stay-green phase in most artificially inoculated plots (Figure 3.4), several SVIs could be identified which displayed similar temporal patterns across treatments (Table 3.3). In particular, the flowering intensity index (FII; Stuckens *et al*., 2011), i.e. the normalized difference of R475 and R365, was found to be almost unaffected by STB (Figure 3.2A). In a previous study, we found that early physiological senescence of wheat canopies results in only (proportionally) small increases in reflectance at wavelengths shorter than 500 nm (unpublished data). Strong increases were observable only towards later stages of senescence. The observed insensitivity of the FII to STB likely results from the fact that STB affects only parts of the vegetation, initially mostly lower leaf layers, while significant amounts of healthy green tissue remain. Thus, FII values should change significantly only with the onset of rapid apical senescence, encompassing a generalized loss of green leaf area. It has been suggested that STB does not accelerate or anticipate apical senescence under a range of environmental conditions (Bancal *et al*., 2016). This is in line with the observed insensitivity of the FII to STB. Interestingly, the dynamic pattern of the structure insensitive pigment index (SIPI; Penuelas *et al*., 1995) was also found to be highly insensitive to STB. In contrast, this SVI was previously suggested as a potential surrogate for crop disease under field conditions (Bajwa *et al*., 2017; Yu *et al*., 2018). This Index was developed at the leaf scale to maximize sensitivity to the

ratio between carotenoid and chlorophyll *a* concentrations (Car/Chl *a* ratio), while minimizing the effect of leaf surface and mesophyll structure (Penuelas *et al.*, 1995). Provided the principles underlying the SIPI hold also for canopy level reflectance, a low sensitivity of the dynamic pattern to the presence of STB would be expected, as there seems to be no reason to expect a significant change at canopy level of the Car/Chl *a* ratio due to STB. STB causes the appearance of localized necrotic lesions; however, a general increase in the Car/Chl *a* ratio is not expected, unless STB accelerates or anticipates apical senescence, which does not seem to be the case (Bancal *et al.*, 2016). The PRInorm was also among the most STB-insensitive SVIs. This SVI is based on the photochemical reflectance index (PRI), initially employed to measure changes in the relative levels of pigments in the xanthophyll cycle (Gamon *et al.*, 1992). Over larger temporal scales, the PRI was shown to be strongly responsive to the Car/Chl ratio (Sims and Gamon, 2002). Zarco-Tejada *et al.* (2013) modified this SVI to decrease the effect of reduced canopy leaf area resulting from water stress. Thus, its insensitivity to STB can probably be explained in an analogous manner as for the SIPI.

The temporal patterns of water-sensitive SVIs such as the water index (WI; Peñuelas and Inoue, 1999) and the normalized difference water index (NDWI; Gao, 1996) were found to be highly sensitive to STB. Similarly, the disease water stress index (DSWI; Apan *et al.*, 2004), which uses information from the water-sensitive SWIR and the NIR, was strongly affected in its temporal dynamics. In particular, water sensitive SVIs decreased much earlier for inoculated than for control plots, and the decrease occurred more gradually than in healthy plots (data not shown). This is in line with findings by Yu *et al.* (2018), who reported both the WI and NDWI to discriminate best between STB-diseased and healthy canopies in early stages of disease development. Several SVIs using reflectance in the RE and NIR also showed strongly contrasting temporal patterns (e.g. DSWI, NDVI, PSRI, and VI700). Similar to the SIPI, the plant senescence reflectance index (PSRI; Merzlyak *et al.*, 1999) is highly sensitive to the Car/Chl ratio at the leaf level. However, NIR reflectance is used to normalize the difference between R677 and R500. It seems highly questionable whether the PSRI is particularly sensitive to the Car/Chl ratio in diverse germplasm at the canopy level. Variation in the PSRI seems to arise primarily from differences in canopy structure among genotypes and from canopy structural changes over time (Anderegg *et al.*, 2020). The modified chlorophyll absorption ratio index (MCARI2; Haboudane *et al.*, 2004), sensitive to green leaf area index, also showed strongly contrasting dynamic patterns (Figure 3.2B). The healthy-index (HI; Mahlein *et al.*, 2013) was developed to separate healthy sugar beet leaf tissue from tissues affected by various foliar diseases. The prominent use of the RE by this index suggests that in our case, HI values are mostly driven by canopy structure and to a lesser extent chlorophyll absorption. Overall, our results thus suggest that SVIs sensitive to leaf internal structure and canopy structure are strongly affected by the presence of STB. This effect has been previously described for various patho-systems (*e.g.*, Yu *et al.*, 2018; Zhang *et al.*, 2012; Zheng *et al.*, 2019).

In summary, it seems that many of the derived spectral-temporal features can be interpreted as robust measures of STB-induced temporal changes in leaf internal structure, canopy structural parameters and canopy water content. These are obtained by normalizing the temporal dynamics

of corresponding SVIs *via* the estimation of temporal changes in pigment ratios and reflectance at short wavelengths centered around 465 nm, likely representing physiological apical senescence. Thus, spectral-temporal features seem to well represent our hypothesis H_2, as STB-affected plant and canopy traits are expressed relative to phenology-related traits.

3.4.4 Robustness of spectral-temporal features

As far as quantifiable in the framework of this experiment, models based on spectral-temporal features were robust to variation in genotype morphology, phenology, canopy cover and canopy 3D-structure as well as genotype-specific temporal changes thereof. Accurate predictions were obtained also for 330 diverse genotypes comprised in the GABI panel and grown in a large field experiment, suggesting robustness of the method to varying growing conditions arising from field heterogeneity (Figure 3.9B). The most predictive feature was also among the best predictors of STB severity in a different experiment. The observed correlation is very similar to the correlations reported by Yu *et al*. (2018) for simple SVI measured at individual time points, however with the advantages discussed throughout the manuscript. The results of this final validation likely underestimate the power of the developed approach for several reasons as described in section 3.2.4.5, but offer evidence for the transferability of our results across sites, environments, genotypes and agricultural practices.

Results from feature selection suggested that a single spectral-temporal feature (i.e. the difference between *M* derived from MCARI2 and SIPI), relating structural changes in leaves and canopies to senescence-induced changes in pigment composition, was sufficient to achieve the performance illustrated above. Other stresses occurring during grain filling such as terminal drought stress and nitrogen shortages are likely to result in a similar decrease in green leaf area index. However, these stresses are also known to accelerate physiological senescence (Bogard *et al*., 2011; Distelfeld *et al*., 2014; Martre *et al*., 2006). Therefore, we speculate that the developed models may be moderately robust against the effect of common other stresses despite their simplicity. Yet, we conclude that our hypothesis H_3 (i.e. that the combination of several spectral-temporal features representing the unique sequence and dynamics of separate events during pathogenesis could increase the specificity of the method) remains to be confirmed in larger experiments including other stress factors. In particular, other diseases causing similar symptoms and prevalent in the same developmental stage of the crop may have similar effects on the temporal evolution of hyperspectral reflectance. Large multifactor experiments will be required to judge the potential of the proposed approach to detect, quantify and delineate individual necrotrophic foliar diseases. Finally, it should be noted that some effects of fungicide applications on canopy reflectance characteristics cannot be excluded in our experiment. However, a fungicide formulation without greening effect was used for the last treatment at BBCH 65. It seems unlikely that this or earlier fungicide applications significantly affected the temporal dynamics of canopy reflectance.

3.4.5 Multiple spectral vegetation indices to exploit temporal dynamics in reflectance

A key component of the proposed approach consists in summarizing hyperspectral data in terms of multiple SVIs and modelling of their temporal dynamics. Though this may result in the loss of some relevant information contained in reflectance spectra (Pauli *et al.*, 2016), the use of SVIs presented a number of advantages here: (i) noise in temporal patterns was much reduced compared to reflectance values at single wavelengths, facilitating the fitting of parametric models; (ii) the inevitable subset selection step preceding feature combination could be based on objective criteria related to the form and purpose of SVIs; (iii) many of the used SVIs were designed specifically to maximize responsiveness to certain vegetation properties while minimizing the effect of common confounding factors, which is likely to also increase the robustness of derived spectral-temporal features (*see e.g.*, Haboudane *et al.*, 2004; Penuelas *et al.*, 1995); and finally (iv) the procedure results in a summary of the hyperspectral dataset that is interpretable in terms of plant physiology and canopy characteristics, which also holds true for derived spectral-temporal features. Fitting parametric models to scaled SVI values may smooth out measurement errors related to single measurement dates, resulting for example from varying sun angle at measurement or short-term variation in illumination conditions. Thus, scaling SVI values and modelling their temporal dynamics reduces the effect of confounding factors on initial reflectance spectra and minimizes the effect of errors related to single measurements in the series.

3.4.6 Context and scope

In this study, we used a non-imaging spectroradiometer and manual feature engineering for disease detection and quantification. A high spatial resolution of imaging sensors has been deemed critical for disease detection, identification and quantification by others (Mahlein, 2016; Mahlein *et al.*, 2012, 2010). The high potential of 2-D information in combination with deep learning methods for disease identification has been demonstrated recently (Fuentes *et al.*, 2017; Mohanty *et al.*, 2016). However, changes in spectral reflectance over time have also been shown to be highly informative at the leaf level (Mahlein *et al.*, 2010; Wahabzada *et al.*, 2015). To make use of the spatial and temporal dimensions under field conditions, individual lesions would arguably have to be tracked across time. Some solutions to this problem have been presented for close-range hyperspectral measurements (Behmann *et al.*, 2018). However, similar solutions at the canopy level may be technically extremely challenging to implement and require extensive studies due to problems in tracking individual pixels or organs over time and in obtaining a clean spectral signal from objects with varying orientation. Existing approaches to make use of spectral, spatial and temporal information rely on automated and data-driven extraction of characteristic spectral features for diseased plants under controlled conditions (Thomas *et al.*, 2018; Wahabzada *et al.*, 2016, 2015). Here, promising results were achieved using a non-imaging sensor and manual feature extraction. This highlights that an improved understanding of potential confounding factors arising under field conditions may equally boost the potential of remote sensing methods for applications in crop breeding.

We developed and validated the presented method to facilitate robust in-field detection and quantification of STB. However, the underlying concepts should be transferrable to different problems, such as the detection and quantification of other foliar diseases. Several features of the proposed approach (e.g. exploiting plot-based relative changes in reflectance over time, combining sensitive and insensitive features, or the SVI-based parameterization of temporal dynamics) may also be valuable in quantifying other breeding-relevant traits, such as the timing and dynamics of nitrogen remobilization.

3.5 Conclusions

Here, we tested the possibility to detect and quantify STB relying exclusively on relative changes in spectral reflectance over time, which is expected to minimize confounding effects on spectral reflectance arising from genotypic diversity and environmental conditions. Our results demonstrated the feasibility of the proposed approach and suggested that resulting models were robust against variation in several common nuisance factors. Specifically, it appears that the temporal dynamics in green leaf area index when set in relation to the dynamics of physiological apical senescence is highly indicative of the presence of STB infections and of STB severity. Time-resolved measurements of the MCARI2 and the SIPI spectral vegetation indices could allow to assess these traits at very high throughput, facilitating time-resolved large-scale screenings of breeding nurseries.

Larger calibration experiments will offer the opportunity to evaluate the inclusion of additional spectral-temporal features that better capture relevant information in different phases of pathogenesis. This is likely to improve sensitivity and specificity of resulting models, which should also be tested in more detail. Furthermore, the evaluation of the scalability to unmanned aerial vehicles will represent a crucial step towards application of such methods in breeding programs.

4 Temporal Trends in Canopy Temperature as Potential Indicators of Resistance to Short-Term Drought Periods in Diverse Wheat Breeding Material

Jonas Anderegg[1, †], Gregor Perich[1, †], Achim Walter[1], Helge Aasen[1, ‡], Andreas Hund[1, ‡]

[1]Crop Science Group, Institute of Agricultural Sciences, ETH Zurich, Zurich, Switzerland

[†]Joint first authors, [‡]Joint last authors

*Corresponding author Email: jonas.anderegg@usys.ethz.ch

Abstract

Transpiration through leaf stomata decreases leaf temperature due to the cooling effect of water evaporation. Drought-resistant genotypes avoid canopy dehydration and the subsequent reduction in transpiration rates under drought. Therefore, measurements of canopy temperature (CT) have proven a useful tool in the selection of drought-resistant genotypes for breeding purposes in wheat. However, besides transpiration rates, variation in shoot biomass, morphology, canopy structure and phenology also strongly affect CT. This study evaluates the potential of repeated measurements to derive temporal trends in CT, independently of absolute values of CT. In a set of 354 registered European wheat cultivars, a moderate to high repeatability was found for temporal trends in CT during the stay-green phase under hot and dry conditions. Temporal trends were much less correlated with biomass, canopy structural parameters and phenology than single time point measurements. A comparison of trends in CT and in spectral indices representing canopy greenness demonstrated that green leaf area and CT evolve independently over time. Among three replicated check cultivars, the contrast between temporal trends in greenness indicators and CT appeared to be genotype-specific. This suggested that a combination of time-resolved thermal and spectral measurements may allow for the identification of different levels of functionality in stay-green.

4.1 Introduction

Drought is a major factor limiting wheat productivity in rain-fed production systems around the world. In Europe, increased climatic variability, in particular the occurrence of heat and drought in high-yielding regions, has recently counter-balanced constant genetic progress in yield potential (Brisson *et al.*, 2010; Oury *et al.*, 2012). As a result of climate change, heat and drought periods are predicted to further increase both in frequency and in severity (Calanca, 2007; Lehner *et al.*, 2006; Trnka *et al.*, 2015).

In regions with predictable early-season rainfall, heat and drought occurring during the late growing season may be escaped by adjusting crop phenology. However, this escape strategy in combination with increased temperatures has led to a considerable shortening of the generative growth phase in modern wheat cultivars (Rezaei *et al.*, 2018). Further adjustments in phenology may become increasingly incompatible with sufficient biomass acquisition and adequate potential yield formation during these stages (Slafer *et al.*, 2015). Thus, developmental adaptation will have to be complemented with genetic progress in plant performance under heat and drought stress (Lobell *et al.*, 2011; Tester and Langridge, 2010).

Drought resistance is often regarded as a complex trait, arising from different underlying constitutive or adaptive traits, each of which potentially under complex genetic and environmental control. However, it has been argued that viewed from an agronomy or breeding perspective, it is a relatively simple phenomenon, manifesting itself primarily as the avoidance of plant dehydration under prolonged periods of limited water availability (Blum, 2011). Dehydration avoidance is equally important under heat stress conditions, because higher transpiration rates reduce leaf and canopy temperature through evaporative cooling (Amani *et al.*, 1996; Ayeneh *et al.*, 2002; Reynolds *et al.*, 1994). Therefore, the assessment of leaf or canopy temperature (CT) has been proposed as a low-cost indirect selection criterion for drought and heat stress resistance (*e.g.*, Mason and Singh, 2014; Reynolds *et al.*, 2009; Singh and Kanemasu, 1983).

So far, CT measurements in genetics or breeding experiments were most often obtained plot-by-plot using hand-held infrared thermometers (*e.g.*, Lopes and Reynolds, 2010; Rebetzke *et al.*, 2013). Lower CT values have been associated with increased yield under various conditions (Li *et al.*, 2019; Lopes and Reynolds, 2010; Rebetzke *et al.*, 2013; Reynolds *et al.*, 1994). However, such measurements are time consuming and often result in a low repeatability if environmental conditions change during the measurements (Deery *et al.*, 2016). More recently, airborne thermography has gained popularity as it allows for the assessment of CT in large breeding or genetics experiments in short time (Deery *et al.*, 2019, 2016; Joalland *et al.*, 2018; Li *et al.*, 2019; Liebisch *et al.*, 2015; Perich *et al.*, 2020), increasing repeatability as compared to manual plot-by-plot measurements (Deery *et al.*, 2016).

Unfortunately, CT is not only a function of canopy water status and transpiration rates, but is also strongly affected by shoot biomass, phenology, morphology and canopy structural parameters such as plant height, leaf area index, ground cover and leaf and spike orientation (*reviewed by* Prashar and Jones, 2014). This is prominently illustrated by a high repeatability of CT even at

physiological maturity, where only dry canopies and no transpiring plant tissues are left (Perich *et al.*, 2020). The presence of confounding factors may complicate the use of CT as a proxy for drought or heat stress resistance (i) when diverse material is screened in order to identify new sources of drought resistance, (ii) when aiming at early selection of genetically diverse material, where confounding factors cannot be tightly controlled, and (iii) in environments with drought or heat stress intermittent and restricted to certain phenophases. For example, if drought is restricted to the grain filling phase, it does not affect stem elongation and biomass acquisition. Still, genotypic differences in final height and biomass have strong effects on CT measured at a later point, thus confounding the possible effect of drought on CT. Therefore, genotype-specific temporal trends in CT may be more informative of drought and heat stress resistance than individual measurements. Genotypes better able to cope with extended periods of water shortage and high temperatures should exhibit little change in CT over time, whereas drought- or heat-susceptible genotypes are expected to tend towards higher CT, as transpiration rates decrease (Figure 4.1A).

The derivation of plot-based trends in CT is complicated by the effect of environmental factors varying from one measurement campaign to the other. Such factors are air and soil temperature, cloud cover or wind speed (Prashar and Jones, 2014). Measurement errors related to single time points may be removed by the use of average CT integrating over a longer period of time, as proposed by Thapa *et al.* (2018). In contrast, the above-mentioned crop-related confounding factors cannot be effectively controlled in this manner. To better compare measurements across different measurement time points, a genotype's CT at a specific measurement time point may be expressed relative to the rest of the population under study, as a normalized relative canopy temperature (NRCT) calculated as (Elsayed *et al.*, 2015; Jackson *et al.*, 1981)

$$NRCT = (CT - CT_{min})/(CT_{max} - CT_{min})$$

where CT is the canopy temperature for a particular experimental plot at a specific measurement time point, and T_{min} and T_{max} are the lowest and highest canopy temperatures measured in the entire experiment at a specific measurement time point, respectively.

Terminal heat and drought stress may trigger premature senescence, resulting in a shortening of the grain filling phase and in reduced assimilate production during grain filling (Gregersen *et al.*, 2013). Under such conditions, a prolonged post-anthesis maintenance of green leaf area (i.e. the "stay-green" phenotype; Thomas and Smart, 1993) has been proposed as a possible selection criterion (Christopher *et al.*, 2008; Lopes and Reynolds, 2010, 2012). However, the stay-green phenotype may equally result from source-sink imbalances (Bogard *et al.*, 2011; Borrell *et al.*, 2003; Jiang *et al.*, 2004), in which case it would be an undesirable trait for selection. Thus, a delayed senescence is expected to be beneficial in wheat only if coupled with a highly productive green leaf area during the stay-green phase and a subsequent fast and efficient remobilization of nutrients (Gregersen *et al.*, 2008), indicating a strong sink demand. Spectral vegetation indices (SVI) can be used to assess stay-green in wheat at high throughput (Anderegg *et al.*, 2020; Christopher *et al.*, 2014; Lopes and Reynolds, 2012). However, they are limited in their capability to assess plant performance in response to environmental conditions (e.g. photosynthesis or
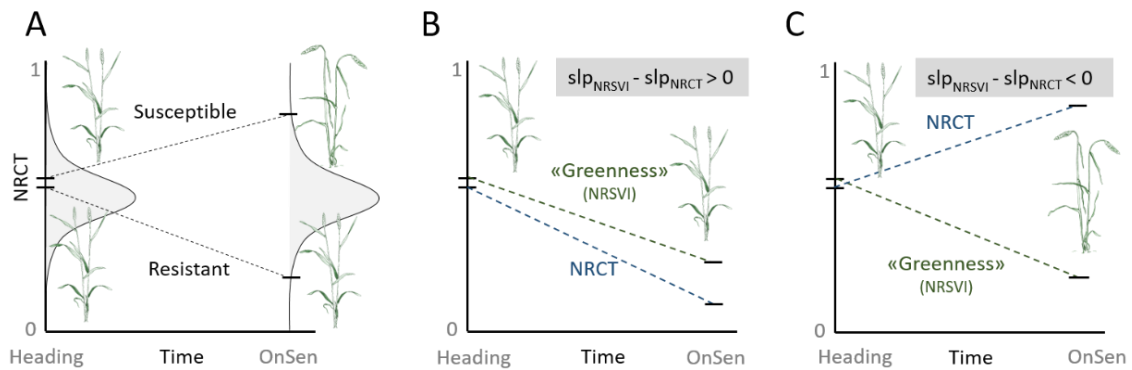
Figure 4.1 Hypotheses underlying this work. **(A)** Temporal trends from heading to late stay-green in normalized relative canopy temperature (NRCT) for a drought susceptible and a drought resistant genotype, in relation to the population (represented by the sketched grey distribution). **(B)** A genotype exhibiting functional stay-green; the low decay in canopy greenness relative to the population (indicated by a normalized relative spectral reflectance index) is accompanied by a relatively decreased NRCT. **(C)** A genotype exhibiting dysfunctional or partly functional stay-green; the relatively low decay in canopy greenness is not accompanied by a decreased NRCT, suggesting that the stay-green phenotype is not paralleled by relatively high transpiration and possibly relatively high photosynthetic rates. Illustrations of wheat plants were modified from Schürch *et al*. (2018).

transpiration rates). In contrast, a combination of spectral and thermal measurements may allow for a more precise characterization of stay-green. Specifically, genotype-specific contrasting trends between SVI commonly used to characterize canopy greenness, green leaf area index, green biomass and senescence on the one hand, and CT on the other hand may allow for a more precise characterization of crop performance during the stay-green phase (Figure 4.1B, Figure 4.1C; Rebetzke *et al*., 2016).

The overall aim of this study was to evaluate the potential of plot-based changes in NCRT over time as a selection criterion for resistance to intermittent or terminal drought conditions in genetically diverse breeding material. The specific objectives were (i) to evaluate the repeatability of plot-based changes in NRCT over time, (ii) to examine the correlation between changes in NRCT and initial NRCT or NRCT integrating CT throughout the investigated phase, (iii) to examine the correlation between changes in NRCT and morphological, canopy structural and phenological traits known to affect CT at specific time points and (iv) to evaluate the potential of a combination of thermal and spectral temporal trends to characterize genotype performance during the stay-green phase. To achieve this, we examined a time series of thermography and spectral reflectance measurements carried out during grain filling using our recently established unmanned aerial vehicle (UAV) measurement platform and image and reflectance data processing pipelines (Anderegg *et al*., 2020; Perich *et al*., 2020).

4.2   Materials and Methods

4.2.1   Plant Materials, experimental design and meteorological data

A field experiment was conducted in the field phenotyping platform FIP (Kirchgessner *et al*., 2017) at the ETH Research Station for Plant Sciences Lindau-Eschikon, Switzerland (47.449N,

8.682E, 520 m a.s.l.; soil type: eutric cambisol) in the wheat growing season of 2017-2018 which was characterized by extraordinarily dry and hot weather from April 2018 on (MeteoSwiss, 2018). Three hundred cultivars comprised in the GABI wheat panel (Kollers *et al*., 2013) obtained from the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) were used and complemented with important Swiss cultivars for a total of 354 cultivars. A detailed description of the experimental design is provided by Anderegg *et al*. (2020). In short, cultivars were grown in two spatially separated replications in plots of 1 m × 1.4 m size, arranged in a two-dimensional incomplete block design with three check varieties. Checks were distributed evenly at 21 locations in each replicate. The experiment was sown with a sowing density of 400 plants m^{-2} on Oct 18, 2017 and crop husbandry was performed according to local agricultural practice. Temperature data was obtained from an on-site weather station. Rainfall data was obtained from a nearby weather station of the federal Swiss meteorological network Agrometeo (www.agrometeo.ch) located at ca. 250 m distance to the field trial.

4.2.2 Phenology, canopy structural and morphological parameters and biomass

The used plant material comprised phenologically, morphologically and structurally diverse genotypes. To characterize the diversity of the panel with respect to traits reported to affect CT, we assessed heading date (HD), the onset of apical senescence (OnSen), glaucousness, flag leaf angle, flag leaf length and width, the presence or absence of awns, final height and above ground biomass. For simplicity, these traits are subsumed under the term "covariates" hereinafter.

HD was recorded when 50% of the spikes were fully emerged from the flag leaf sheath (BBCH 59, Lancashire *et al*., 1991). Following Gooding *et al*. (2000) and Pask *et al*. (2012), OnSen was extracted from a Gompertz model fitted to repeated visual canopy senescence scorings as described by Anderegg *et al*. (2020). To account for contrasting temperatures experienced by early and late genotypes, HD and OnSen were converted to growing degree days after sowing (GDDAS) and GDD after heading (GDDAH), respectively, using hourly mean temperatures and a base temperature of 0°C. Glaucousness, flag leaf angle, length and width were assessed strictly following the guidelines provided by Pask *et al*. (2012). Final height was determined over the three preceding years (growing seasons of 2014-2015 until 2016-2017) for most of the genotypes comprised in the experiment based on repeated terrestrial laser scans (*for details refer to* Kronenberg *et al*., 2017). Here, the derived three-year genotypic best linear unbiased estimators (BLUE) were used, as within-year repeatability of final height was 0.98-0.99 and across-year heritability was 0.98 (Kronenberg *et al*., 2019). Finally, above-ground biomass was estimated using a calibration for spectral reflectance created in parallel as follows: For a subset of 49 genotypes grown in one replicate next to the main experiment, above-ground biomass was determined destructively as the average of eight samples of an area of 0.25 m^2. Technical replicates deviating from the overall mean by more than two standard deviations were excluded from the analysis as outliers. Spectral reflectance was measured on 13 dates covering the entire grain filling phase on the same plots. For each measurement date, a large number of spectral vegetation indices were computed and evaluated for their correlation to destructively determined above-ground biomass, and the optimal date-by-SVI combination was used to infer biomass for

all plots of the GABI panel. The highest correlation ($R^2 = 0.36$, Figure S 7.2) was observed for the reflectance ratio at 1200 nm, calculated as (Pu, 2011)

$$Ratio_{1200} = \frac{2 \times Avg(R1180to1220)}{Avg(R1090to1110) + Avg(R1265to1285)}$$

for measurements carried out on 22 June, 2018 (i.e. during late stay-green/early senescence). The use of this SVI was preferred over full-spectrum models due to the instability of the latter, probably resulting from the relatively small sample size (n = 49; *for details on measurement procedures and calculation of SVI refer to* Anderegg *et al.*, 2020).

4.2.3   Canopy temperature and derived time-integrated traits

CT data was obtained from a radiometrically calibrated FLIR A65 thermal imaging camera (FLIR integrated Imaging Solutions Inc., Canada) with a field of view field of 25° x 20° and a resolution of 640 x 512 pixels. The camera was mounted on a DJI Matrice 600 Pro (SZ DJI Technology Co. Ltd., China). Both the UAV platform and the processing of the thermal data were described in detail by Perich *et al.* (2020). In short, 24 flights were carried out during the grain filling and ripening phase (BBCH 73-92, Lancashire *et al.*, 1991) at 80 m height above ground level with an image overlap of >70% across flight direction and >90% along the direction of the flight path, resulting in a flight duration of approximately 8 min to cover the two replications. Thermal orthomosaics were generated from individual thermal images in the following steps implemented in Agisoft PhotoScan Professional 1.4.3 (Agisoft LLC, St.Petersburg, Russia): (i) generation of sparse point clouds using the Structure from Motion algorithm, (ii) geo-referencing of sparse point clouds using custom-made thermal ground-control points, (iii) generation of digital surface models, (iv) generation of dense point clouds, (v) generation of digital surface models and finally (vi) generation of georeferenced thermal orthomosaics using the blending mode 'average'. Plot-wise canopy temperature was then extracted using polygons describing the plot shape and location. QGIS 3.2.3 Geographic Information System Software (QGIS Development Team, 2019) was used to create an inward buffer of 40 cm from the shapes to omit edge effects. Based on a Python 3.6 script, the median of this area was then used as CT for a plot (*for details refer to* Perich *et al.*, 2020).

Time-integrated measures of CT were derived from NRCT obtained from flights carried out in June 2018. According to the recommendations of Perich *et al.* (2020), flights performed at around 14:00h local time were selected for the analysis. For each plot, the following time-integrated CT traits were extracted: (i) the slope of the linear model fitted to date-specific NRCT, (ii) based on the same linear model fit, the predicted NRCT at mid-stay-green and (iii) the predicted NRCT at OnSen. Unfortunately, the presence of significant variation in heading date (12 d between the earliest and latest genotype), as well as variation in the duration of the stay-green phase (i.e. between HD and OnSen), allowed only a limited number of flights covering the stay-green phase for most genotypes to be retained for this part of the analysis. The selected five flights were carried out prior to OnSen for 539 experimental plots (i.e. 72% of all plots). Thus, the selection of these five flights represented a trade-off between limiting the effect of phenology on the

characterization of the stay-green phase and retaining as much data as possible to enable a reliable derivation of temporal trends.

4.2.4 Spatial correction and calculation of repeatability

Spatial heterogeneity of NRCT values at each measurement date, for each of the time-integrated NRCT traits, and for each of the covariates was corrected by applying two-dimensional P-splines to the raw data using the R-package SpATS (Rodriguez-Alvarez *et al.*, 2019). Repeatability (w^2), genotypic BLUEs and spatially corrected plot values were obtained as described previously (Anderegg *et al.*, 2020). Spatial correction for time-integrated NRCT traits was performed either on derived traits extracted from uncorrected NRCT values at each time point or on derived traits extracted from spatially corrected time-point-specific NRCT values (in this case, spatial correction was applied at the level of individual NRCT values per time point as well as at the level of derived NRCT traits). For simplicity, date-wise spatially corrected NRCT values are denoted as $NRCT_{corr}$, whereas uncorrected NRCT values are denoted as $NRCT_{raw}$.

4.2.5 Correlation of Canopy temperature and derived traits with covariates

Covariates were measured at one point in time and correlated with CT on different measurement dates by calculating pair-wise Pearson product moment correlation coefficients. For correlation analyses at each measurement time point, uncorrected CT and covariate values were used. To examine correlations between time-integrated measures of CT and covariates, spatially corrected values were used. This approach was motivated by the fact that linear models fitted to date-wise spatially corrected NRCT yielded higher R^2 and lower p-values on average, and derived time-integrated measures of NRCT had a higher repeatability (see section 4.3 for details).

To quantify the portion of the variance in NRCT explained by covariates, multiple linear regression using all covariates was performed for NRCT values at each time point as well as for time-integrated measures of NRCT. All calculations and data analyses were performed in the R environment for statistical computing (R Core Team, 2018).

4.2.6 Combining thermal and spectral measurements to characterize genotype performance during stay-green

To estimate the potential of thermal and spectral measurements to provide complementary information regarding plant performance during the stay-green phase, we first examined the correlation in the temporal trends of the respective signals. We selected the three frequently used indices NDVI (Tucker, 1979), MCARI2 (Haboudane *et al.*, 2004) and PSRI (Merzlyak *et al.*, 1999) and determined their dynamic development (slope) during the stay-green phase as was done for NRCT (*for details on measurements of spectral reflectance and spectral data processing refer to* Anderegg *et al.*, 2020). To quantify the contrast between thermal and spectral trends, we calculated the difference between the resulting slopes as $slope_{SVI} - slope_{NRCT}$. In this way, a difference greater than zero hypothetically indicates a genotype maintaining a relatively low canopy temperature, though not necessarily a high canopy greenness (Figure 4.1B). Conversely, a difference smaller than zero hypothetically indicates a genotype maintaining a relatively high
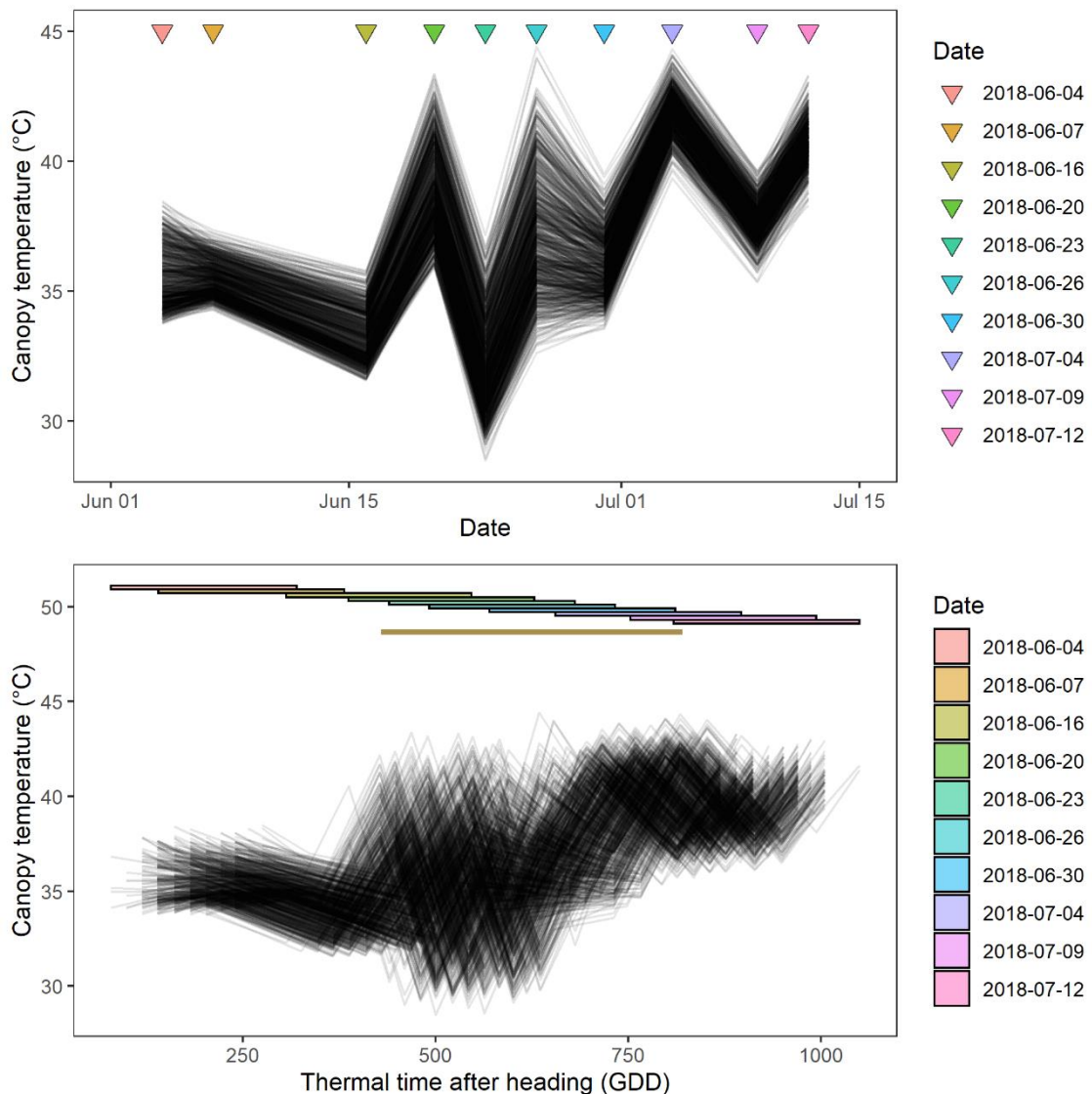
Figure 4.2 Trends in canopy temperature across measurement time points for 756 experimental plots, sown with two replications of 354 genotypes plus checks. The first five measurement dates were used for the derivation of time-integrated CT traits. Upper panel: Canopy temperature across measurement dates. Lower panel: Canopy temperature across measurement time points as expressed in thermal time after heading. Here, measurement dates correspond to a range in thermal time after heading (represented by colored bars) due to the phenological diversity of the screened material. The dark yellow bar corresponds to the observed range in the onset of rapid apical senescence. Due to contrasting average daily temperature sums between measurement dates, intervals are stretched or contracted.

canopy greenness, but not a low canopy temperature (Figure 4.1C). As spectral measurements were performed only on one replicate (i.e. one Lot), no repeatabilities for the observed patterns could be calculated. Instead, we compared the distribution of the derived trait observed for three replicated check cultivars with the distribution observed for all 354 genotypes in the experiment to obtain an indication of whether genotype-specific differences may be measureable (see section 4.2.1).

4.3 Results

4.3.1 Trends in canopy temperature and correlation with covariates

A clear temporal trend was observable for CT values across measurement dates (Figure 4.2, upper panel). This trend is primarily attributable to environmental conditions at the time of measurement or during a short period preceding the measurement, as demonstrated by the temporal trends in CT expressed in thermal time after heading (Figure 4.2, lower panel). Hence, the time series examined here demonstrates that the interpretation of temporal trends in absolute CT values is greatly complicated by fluctuations in mean canopy temperature across dates and its variation within each date making modelling of the temporal trend and extraction of interpretable parameters challenging.

Besides environmental conditions, different crop phenological, morphological and canopy structural traits (i.e. "covariates") seemed to affect CT at specific points in time (Figure 4.3). Of the examined covariates, shoot biomass as approximated by the Ratio1200 was highly correlated to CT, reaching a maximum value of $r = 0.81$ on 23 June. Flag leaf length and OnSen also strongly correlated with CT (r-values of circa -0.5). In contrast, flag leaf angle, glaucousness, and width were only weakly correlated with CT (Table 4.1). Final height correlated weakly with CT, except at the final measurement time points. The correlation between heading time and CT showed an increasing trend towards later measurement time points, but remained relatively weak throughout the assessment period. The portion of the variance in normalized CT explained by the covariates in multiple linear regression models varied across time points. It was particularly large in the period between 20 June and 30 June, i.e. during the late stay-green and early senescence phase
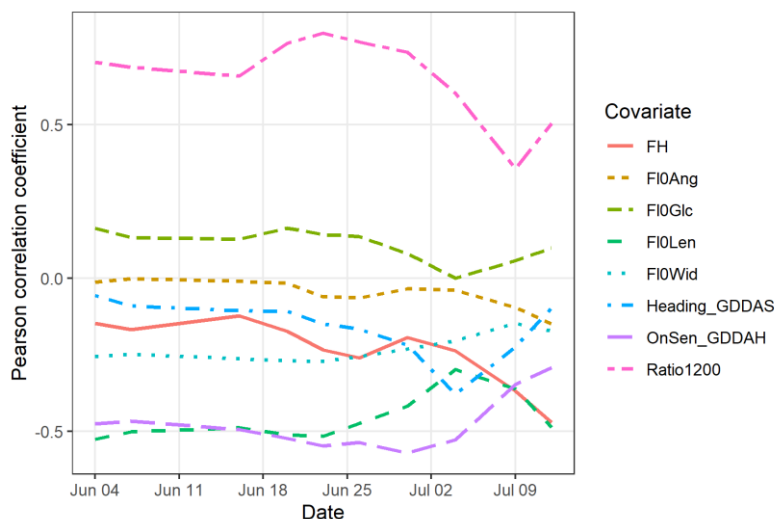

Figure 4.3 Correlation between canopy temperature and final canopy height (FH), flag leaf angle (Fl0Ang), flag leaf glaucousness (Fl0Glc), flag leaf length (Fl0Len) and width (Fl0width), heading (expressed in growing degree days after sowing), onset of apical senescence (OnSen, expressed in growing degree days after heading), the ratio of reflectance at 1200 nm (Ratio1200) as a biomass proxy for all measurement dates.

(0.62 ≤ adj. R^2 ≤ 0.67, p < 0.001), but somewhat smaller during the early stay-green phase (0.43 ≤ adj. R^2 ≤ 0.54, p < 0.001).

4.3.2 Time-integrated analysis of CT

We hypothesized that time-integrated measures of CT may represent useful alternatives to absolute values of CT at specific points in time, particularly when genetically diverse material is examined and/or when the stress is restricted to certain pheno-phases (see Section 4.1 for details). To evaluate this hypothesis, we calculated repeatability of trends in NRCT and evaluated their correlation with factors known to confound the CT signal at specific points in time. Linear regression models were fitted to date-wise spatially correct NRCT values, as well as to raw derived from uncorrected CT values over thermal time for each individual plot.

Fitted linear models often provided a poor fit to the data, but the slopes of these regressions, representing changes in NRCT during the period between June 4 and June 23, showed a moderate to high repeatability (w^2 = 0.27 if slopes were obtained for $NCRT_{raw}$, w^2 = 0.64 if slopes were obtained for $NRCT_{corr}$; a summary of the linear model fits is given in Figure S 7.3 and Figure S 7.4). The predicted NRCT at OnSen had a high heritability, with w^2 = 0.76. Thus, both trends in NRCT and average NRCT during the stay-green phase appear interesting as selection tools from this perspective.

Table 4.1 Pearson correlation between time-integrated normalized relative canopy temperature (NRCT) traits and covariates as well as the initial NRCT. Time-integrated NRCT traits were extracted for date-wise spatially corrected NRCT values. NRCT at mid stay-green ($NRCT_{midstg}$) and NRCT at the onset of senescence ($NRCT_{OnSen}$) were predicted from the linear model, whereas the initial NRCT ($NRCT_{init}$) represents the NRCT at the first measurement date, i.e. on 4 June, 2018.

IntTrait	$NRCT_{init}$	Ratio1200	FH	Heading	OnSen	Fl0Ang	Fl0Glc	Fl0Len	Fl0Wid	[1)] Adj. R^2	[2)] Adj. R^2
$NRCT_{init}$	-	0.47	0.41	0.13	0.12	0.22	0.03	0.19	0.09	0.43	-
$NRCT_{midstg}$	0.95	0.56	0.43	0.20	0.19	0.23	0.00	0.20	0.12	0.46	0.91
$NRCT_{OnSen}$	0.84	0.58	0.56	0.32	0.19	0.24	0.01	0.22	0.11	0.61	0.86
$Slope_{NRCT\sim GDD}$	0.30	0.01	0.24	0.24	0.02	0.05	0.04	0.03	0.03	0.21	0.39

[1)] Adjusted R^2 of multiple linear regression models with all covariates as predictors

[2)] Adjusted R^2 of multiple linear regression models with all covariates and the initial rank as predictors

Ratio1200 Ratio of reflectance at 1200 nm

FH final height

OnSen onset of senescence (GDDAH)

Heading Heading (GDDAS)

Fl0Ang flag leaf angle

Fl0Glc flag leaf glaucousness

Fl0Len Flag leaf length

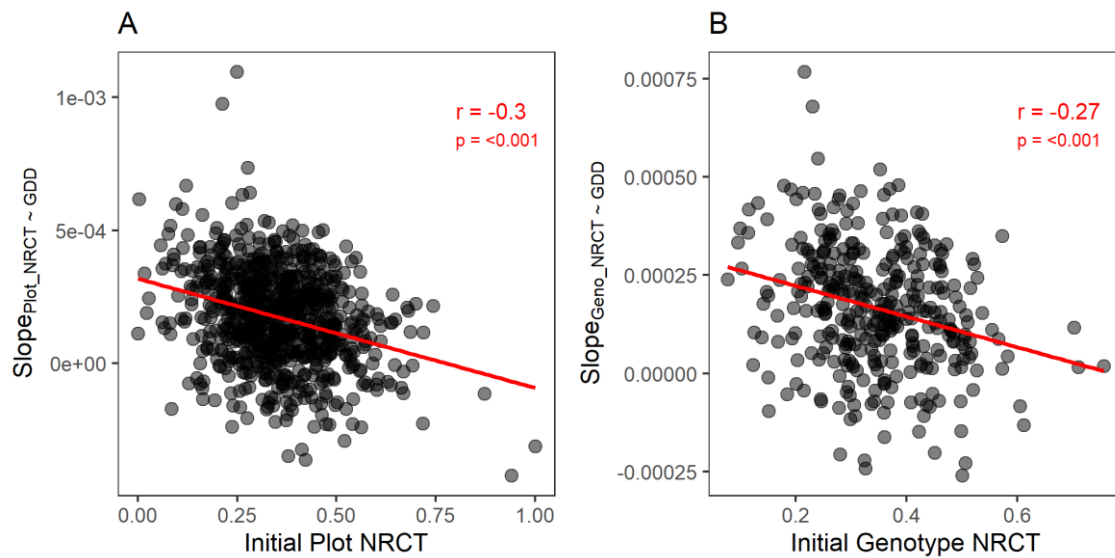**Fl0Wid** flag leaf width

Figure 4.4 Slopes of linear regression models fitted to NRCT over thermal time plotted against initial NRCT. **(A)** Plot-level correlation for NRCT based on date-wise spatially corrected NRCT values (compare Figure S 7.4B); **(B)** Genotype-level correlation for spatially corrected initial NRCT and slopes derived from date-wise spatially corrected NRCT values; The red line represents the least squares line, reported correlation coefficients (r) are Pearson product moment correlation coefficients.

At the plot level, slopes were negatively correlated with initial NRCT if slopes were extracted for NRCT$_{corr}$ (r = -0.30, Figure 4.4A). This pattern was observable also at the genotype level, i.e. using average spatially corrected slopes and initial NRCT (Figure 4.4B).

Only weak correlations were found between slopes fitted to NRCT$_{corr}$ and any of the examined covariates ($0.07 \leq r \leq 0.24$), with the highest correlation found for heading and final height (Table 4.1). Instead, time-integrated measures of NRCT (predicted NRCT at OnSen and at mid stay-green) were moderately correlated with FH and the Ratio1200 (r = 0.56 and r = 0.58; and r = 0.43 and r = 0.56, respectively; Table 4.1), and less with heading (r = 0.20 and r = 0.32, Table 4.1). Overall, the correlation between time-integrated measures of NRCT and covariates was lower than at specific time points (Table 4.1, Figure 4.3).

The portion of variance in slopes fitted to NRCT$_{corr}$ explained by covariates (adj. $R^2 = 0.21$) was much lower than for NRCT at individual time points ($0.43 \leq$ adj. $R^2 \leq 0.67$). When the initial NRCT was added as a predictor, the portion of the explained variance increased (adj. $R^2 = 0.39$, Table 4.1), due to the correlation between the initial NRCT and the extracted slope (Figure 4.4).

4.3.3 Combination of spectral and thermal measurements

No significant correlations were observed between any of the slope$_{SVI}$ and the slope$_{CT}$ (Figure 4.5). Therefore, we aimed to determine whether these contrasting trends could be genotype-specific, based on the three check cultivars included in the experimental design. A rather large variation in the difference of slopes was observed for SURETTA and CH CLARO, whereas for CH NARA values were more similar across replicates (Figure S 7.7, Figure S 7.6). CH NARA showed a clearly different contrast between the trend in NRCT and NRSVI, particularly if the NDVI was used as a greenness indicator (Figure 4.6, Figure S 7.7, Figure S 7.6). In spite of the relatively
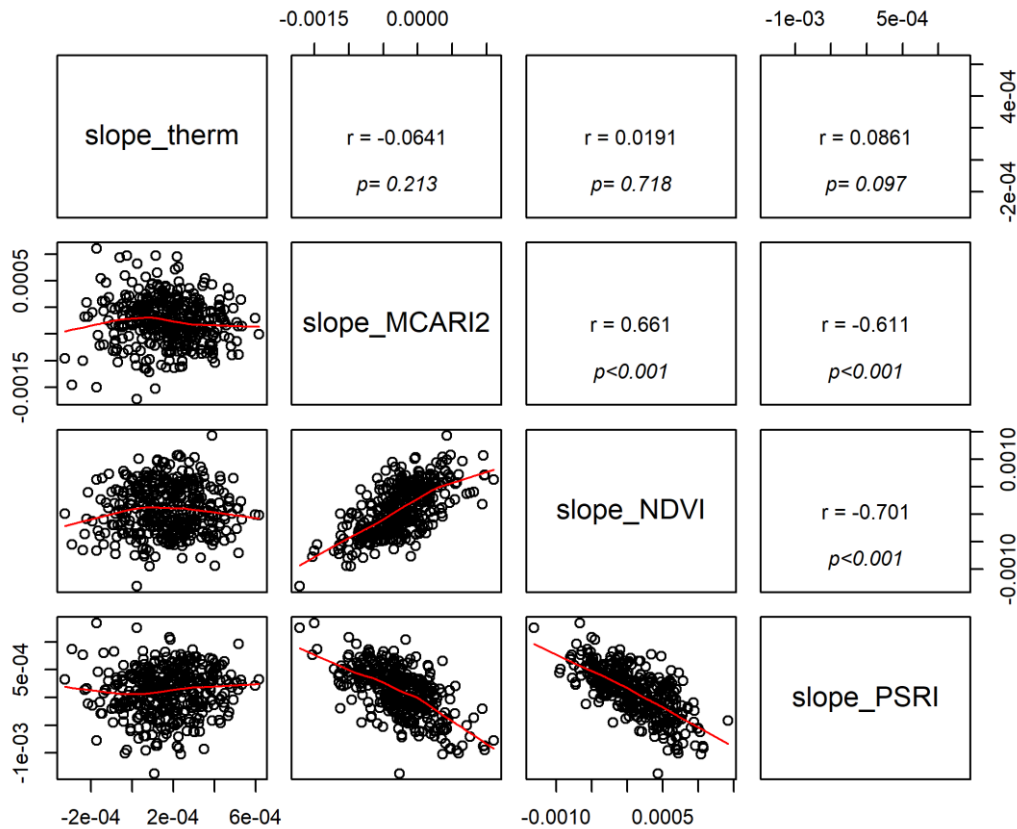
Figure 4.5 Correlations among slopes of the linear regression models fitted to normalized relative canopy temperature (NRCT) and normalized relative spectral vegetation indices (NRSVI). Slopes were obtained for date-wise spatially corrected NRCT and NRSVI against thermal time.

large variation, these preliminary results suggest that it may be possible to identify population extremes, i.e. genotypes showing strongly contrasting patterns in frequently used canopy greenness or senescence proxies and canopy temperature during the stay-green phase.

4.4 Discussion

Canopy temperature (CT) has been proposed as an indirect selection criterion under heat and drought conditions at least since the 1980s (*e.g.*, Singh and Kanemasu, 1983) due to the effect of canopy water status and transpiration rates on leaf temperature. However, strong effects of phenology, morphology and canopy structural parameters on canopy temperature have been demonstrated in a multitude of studies (*e.g.*, Ayeneh *et al.*, 2002; Giunta *et al.*, 2008; Rebetzke *et al.*, 2013; *see* Prashar and Jones, 2014 *for a review*). To limit such effects, it has been recommended to tightly control for confounding factors and restrict comparisons to material with very similar phenology and morphology (Lopes and Reynolds, 2010; Prashar and Jones, 2014). However, if CT is to be incorporated as a secondary trait indicative of performance under heat or drought in early selection as proposed by e.g. Rutkoski *et al.* (2016) and Sun *et al* (2019), the presence of such confounding factors, particularly differences in above-ground biomass or leaf area index, must be expected. An accurate characterization of the screened material for all factors
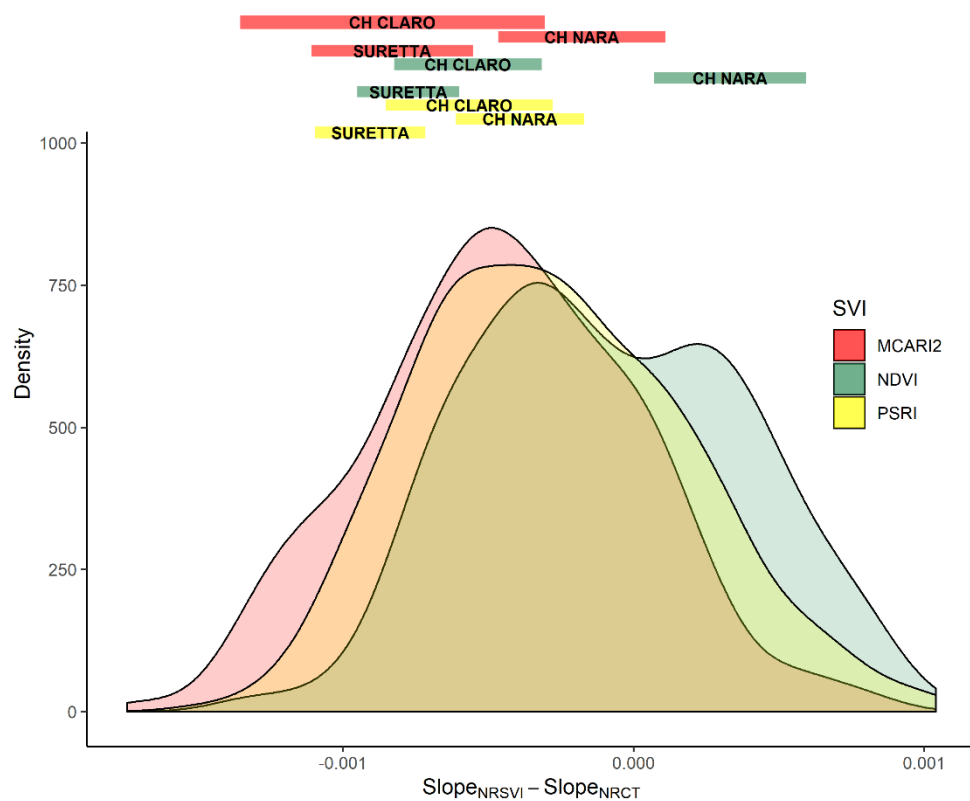
Figure 4.6 Distribution of the differences between slopes fitted to date-wise spatially corrected normalized relative spectral vegetation indices (NRSVI) and normalized relative canopy temperature (NRCT). The density plot shows the distribution based on three different SVI for all 378 experimental plots, sown with 354 different wheat genotypes. The colored bars represent the range observed for the three check cultivars CH CLARO, CH NARA and SURETTA, grown each in seven replicate plots, i.e. in a total of 21 plots, evenly distributed across the entire experimental field. Density curves represent smoothed histograms, using a Gaussian kernel, with the area under the curves standardized to equal unity.

with sizable confounding effects on CT as the trait of interest itself (i.e. water status or transpiration rates) may offer a solution as suggested by Deery *et al*. (2019). These factors may be used as covariates in a statistical model. Alternatively, relative temporal changes in CT at the level of individual plots may be much less affected by confounding factors, and therefore reflect dynamic genotype-specific responses to developing stress conditions. A similar approach was previously used successfully to track senescence dynamics and detect and quantify foliar diseases in diverse wheat germplasm using repeated spectral reflectance measurements (Anderegg *et al*., 2019a, 2020). In this study, we used a large set of phenologically and morphologically diverse genotypes – hence not ideally suited for the analysis of CT in terms of drought or heat stress resistance – to evaluate the potential and feasibility of this approach.

4.4.1   Repeatability of trends and effects of covariates

A key finding of this study is the high repeatability of temporal trends in CT, which are relatively independent of important covariates compared to measurements at individual time points (Figure 4.3, Table 4.1). In particular, the repeatability of 0.64 for slope$_{NCRT}$ was higher than the repeatability of date-specific NRCT values obtained for the same experiment during the same

period ($0.4 \leq w^2 \leq 0.61$). This finding is promising, particularly in light of the low correlation between covariates and trends. The investigated covariates were assumed not to vary greatly during the measurement period. This is a valuable assumption for final height, shoot biomass and flag leaf glaucousness, length and width. However, flag leaf angles and other canopy structural parameters may undergo significant changes over time, particularly towards the end of the stay-green phase. For example spikes may increasingly bend over and cover a larger fraction of the plot area. Such effects were identified as a likely cause of imprecise measurement of the onset of senescence by means of spectral reflectance in the same experiment (Anderegg *et al.*, 2020). If such changes in canopy structure are heritable, for example because they are closely related to plant height, lodging resistance, peduncle length or ear size and weight, this may also contribute to the observed repeatability, as ears differ strongly in their surface temperature from the rest of the canopy (Ayeneh *et al.*, 2002; Fernandez-Gallego *et al.*, 2019). Similarly, such canopy structural changes may affect the soil fraction in images over time, likely with consequences on extracted plot-based CT values (Deery *et al.*, 2019). This highlights the importance of restricting the analysis of temporal trends to the period of early grain filling before major changes in covariates occur.

A significant correlation was observed between the slope$_{NRCT}$ and the initial NRCT (Figure 4.4). Since the initial NRCT is based on a single time point measurement of CT, it follows that covariates with an effect on CT at this stage will equally affect the slope. This is demonstrated by the results of multiple linear regressions (Table 4.1). A possible explanation for this correlation may be that genotypes producing less biomass (thus with a higher initial NRCT) have consumed less water and therefore run out of water later than genotypes that have accumulated much more biomass (and thus have a lower initial NRCT). Another possible explanation for the observed correlation is that the slopes for very high and very low initial NRCT cannot be positive and negative, respectively. It is thus at least partly attributable to the approach taken here to cope with environmental variability across different measurement time points. This issue could be avoided by using non-linear models to derive temporal trends in NRCT. This was not possible here due to the inherently limited period of time that could be used for the analysis, resulting in merely five measurement time points. The window of opportunity could, however, be significantly enlarged by decreasing the variation in phenology. Here, a difference of 12 d was observed for heading between the earliest and the latest genotypes. In addition, given the low operational costs of the used platform (Perich *et al.*, 2020), frequency of flights could be significantly increased.

4.4.2 Combined thermal and spectral measurements to characterize stay-green

Particularly under heat and drought stress, the stay-green phenotype (in this case interpretable as the avoidance of premature senescence) has been proposed as a promising selection criterion in wheat (Christopher *et al.*, 2008; Lopes and Reynolds, 2012). Stay-green and the dynamics of senescence can be tracked using repeated spectral reflectance measurements (Anderegg *et al.*, 2020). However, the productivity of the retained green leaf area cannot be accurately quantified using spectral measurements. In contrast, thermal imaging has the potential to identify genotypes with high and stable transpiration rates (Rebetzke *et al.*, 2013), which is presumably related to a high photosynthetic activity (Fischer *et al.*, 1998; Richards, 1996). Thus, a combination of

spectral and thermal sensors may facilitate the selection of genotypes exhibiting ideotypic stay-green coupled with a high productivity of the retained green leaf area (Gregersen *et al.*, 2008; Jagadish *et al.*, 2015; Rebetzke *et al.*, 2016).

The absence of a significant correlation between any of the slope$_{SVI}$ and the slope$_{CT}$ indicated an independent temporal evolution of canopy reflectance and canopy temperature during the stay-green phase (Figure 4.5). This provides additional support for the conclusion drawn earlier that trends in the analyzed data set were not affected by the differences in phenology. If that were the case, both SVI and CT trends would be strongly driven by the onset of senescence, necessarily resulting in correlation between the two. The absence of such a correlation, thus, allowed exploring the potential to derive genotype-specific contrasts between these two temporal trends.

Replications of the check cultivars showed a relatively large variation for the contrast between the two temporal trends (Figure 4.6). In particular, this variation was rather large compared to the overall variation observed for all 378 experimental plots (Figure 4.6), especially for the cultivars CH CLARO and SURETTA. Nonetheless, it appeared that differences between genotypes may be present and that at least population extremes should be clearly separable. Since increased sink capacity appears to increase photosynthetic activity and stomatal conductance (Richards, 1996), we hypothesize that it may be possible to separate functional stay-green combined with a high photosynthetic activity from dysfunctional or partly functional stay-green or a stay-green phenotype resulting from source – sink imbalances. However, it should be noted that variation in the examined plant material is larger than what would be typically observed in a commercial breeding program. As pointed out previously, it will be critical to clarify whether observed trends may be affected by canopy structural changes over time, which could also result in genotype-specific contrasts between the two temporal trends even in the absence of genotypic differences in photosynthetic activity during the stay-green.

4.5 Conclusions

This study evaluated the potential of time-resolved canopy temperature measurements as indicators of resistance to short-term drought periods. We found temporal trends in normalized relative canopy temperature to be repeatable and not strongly affected by genotypic diversity and varying environmental conditions across measurement dates, as opposed to single time point measurements. Such trends may thus represent genotype-specific reactions to increasing stress levels over time. Furthermore, in combination with spectral measurements, a time-integrated analysis of CT may enable an accurate characterization of the degree of functionality of the stay-green phase. Further evaluations will need to clarify the degree of genetic correlation between observed trends and primary target traits under terminal drought conditions.

5   General Discussion

5.1   Multiple aspects of stay-green and consequences for its remote assessment

The main aim of this thesis was to evaluate the potential of remote sensing based high throughput methods to characterize genetically diverse material with respect to important secondary breeding traits during the grain filling stage. In this context, the most obvious phenomenon to investigate is stay-green and the dynamics of senescence (*see* Chapter 1 *for a discussion of the relevance of such assessments*). From a technical remote sensing perspective, this is in itself a readily achievable objective, given the obvious changes in canopy characteristics associated with the transition from carbon assimilation during the stay-green phase to N remobilization during senescence. Indeed, a multitude of studies has successfully used remote sensing based canopy indicators to distinguish stay-green from early-senescing genotypes in various crops, including maize, sorghum and wheat, using relatively low-cost equipment (*e.g.*, Christopher *et al*., 2014; Lopes and Reynolds, 2012; Makanza *et al*., 2018; Potgieter *et al*., 2017). However, as discussed in Chapters 1 and 2, the regulation of senescence and green leaf area dynamics is highly complex. The effects of this dynamics on key primary breeding traits such as GY and GPC are prone to strong genotype-by-environment interactions. The genetic and physiological determinants of the dynamics are still poorly understood. This is stressed here not to question the relevance of work interpreting stay-green as a stress-adaptive trait contributing to the avoidance of sink limitation during grain yield formation. On the contrary, such stress situations during grain filling are predicted to occur with increasing frequency and severity in many major agricultural production systems around the world (Lehner *et al*., 2006; Trnka *et al*., 2015). As a consequence, there is an urgent need to develop adapted genotypes. It is widely recognized that the stay-green phenotype and a high throughput method to assess senescence dynamics may significantly contribute to crop adaptation to changing climatic conditions (Thomas and Ougham, 2014). However, as discussed in Chapters 1 and 2, the stay-green phenotype may arise from a multitude of underlying factors, some of which are deemed desirable in specific environments, others entirely undesirable (Borrell *et al*., 2003; Sadras *et al*., 2019). This is the case in particular for bread wheat grown in temperate high-yielding environments with a low probability of encountering yield-relevant terminal heat and drought stress. Under such conditions, green leaf area dynamics are strongly affected by an interplay of phenology driven by developmental signals, the N economy at the canopy level, water availability and foliar diseases. Thus, stay-green cannot simply be interpreted as the avoidance of a premature senescence that would hamper grain filling and possibly result in source limitation. Furthermore, significant effects on GPC are expected and must equally be considered.

With this complexity, it appears difficult to capitalize on variation in stay-green without a basic understanding of the underlying physiological and possibly genetic determinants. From an extensive literature review, Gregersen *et al*. (2008) concluded that "a large and highly productive green leaf area should be kept as long as possible, but should also be able to respond as quickly as possible to environmental stress such as heat or desiccation and mobilize nutrients with high

efficiency". This highlights that in high-yielding environments, optimizing stay-green and the dynamics of senescence requires a more detailed characterization of green leaf area dynamics and its determinants than in environments with well-defined stress patterns and constraints to grain yield improvement.

This requirement motivated the choice of sensors, their combined use as well as the data analysis approaches taken in this thesis. We believe, like others (*e.g.*, Deery *et al*., 2019; Gregersen *et al*., 2008; Rebetzke *et al*., 2016), that a multidimensional characterization of the stay-green and senescence phase (w.r.t. green leaf area, its productivity, remobilization of nutrients and the underlying determinants) is required and requires in turn a time-integrated multi-sensor approach.

With this thesis, we focused on the development of methods to enable accurate phenotyping in the final growth stages of wheat. A detailed investigation of the physiological and genetic basis of stay-green and senescence and corresponding effects on GY and GPC was beyond the scope of this thesis and cannot be reliably addressed based on the data collected here. Investigations into the regulation of senescence and direct effects on GY and GPC will have to be carried out using genetic material not segregating for major genes regulating development (i.e. *Vrn* and *Ppd*; Bogard *et al*., 2010; Lopes and Reynolds, 2012, 2010; Pinto *et al*., 2010). Given the effect of varying plant height through the *Rht* genes on the sink- source ratio, segregation in these genes should equally be avoided (Borrell *et al*., 2003; Miralles and Slafer, 1996). High throughput phenotyping will enable the screening of multiple homogeneous populations or large diverse populations which can be analyzed in subsets with similar phenology. Finally, experiments should arguably be carried out under environmental conditions expected to maximize direct effects of the grain filling phase on primary traits (Jackson *et al*., 1996, V. Allard, personal communication).

5.2 Assessments of visually observed senescence dynamics

In Chapter 2, we evaluated the potential of hyperspectral reflectance measurements to accurately track visually observed canopy senescence dynamics and possibly to provide a better representation of the most relevant (in terms of their effect on GY and GPC) changes occurring at the canopy level during grain filling. The PSRI was identified as the optimal spectral vegetation index in terms of precision and robustness to track visually observed canopy senescence dynamics.

The use of the PSRI does not *per se* provide additional information when compared to visual scorings. A more accurate characterization of the stay-green phase and the elusive switch from carbon assimilation to remobilization (Jagadish *et al*., 2015; Thomas and Ougham, 2014) cannot be provided. Instead, benefits may arise primarily from an increased temporal resolution and throughput (*see* Chapter 2). However, this alone may be of value in several ways:

(i)   An often proposed physiological strategy to improve wheat yields consists in adjusting phasic development through photoperiod-sensitivity (Pérez-Gianmarco *et al*., 2018; Slafer *et al*., 2001; Slafer and Rawson, 1996) and temperature-sensitivity of developmental processes (Atkinson and Porter, 1996). Given the large effects of

major genes regulating development on senescence dynamics (*e.g.*, Bogard *et al.*, 2011), the evaluation of such strategies will have to take effects on the entire life cycle into account. In this context, the onset of senescence is a cardinal point requiring assessment.

(ii)   The onset of senescence represents a key junction between C and N metabolism at canopy level (Bogard *et al.*, 2011), and trade-offs are likely to contribute to the observed negative correlation between GY and GPC (*see* Chapter 1). Due to the complexity of the regulation of senescence *per se* and strong context-dependency of its effects on GY and GPC, it seems likely that only intense field testing under contrasting environmental conditions will allow for a definition and possibly selection of environment-specific ideotypes.

(iii)   Under well-characterized terminal stress conditions, a delayed senescence may be safely interpreted as an adaptive trait (*see* Chapter 1), and used as an indirect selection criterion for stress resistance.

(iv)   Phase-dependent environmental indices (Millet *et al.*, 2019) require a genotype specific timing of developmental stages, one of which is physiological maturity, which may be approximated by the onset or end of the rapid senescence phase (Pask *et al.*, 2012).

For these purposes (i.e. to quantify the onset of senescence as a cardinal point in plant development), identifying the onset of senescence as the onset of visually observable chlorophyll breakdown is likely to be precise enough. In this context, the question arises whether this could be achieved using RGB photography, which would reduce equipment costs and simplify data processing and analysis. This approach has not been tested here and can only be evaluated on color-calibrated images to remove the effect of unstable ambient illumination conditions varying from one measurement campaign to another, as done for example by Grieder *et al.* (2015). In contrast, the exploitation of color information from uncalibrated images to derive time-integrated descriptors of vegetation is challenging, even for comparably simple tasks such as the segmentation of vegetation from the soil background, where absolute color information is not used explicitly (*see e.g.*, Sadeghi-Tehran *et al.*, 2017; Yu *et al.*, 2017, *compare with* Figure 2.6 *for an illustration of such effects*).

5.3   Beyond visual - Assessing GPC and the N economy of canopies

As pointed out in Chapter 1, the stay-green phenotype is unlikely to confer yield benefits under the conditions of our experiments, because source capacity is largely adequate to allow for complete grain filling. In contrast, the N economy of developing grains and the canopy appears to be more strongly interconnected with post-anthesis processes and senescence in particular, even under high-yielding conditions. In this context, several physiological traits have been identified

which may contribute to improvements of NUE and GPC. These include post-anthesis N uptake, which may be of particular value because it increases GPC without reducing GY (Bogard *et al*., 2010; Kichey *et al*., 2007; Taulemesse *et al*., 2016), and N remobilization dynamics, which have been identified as an important determinant of N utilization efficiency and consequently, overall N use efficiency and GPC (Gaju *et al*., 2011, 2014; Kong *et al*., 2016).

Here, the spectroradiometer was chosen for evaluations due to its (at least theoretically) unrivalled potential to characterize vegetation in terms of biochemical properties, physiological status and healthiness (*see e.g.*, Araus *et al*., 2018; Mahlein, 2016; Yendrek *et al*., 2017). Several studies have evaluated spectral measurements during grain filling as a tool to assess traits related to the N economy of wheat canopies and developing grains. Many studies revealed that traits related to N content and N uptake of individual organs or the canopy are best assessed during the milk- or dough ripeness stages using red-edge based SVI (Barmeier and Schmidhalter, 2017; Frels *et al*., 2018; Prey *et al*., 2020; Prey and Schmidhalter, 2019a, 2019b). This holds true for predictions of N traits at anthesis (in this case sometimes referred to as "post-dictions", *e.g.*, Prey *et al*., 2020), for predictions of N traits at maturity and for predictions of N traits integrating over the entire grain filling phase (e.g. post-anthesis N uptake, N remobilization or N utilization efficiency). While such studies have a long tradition in the domains of plant nutrition and precision agriculture, more recently the interest in variety differentiation has increased (Frels *et al*., 2018; Prey *et al*., 2020; Prey and Schmidhalter, 2019a). For N uptake at anthesis and maturity, some potentially useful correlations with spectral indicators were observed. In contrast, correlations are substantially lower for N translocation efficiency, post-anthesis N uptake and N harvest index (*e.g.*, Prey and Schmidhalter, 2019a). N uptake after anthesis is closely related to continued biomass acquisition by stems, serving as a temporary sink (Taulemesse *et al*., 2016), which may help to maintain a high root activity (V. Allard, personal communication). In contrast, sensors detect the horizontally aligned leaf blades whereas occluded organs may be assessed only indirectly *via* correlation in biochemical characteristics with measured organs. Therefore, N uptake in stems is often only weakly correlated with spectral indicators (Barmeier and Schmidhalter, 2017; Prey and Schmidhalter, 2019a), suggesting it may be difficult to assess post-anthesis N uptake *via* a direct functional relationship with spectral reflectance.

As pointed out by Prey *et al*. (2020) and Prey and Schmidhalter (2019a), correlations are likely to arise as a consequence of contrasting senescence status. Such correlations can arguably arise (i) directly, because senescence and N remobilization are tightly linked processes and their timing is often reported to have significant effects on N traits (Masclaux *et al*., 2001; Sinclair and Wit, 1975; Uauy *et al*., 2006); (ii) as a spurious relationship, because both N traits and senescence dynamics are strongly related to anthesis date (Bogard *et al*., 2011; Foulkes *et al*., 2004; Gaju *et al*., 2011, 2014; Xie *et al*., 2016), (iii) through interactions between management interventions and anthesis date. These result in genotypes with contrasting phenology receiving treatments at different growth stages, affecting the partitioning of N uptake between phases and the duration of the stay-green phase *via* modified relative N availability across growth stages (Bogard *et al*., 2010, 2011); finally (iv) through interactions between anthesis date and environment, resulting in genotypes with contrasting phenology perceiving contrasting environments in given phenophases.

Variation in flowering time is often not taken into account when evaluating correlations between spectral indicators and N traits. Instead, average growth stages at treatment interventions and measurements are normally reported. Consequently, correlations (even genetic correlations as reported by e.g. Frels *et al.* (2018)) may suffer from a strong dependency on genotype and environment, because the timing of senescence (in chronological time after sowing) is associated with N traits in an unstable manner, as it critically depends on anthesis date. This potentially limits the generalizability of the developed secondary traits (*see* Chapter 1). Such effects may be a cause of the frequently observed environment-specificity of optimal growth stages for assessments of N traits (Erdle *et al.*, 2013; Frels *et al.*, 2018; F. Li *et al.*, 2014; Pavuluri *et al.*, 2015).

Under the hypothetical scenario of differences in senescence dynamics representing indeed the dominant factor underlying the predictability of N traits, the approach taken in Chapter 2 of this thesis may be advantageous because the dynamics of senescence are explicitly described and expressed relative to heading date (a proxy measure for flowering). In this way, several consecutive measurements are functionally integrated over time based on prior physiological knowledge on the dynamics of the process (Gaju *et al.*, 2011; Gooding *et al.*, 2000; Moreau *et al.*, 2008). This may have the additional benefit of reducing the effect of measurement errors related to a particular measurement date on the estimation of the target trait. Furthermore, given the dynamic nature of senescence, modelling its dynamics allows to take spatial patterns in field experiments affecting these dynamics into account when estimating genetic effects (*see e.g.*, van Eeuwijk *et al.*, 2018). A time-integrated analysis allows for the derivation of independent parameters describing different aspects of the dynamics of senescence (*see e.g.*, Kong *et al.*, 2016; Moreau *et al.*, 2008; Xie *et al.*, 2016*; see* Chapter 2 *for details*). Furthermore, we propose the exploitation of relative temporal changes in reflectance characteristics at the level of individual plots as a means to control for confounding effects on spectral reflectance caused by genotypic diversity and varying environmental conditions. Frequent measurements are becoming increasingly feasible and affordable with technical advances in platform and sensor technology (*see e.g.*, Aasen *et al.*, 2015; Aasen and Bolten, 2018).

Due to the context-specificity of effects of senescence on N traits and GPC, it would be preferable to measure relevant physiological processes occurring during the grain filling phase that are directly linked with N traits (e.g. post-anthesis N uptake and N remobilization) "as they occur". Unfortunately, this appears infeasible given the spectacular changes in canopy structural characteristics during late development. These changes are functionally largely unrelated with N uptake and remobilization processes. However, they have massive effects on reflectance characteristics of canopies in the entire measureable spectrum. In Chapter 3, we showed that even in the absence of visually apparent changes in canopy structure during the stay-green phase, values of spectral indices such as the structure-insensitive pigment index (SIPI) change quite drastically in short time when set in relation to the time span between early and late flowering genotypes (Figure 3.4B, Figure 3.4C). In fact, it seemed that even treatments such as inoculations only moderately affected canopy reflectance, compared to changes over time, highlighting the importance of accounting for contrasting phenology among genotypes. It has been argued that the use of the full spectrum instead of just a few wavelengths may facilitate the detection of subtle

differences in canopies (Araus *et al.*, 2018). However, these methods suffer from a lack of generalizability across genotypes and environments, which is a key requirement for an efficient implementation in breeding (Araus *et al.*, 2008). This problem was obvious also in our work despite the apparent simplicity of the task (Figure 2.7, Figure 2.8). This led us to conclude that the extraction of tailored and thoroughly validated spectral features may be preferable over a "brute-force" modelling approach. Several spectral indices were developed taking the effect of common confounding factors explicitly into account (*e.g.*, Haboudane *et al.*, 2004). Nonetheless, many spectral indices are also strongly affected by canopy structural changes, limiting their predictive ability for biochemical or physiological changes.

Possibly a part of the predictive ability of spectral indicators stems from a direct relationship between N traits of canopies and spectral reflectance (as opposed to an indirect relationship *via* the timing of senescence). However, given the strong effect of phenology, it appears that only a time-integrated analysis may be able to reveal relevant differences among genotypes. We found marked differences in the dynamics of different spectral indices, suggesting the potential to capture different aspects of senescence (Figure 2.4). However, our analysis of time-courses of spectral indices delivered rather inconclusive results. It appeared that visual scorings best captured GY-relevant aspects of senescence when a correlation existed between senescence dynamics and GY. This was probably because scorings were less affected by canopy structural changes over time (particularly spike orientation). This finding was in contrast to the findings of the same analysis for GPC. Here, NIR/VIS indices apparently better captured relevant processes, a pattern which was comparably stable across the two years for which data was available (Figure 2.10B). In time point specific analyses, most authors found a superior predictive power of red-edge based indices over NIR-based indices. The superior performance of these indices in assessing N traits during grain filling is frequently ascribed to the absence of saturation effects as opposed to NIR-based indices such as the NDVI (Barmeier and Schmidhalter, 2017; Frels *et al.*, 2018; Prey *et al.*, 2020; Prey and Schmidhalter, 2019a). We observed decreasing NDVI-values before any visual symptom of senescence could be observed, which seemingly contradicts the assumption of saturation (Figure 2.5). However, it also seemed that NIR-based indices react sensitively to canopy structural changes (Figure 2.6), although we were not able to analyze this relationship quantitatively. Thus, it seemed that an increased contribution of spikes to the signal rather than biochemical changes occurring in the subtending canopy were responsible for the observed decline in NDVI values.

5.4 A time-integrated multi-sensor approach to characterize green leaf area dynamics - Towards an understanding of factors driving senescence under field conditions?

Despite its importance for grain yield and N yield formation, senescence and its physiological and genetic determinants under field conditions remain relatively poorly understood (Jagadish *et al.*, 2015). In the special case of the Swiss federal wheat breeding program of Agroscope, effects of foliar diseases may interact with or overlay physiological and genetic factors regulating

senescence, as breeders select without fungicide applications. Given this complexity, an improved understanding of mechanisms determining GLAD under specific environmental conditions may be required to better understand and exploit the functional relationship with primary traits for breeding purposes. We hypothesized that a combination of different sensors and a time-integrated approach may offer insights due to the potential to track multiple aspects of GLAD over time.

In Chapter 3, we demonstrated the potential to distinguish between disease-driven GLAD and GLAD primarily affected by physiological senescence. Such a distinction may allow for a better interpretation of the correlations between GLAD and primary traits. This is prominently illustrated by the difference observed between the wet season of 2016 and the dry and hot seasons of 2017 and 2018 in our experiments. Although in the three year-locations investigated in this work, a well-informed breeder would probably be able to provide a solid interpretation of the observed correlations between GLAD and primary traits without much phenotyping data, the characterization of environments may often not be as simple as in these relatively extreme cases. Thus, a presence of several different factors influencing GLAD should always be expected. Accordingly, a larger number of possible mechanisms conferring stay-green or early senescence may be present simultaneously within experiments, some of which may be more or less beneficial. A coarse, overall analysis of the effects of GLAD on primary traits may therefore mask beneficial genetic variation, if a more detailed characterization is not achieved.

In Chapter 4, we evaluated the potential to combine spectral and thermal measurements to more accurately characterize genotype performance during the stay-green phase. Stay-green is a broad phenotype, which can represent the underlying genotypic driver of assimilation, but can also simply reflect slowed water use, greater N uptake or slowed N remobilization (Rebetzke *et al.*, 2016 *and citations therein*). As a proxy of transpiration rates, CT measurements may therefore complement measurements of canopy greenness to distinguish between different physiological determinants of stay-green. We were able to extract repeatable temporal trends of CT, in the dry and hot summer of 2018, without applying drought treatments. These trends were not correlated with trends in spectral canopy greenness indicators. This suggested that multiple temporal trends derived from different sensors may indeed provide complementary information to more accurately characterize stay-green. However, it is stressed here that the value of this approach requires further evaluation, including assessments of repeatability and heritability and investigations into the genetic correlation with primary breeding target traits (Araus *et al.*, 2008; Jackson *et al.*, 1996).

5.5 Potential of high throughput phenotyping of green leaf area dynamics for physiological wheat breeding

As pointed out previously, the development of an ideotype for the tested environments is beyond the scope of this thesis and will require extensive additional investigations. Consequently, the question whether GLAD, stay-green and senescence dynamics may represent a valuable secondary trait for a physiological breeding strategy can only be addressed in terms of its technical feasibility.

Despite the small plot size used in our work, we found moderate to high repeatability within years and a moderate to high heritability across years for senescence dynamics parameters extracted from repeated canopy reflectance measurements. Importantly, this also held true for parameters describing the duration (and thus the rate) of senescence (Table 2.1). A similar correlation for spectral and visual parameters with primary traits (derived genotypic BLUEs) appears to justify the use of spectral indices instead of laborious visual scorings. However, only the upscaling to phenotyping platforms such as UAVs will deliver the required throughput to investigate the value of senescence dynamics as a secondary trait in breeding programs. Such an implementation will require additional prior validation of spectral indices. Furthermore, trait heritability may be overestimated here because it was estimated in an unrealistically genetically diverse population that is representative of breeding populations (Jackson *et al*., 1996).

There is some evidence that senescence dynamics may affect GY in other ways than by relaxing a source capacity constraint. For example, grain filling dynamics and senescence dynamics may be related. For example, a delayed but short and fast senescence was related to a shortened grain filling phase but higher grain filling rates in a doubled haploid mapping population (derived from 'Forno'×'Oberkulmer'; Xie *et al*., 2016). Final individual grain weight appears to be more strongly related to grain filling rates than grain filling duration (Xie *et al*., 2016a; Zahedi and Jenner, 2003). Thus, if senescence dynamics and grain filling dynamics are related, senescence dynamics could be used as a readily accessible indicator for grain filling dynamics. This appears to be of particular interest under heat stress conditions, where temporal constraints to grain filling may become limiting, even in the absence of source limitation (García *et al*., 2016). However, under increased night temperatures, the shortened duration of grain filling was not associated with changes in green leaf area dynamics (García *et al*., 2016).

Detailed direct assessments of yield and quality formation during grain filling are not possible using current high throughput phenotyping techniques. However, as discussed throughout this thesis, correlations of green leaf area dynamics with primary target traits or with secondary traits conferring yield or quality benefits under certain environmental conditions are frequently observed. This suggests that senescence dynamics may be used as a readily available indicator for important physiological and developmental processes determining yield and quality. Relatively low-cost and high throughput post-harvest assessments of yield components (particularly individual grain weight) and N traits at maturity will help to elucidate these correlations and their dependency on environmental conditions. From our work, a time-integrated multi-sensor approach emerges as the most promising way to characterize green leaf area dynamics and investigate causal genetic and physiological factors. Upscaling to mobile phenotyping platforms will provide the required throughput. We anticipate that this will allow for the identification of combinations of stay-green phenotypes and environments with stable positive effects on yield and quality.

6 References

Aasen, H., Bolten, A., 2018. Multi-temporal high-resolution imaging spectroscopy with hyperspectral 2D imagers – From theory to application. Remote Sensing of Environment 205, 374–389. https://doi.org/10.1016/j.rse.2017.10.043

Aasen, H., Burkart, A., Bolten, A., Bareth, G., 2015. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. ISPRS Journal of Photogrammetry and Remote Sensing 108, 245–259. https://doi.org/10.1016/j.isprsjprs.2015.08.002

Aasen, H., Honkavaara, E., Lucieer, A., Zarco-Tejada, P.J., 2018. Quantitative Remote Sensing at Ultra-High Resolution with UAV Spectroscopy: A Review of Sensor Technology, Measurement Procedures, and Data Correction Workflows. Remote Sensing 10, 1091. https://doi.org/10.3390/rs10071091

Acreche, M.M., Slafer, G.A., 2009. Variation of grain nitrogen content in relation with grain yield in old and modern Spanish wheats grown under a wide range of agronomic conditions in a Mediterranean region. The Journal of Agricultural Science 147, 657–667. https://doi.org/10.1017/S0021859609990190

Amani, I., Fischer, R.A., Reynolds, M.P., 1996. Canopy Temperature Depression Association with Yield of Irrigated Spring Wheat Cultivars in a Hot Climate. Journal of Agronomy and Crop Science 176, 119–129. https://doi.org/10.1111/j.1439-037X.1996.tb00454.x

Ambroise, C., McLachlan, G.J., 2002. Selection bias in gene extraction on the basis of microarray gene-expression data. PNAS 99, 6562–6566. https://doi.org/10.1073/pnas.102102699

Anderegg, J., Hund, A., Karisto, P., Mikaberidze, A., 2019. In-Field Detection and Quantification of Septoria Tritici Blotch in Diverse Wheat Germplasm Using Spectral–Temporal Features. Front. Plant Sci. 10. https://doi.org/10.3389/fpls.2019.01355

Anderegg, J., Yu, K., Aasen, H., Walter, A., Liebisch, F., Hund, A., 2020. Spectral Vegetation Indices to Track Senescence Dynamics in Diverse Wheat Germplasm. Front. Plant Sci. 10. https://doi.org/10.3389/fpls.2019.01749

Apan, A., Held, A., Phinn, S., Markley, J., 2004. Detecting sugarcane 'orange rust' disease using EO-1 Hyperion hyperspectral imagery. International Journal of Remote Sensing 25, 489–498. https://doi.org/10.1080/01431160310001618031

Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: the new crop breeding frontier. Trends in Plant Science 19, 52–61. https://doi.org/10.1016/j.tplants.2013.09.008

Araus, J.L., Kefauver, S.C., Zaman-Allah, M., Olsen, M.S., Cairns, J.E., 2018. Translating High-Throughput Phenotyping into Genetic Gain. Trends in Plant Science 23, 451–466. https://doi.org/10.1016/j.tplants.2018.02.001

Araus, J.L., Slafer, G.A., Royo, C., Serret, M.D., 2008. Breeding for Yield Potential and Stress Adaptation in Cereals. Critical Reviews in Plant Sciences 27, 377–412. https://doi.org/10.1080/07352680802467736

Ashourloo, D., Mobasheri, M.R., Huete, A., 2014. Developing Two Spectral Disease Indices for Detection of Wheat Leaf Rust (Pucciniatriticina). Remote Sensing 6, 4723–4740. https://doi.org/10.3390/rs6064723

Asrar, G., Fuchs, M., Kanemasu, E.T., Hatfield, J.L., 1984. Estimating Absorbed Photosynthetic Radiation and Leaf Area Index from Spectral Reflectance in Wheat 1. Agronomy Journal 76, 300–306. https://doi.org/10.2134/agronj1984.00021962007600020029x

Atkinson, D., Porter, J.R., 1996. Temperature, plant development and crop yields. Trends in Plant Science 1, 119–124. https://doi.org/10.1016/S1360-1385(96)90006-0

Austin, R.B., 1980. Physiological limitations to cereal yields and ways of reducing them by breeding. Physiological limitations to cereal yields and ways of reducing them by breeding. 3–19.

Ayeneh, A., van Ginkel, M., Reynolds, M.P., Ammar, K., 2002. Comparison of leaf, spike, peduncle and canopy temperature depression in wheat under heat stress. Field Crops Research 79, 173–184. https://doi.org/10.1016/S0378-4290(02)00138-7

Bajwa, S.G., Rupe, J.C., Mason, J., 2017. Soybean Disease Monitoring with Leaf Reflectance. Remote Sensing 9, 127. https://doi.org/10.3390/rs9020127

Bancal, M.-O., Ben Slimane, R., Bancal, P., 2016. Zymoseptoria tritici development induces local senescence in wheat leaves, without affecting their monocarpic senescence under two contrasted nitrogen nutrition. Environmental and Experimental Botany 132, 154–162. https://doi.org/10.1016/j.envexpbot.2016.09.002

Bancal, M.-O., Robert, C., Ney, B., 2007. Modelling Wheat Growth and Yield Losses from Late Epidemics of Foliar Diseases using Loss of Green Leaf Area per Layer and Pre-anthesis Reserves. Ann Bot 100, 777–789. https://doi.org/10.1093/aob/mcm163

Barmeier, G., Schmidhalter, U., 2017. High-Throughput Field Phenotyping of Leaves, Leaf Sheaths, Culms and Ears of Spring Barley Cultivars at Anthesis and Dough Ripeness. Front. Plant Sci. 8. https://doi.org/10.3389/fpls.2017.01920

Barnes, E.M., Clarke, T.R., Richards, S.E., Colaizzi, P.D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R.J., Li, H., Moran, M.S., 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. Proceedings of the 5th International Conference on Precision Agriculture, Bloomington, Minnesota, USA, 16-19 July, 2000 1–15.

Becker, E., Schmidhalter, U., 2017. Evaluation of Yield and Drought Using Active and Passive Spectral Sensing Systems at the Reproductive Stage in Wheat. Front. Plant Sci. 8. https://doi.org/10.3389/fpls.2017.00379

Behmann, J., Bohnenkamp, D., Paulus, S., Mahlein, A.-K., 2018. Spatial Referencing of Hyperspectral Images for Tracing of Plant Disease Symptoms. Journal of Imaging 4, 143. https://doi.org/10.3390/jimaging4120143

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M.L., Bareth, G., 2015. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. International Journal of

Applied Earth Observation and Geoinformation 39, 79–87. https://doi.org/10.1016/j.jag.2015.02.012

Bernardo, R., 2016. Bandwagons I, too, have known. Theor Appl Genet 129, 2323–2332. https://doi.org/10.1007/s00122-016-2772-5

Bingham, I.J., Young, C., Bounds, P., Paveley, N.D., 2019. In sink-limited spring barley crops, light interception by green canopy does not need protection against foliar disease for the entire duration of grain filling. Field Crops Research 239, 124–134. https://doi.org/10.1016/j.fcr.2019.04.020

Blackburn, G.A., 1998. Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. International Journal of Remote Sensing 19, 657–675. https://doi.org/10.1080/014311698215919

Blake, N.K., Lanning, S.P., Martin, J.M., Sherman, J.D., Talbert, L.E., 2007. Relationship of Flag Leaf Characteristics to Economically Important Traits in Two Spring Wheat Crosses. Crop Science 47, 491–494. https://doi.org/10.2135/cropsci2006.05.0286

Blum, A., 2011. Drought resistance – is it really a complex trait? Functional Plant Biol. 38, 753–757. https://doi.org/10.1071/FP11101

Blum, A., Poiarkova, H., Golan, G., Mayer, J., 1983. I. Effects of translocation and kernel growth. Field Crops Research 6, 51–58. https://doi.org/10.1016/0378-4290(83)90047-3

Bogard, M., Allard, V., Brancourt-Hulmel, M., Heumez, E., Machet, J.-M., Jeuffroy, M.-H., Gate, P., Martre, P., Le Gouis, J., 2010. Deviation from the grain protein concentration–grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. J Exp Bot 61, 4303–4312. https://doi.org/10.1093/jxb/erq238

Bogard, M., Jourdan, M., Allard, V., Martre, P., Perretant, M.R., Ravel, C., Heumez, E., Orford, S., Snape, J., Griffiths, S., Gaju, O., Foulkes, J., Gouis, J.L., 2011. Anthesis date mainly explained correlations between post-anthesis leaf senescence, grain yield, and grain protein concentration in a winter wheat population segregating for flowering time QTLs. J. Exp. Bot. 62, 3621–3636. https://doi.org/10.1093/jxb/err061

Borrás, L., Slafer, G.A., Otegui, M.E., 2004. Seed dry weight response to source–sink manipulations in wheat, maize and soybean: a quantitative reappraisal. Field Crops Research 86, 131–146. https://doi.org/10.1016/j.fcr.2003.08.002

Borrell, A.K., Hammer, G.L., Henzell, R.G., 2000. Does Maintaining Green Leaf Area in Sorghum Improve Yield under Drought? II. Dry Matter Production and Yield. Crop Science 40, 1037–1048. https://doi.org/10.2135/cropsci2000.4041037x

Borrell, A.K., Oosterom, E.V., Hammer, G.L., Jordan, D.R., Douglas, A., 2003. The physiology of "stay-green" in sorghum., in: Solutions for a Better Environment: Proceedings of the 11th Australian Agronomy Conference.

Borrill, P., Harrington, S.A., Simmonds, J., Uauy, C., 2019. Identification of Transcription Factors Regulating Senescence in Wheat through Gene Regulatory Network Modelling. Plant Physiology 180, 1740–1755. https://doi.org/10.1104/pp.19.00380

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brisson, N., Gate, P., Gouache, D., Charmet, G., Oury, F.-X., Huard, F., 2010. Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. Field Crops Research 119, 201–212. https://doi.org/10.1016/j.fcr.2010.07.012

Brocklehurst, P.A., 1977. Factors controlling grain weight in wheat. Nature 266, 348–349. https://doi.org/10.1038/266348a0

Brown, J.K.M., Chartrain, L., Lasserre-Zuber, P., Saintenac, C., 2015. Genetics of resistance to Zymoseptoria tritici and applications to wheat breeding. Fungal Genetics and Biology, Septoria tritici blotch disease of wheat: Tools and techniques to study the pathogen Zymoseptoria tritici 79, 33–41. https://doi.org/10.1016/j.fgb.2015.04.017

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.J., Thompson, R., 2018. ASReml estimates variance components under a general linear 188.

Calanca, P., 2007. Climate change and drought occurrence in the Alpine region: How severe are becoming the extremes? Global and Planetary Change, Extreme Climatic Events 57, 151–160. https://doi.org/10.1016/j.gloplacha.2006.11.001

Cao, X., Luo, Y., Zhou, Y., Duan, X., Cheng, D., 2013. Detection of powdery mildew in two winter wheat cultivars using canopy hyperspectral reflectance. Crop Protection 45, 124–131. https://doi.org/10.1016/j.cropro.2012.12.002

Cartelle, J., Pedró, A., Savin, R., Slafer, G.A., 2006. Grain weight responses to post-anthesis spikelet-trimming in an old and a modern wheat under Mediterranean conditions. European Journal of Agronomy 25, 365–371. https://doi.org/10.1016/j.eja.2006.07.004

Chen, D., Huang, J., Jackson, T.J., 2005. Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near- and short-wave infrared bands. Remote Sensing of Environment 98, 225–236. https://doi.org/10.1016/j.rse.2005.07.008

Chen, P., Haboudane, D., Tremblay, N., Wang, J., Vigneault, P., Li, B., 2010. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. Remote Sensing of Environment 114, 1987–1997. https://doi.org/10.1016/j.rse.2010.04.006

Chenu, K., Chapman, S.C., Tardieu, F., McLean, G., Welcker, C., Hammer, G.L., 2009. Simulating the Yield Impacts of Organ-Level Quantitative Trait Loci Associated With Drought Response in Maize: A "Gene-to-Phenotype" Modeling Approach. Genetics 183, 1507–1523. https://doi.org/10.1534/genetics.109.105429

Chenu, K., Porter, J.R., Martre, P., Basso, B., Chapman, S.C., Ewert, F., Bindi, M., Asseng, S., 2017. Contribution of Crop Models to Adaptation in Wheat. Trends in Plant Science 22, 472–490. https://doi.org/10.1016/j.tplants.2017.02.003

Christopher, J.T., Christopher, M.J., Borrell, A.K., Fletcher, S., Chenu, K., 2016. Stay-green traits to improve wheat adaptation in well-watered and water-limited environments. J. Exp. Bot. 67, 5159–5172. https://doi.org/10.1093/jxb/erw276

Christopher, J.T., Manschadi, A.M., Hammer, G.L., Borrell, A.K., 2008. Developmental and physiological traits associated with high yield and stay-green phenotype in wheat. Crop and Pasture Science 59, 354–364. https://doi.org/10.1071/AR07193

Christopher, J.T., Veyradier, M., Borrell, A.K., Harvey, G., Fletcher, S., Chenu, K., 2014. Phenotyping novel stay-green traits to capture genetic variation in senescence dynamics. Functional Plant Biology 41, 1035–1048. https://doi.org/10.1071/FP14052

Christopher, M., Chenu, K., Jennings, R., Fletcher, S., Butler, D., Borrell, A., Christopher, J., 2018. QTL for stay-green traits in wheat in well-watered and water-limited environments. Field Crops Research 217, 32–44. https://doi.org/10.1016/j.fcr.2017.11.003

Clark, R.N., Roush, T.L., 1984. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. Journal of Geophysical Research: Solid Earth 89, 6329–6340. https://doi.org/10.1029/JB089iB07p06329

Clay, D.E., Kim, K.-I., Chang, J., Clay, S.A., Dalsted, K., 2006. Characterizing Water and Nitrogen Stress in Corn Using Remote Sensing. Agronomy Journal 98, 579–587. https://doi.org/10.2134/agronj2005.0204

Coombes, N., 2009. DiGGer.

Cormier, F., Foulkes, J., Hirel, B., Gouache, D., Moënne-Loccoz, Y., Le Gouis, J., 2016. Breeding for increased nitrogen-use efficiency: a review for wheat (T. aestivum L.). Plant Breed 135, 255–278. https://doi.org/10.1111/pbr.12371

Crain, J., Reynolds, M., Poland, J., 2017. Utilizing High-Throughput Phenotypic Data for Improved Phenotypic Selection of Stress-Adaptive Traits in Wheat. Crop Science 57, 648–659. https://doi.org/10.2135/cropsci2016.02.0135

Cullis, B.R., Smith, A.B., Coombes, N.E., 2006. On the design of early generation variety trials with correlated data. J. Agric. Biol. Environ. Stat. 11, 381–393. https://doi.org/10.1198/108571106X154443

Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. International Journal of Remote Sensing 25, 5403–5413. https://doi.org/10.1080/0143116042000274015

Datt, B., 1999. A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests using Eucalyptus Leaves. Journal of Plant Physiology 154, 30–36. https://doi.org/10.1016/S0176-1617(99)80314-9

Deery, D.M., Rebetzke, G.J., Jimenez-Berni, J.A., Bovill, W.D., James, R.A., Condon, A.G., Furbank, R.T., Chapman, S.C., Fischer, R.A., 2019. Evaluation of the Phenotypic Repeatability of Canopy Temperature in Wheat Using Continuous-Terrestrial and Airborne Measurements. Front. Plant Sci. 10. https://doi.org/10.3389/fpls.2019.00875

Deery, D.M., Rebetzke, G.J., Jimenez-Berni, J.A., James, R.A., Condon, A.G., Bovill, W.D., Hutchinson, P., Scarrow, J., Davy, R., Furbank, R.T., 2016. Methodology for High-Throughput Field Phenotyping of Canopy Temperature Using Airborne Thermography. Front. Plant Sci. 7. https://doi.org/10.3389/fpls.2016.01808

Delalieux, S., Somers, B., Hereijgers, S., Verstraeten, W.W., Keulemans, W., Coppin, P., 2008. A near-infrared narrow-waveband ratio to determine Leaf Area Index in orchards. Remote Sensing of Environment 112, 3762–3772. https://doi.org/10.1016/j.rse.2008.05.003

Delalieux, S., van Aardt, J., Keulemans, W., Schrevens, E., Coppin, P., 2007. Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications. European Journal of Agronomy 27, 130–143. https://doi.org/10.1016/j.eja.2007.02.005

Demetriades-Shah, T.H., Steven, M.D., Clark, J.A., 1990. High resolution derivative spectra in remote sensing. Remote Sensing of Environment 33, 55–64. https://doi.org/10.1016/0034-4257(90)90055-Q

Derkx, A.P., Orford, S., Griffiths, S., Foulkes, M.J., Hawkesford, M.J., 2012. Identification of differentially senescing mutants of wheat and impacts on yield, biomass and nitrogen partitioning. J Integr Plant Biol 54, 555–566. https://doi.org/10.1111/j.1744-7909.2012.01144.x

Devadas, R., Lamb, D.W., Backhouse, D., Simpfendorfer, S., 2015. Sequential application of hyperspectral indices for delineation of stripe rust infection and nitrogen deficiency in wheat. Precision Agric 16, 477–491. https://doi.org/10.1007/s11119-015-9390-0

Deventer, A.P. van (Geogroep I, Ward, A.D., Gowda, P.H., Lyon, J.G., 1997. Using thematic mapper data to identify contrasting soil plains and tillage practices. Photogrammetric engineering and remote sensing (USA).

Distelfeld, A., Avni, R., Fischer, A.M., 2014. Senescence, nutrient remobilization, and yield in wheat and barley. J. Exp. Bot. 65, 3783–3798. https://doi.org/10.1093/jxb/ert477

Eitel, J.U.H., Long, D.S., Gessler, P.E., Smith, A.M.S., 2007. Using in-situ measurements to evaluate the new RapidEye^{TM} satellite series for prediction of wheat nitrogen status. International Journal of Remote Sensing 28, 4183–4190. https://doi.org/10.1080/01431160701422213

Elsayed, S., Rischbeck, P., Schmidhalter, U., 2015. Comparing the performance of active and passive reflectance sensors to assess the normalized relative canopy temperature and grain yield of drought-stressed barley cultivars. Field Crops Research 177, 148–160. https://doi.org/10.1016/j.fcr.2015.03.010

Erdle, K., Mistele, B., Schmidhalter, U., 2013. Spectral high-throughput assessments of phenotypic differences in biomass and nitrogen partitioning during grain filling of wheat under high yielding Western European conditions. Field Crops Research 141, 16–26. https://doi.org/10.1016/j.fcr.2012.10.018

Fensholt, R., Sandholt, I., 2003. Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment. Remote Sensing of Environment 87, 111–121. https://doi.org/10.1016/j.rse.2003.07.002

Fernandez-Gallego, J.A., Buchaillot, M.L., Aparicio Gutiérrez, N., Nieto-Taladriz, M.T., Araus, J.L., Kefauver, S.C., 2019. Automatic Wheat Ear Counting Using Thermal Imagery. Remote Sensing 11, 751. https://doi.org/10.3390/rs11070751

Finger, R., 2010. Evidence of slowing yield growth – The example of Swiss cereal yields. Food Policy 35, 175–182. https://doi.org/10.1016/j.foodpol.2009.11.004

Fiorani, F., Schurr, U., 2013. Future Scenarios for Plant Phenotyping. Annual Review of Plant Biology 64, 267–291. https://doi.org/10.1146/annurev-arplant-050312-120137

Fischer, A., Feller, U., 1994. Senescence and protein degradation in leaf segments of young winter wheat: influence of leaf age. J Exp Bot 45, 103–109. https://doi.org/10.1093/jxb/45.1.103

Fischer, R.A., 2008. The importance of grain or kernel number in wheat: A reply to Sinclair and Jamieson. Field Crops Research 105, 15–21. https://doi.org/10.1016/j.fcr.2007.04.002

Fischer, R.A., Rees, D., Sayre, K.D., Lu, Z.-M., Condon, A.G., Saavedra, A.L., 1998. Wheat Yield Progress Associated with Higher Stomatal Conductance and Photosynthetic Rate, and Cooler Canopies. Crop Science 38, 1467. https://doi.org/10.2135/cropsci1998.0011183X003800060011x

Flisch, R., Sinaj, S., Charles, R., Richner, W., 2009. GRUDAF 2009. Principles for fertilisation in arable and fodder production. Agrarforschung Schweiz 1–100.

Fones, H., Gurr, S., 2015. The impact of Septoria tritici Blotch disease on wheat: An EU perspective. Fungal Genetics and Biology, Septoria tritici blotch disease of wheat: Tools and techniques to study the pathogen Zymoseptoria tritici 79, 3–7. https://doi.org/10.1016/j.fgb.2015.04.004

Foulkes, M.J., Sylvester-Bradley, R., Worland, A.J., Snape, J.W., 2004. Effects of a photoperiod-response gene Ppd-D1 on yield potential and drought resistance in UK winter wheat. Euphytica 135, 63–73. https://doi.org/10.1023/B:EUPH.0000009542.06773.13

Frels, K., Guttieri, M., Joyce, B., Leavitt, B., Baenziger, P.S., 2018. Evaluating canopy spectral reflectance vegetation indices to estimate nitrogen use traits in hard winter wheat. Field Crops Research 217, 82–92. https://doi.org/10.1016/j.fcr.2017.12.004

Friedli, M., Kirchgessner, N., Grieder, C., Liebisch, F., Mannale, M., Walter, A., 2016. Terrestrial 3D laser scanning to track the increase in canopy height of both monocot and dicot crop species under field conditions. Plant Methods 12, 9. https://doi.org/10.1186/s13007-016-0109-7

Fuentes, A., Yoon, S., Kim, S.C., Park, D.S., 2017. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. Sensors 17, 2022. https://doi.org/10.3390/s17092022

Fuentes, D.A., Gamon, J.A., Qiu, H., Sims, D.A., Roberts, D.A., 2001. Mapping Canadian boreal forest vegetation using pigment and water absorption features derived from the AVIRIS sensor. Journal of Geophysical Research: Atmospheres 106, 33565–33577. https://doi.org/10.1029/2001JD900110

Furbank, R.T., Tester, M., 2011. Phenomics – technologies to relieve the phenotyping bottleneck. Trends in Plant Science 16, 635–644. https://doi.org/10.1016/j.tplants.2011.09.005

Gaju, O., Allard, V., Martre, P., Le Gouis, J., Moreau, D., Bogard, M., Hubbart, S., Foulkes, M.J., 2014. Nitrogen partitioning and remobilization in relation to leaf senescence, grain yield and grain nitrogen concentration in wheat cultivars. Field Crops Research 155, 213–223. https://doi.org/10.1016/j.fcr.2013.09.003

Gaju, O., Allard, V., Martre, P., Snape, J.W., Heumez, E., LeGouis, J., Moreau, D., Bogard, M., Griffiths, S., Orford, S., Hubbart, S., Foulkes, M.J., 2011. Identification of traits to improve the nitrogen-use efficiency of wheat genotypes. Field Crops Research 123, 139–152. https://doi.org/10.1016/j.fcr.2011.05.010

Galvão, L.S., Formaggio, A.R., Tisot, D.A., 2005. Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. Remote Sensing of Environment 94, 523–534. https://doi.org/10.1016/j.rse.2004.11.012

Gamon, J.A., Peñuelas, J., Field, C.B., 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sensing of Environment 41, 35–44. https://doi.org/10.1016/0034-4257(92)90059-S

Gamon, J.A., Surfus, J.S., 1999. Assessing Leaf Pigment Content and Activity with a Reflectometer. The New Phytologist 143, 105–117.

Gao, B., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment 58, 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

García, G.A., Serrago, R.A., Dreccer, M.F., Miralles, D.J., 2016. Post-anthesis warm nights reduce grain weight in field-grown wheat and barley. Field Crops Research 195, 50–59. https://doi.org/10.1016/j.fcr.2016.06.002

Gitelson, A., Merzlyak, M.N., 1994. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. Journal of Plant Physiology 143, 286–292. https://doi.org/10.1016/S0176-1617(11)81633-0

Gitelson, A.A., 2004. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. Journal of Plant Physiology 161, 165–173. https://doi.org/10.1078/0176-1617-01176

Gitelson, A.A., Gritz †, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. Journal of Plant Physiology 160, 271–282. https://doi.org/10.1078/0176-1617-00887

Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002a. Novel algorithms for remote estimation of vegetation fraction. Remote Sensing of Environment 80, 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9

Gitelson, A.A., Keydan, G.P., Merzlyak, M.N., 2006. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophysical Research Letters 33. https://doi.org/10.1029/2006GL026457

Gitelson, A.A., Merzlyak, M.N., Chivkunova, O.B., 2001. Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. Photochemistry and

Photobiology 74, 38–45. https://doi.org/10.1562/0031-8655(2001)0740038OPANEO2.0.CO2

Gitelson, A.A., Viña, A., Ciganda, V., Rundquist, D.C., Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. Geophysical Research Letters 32. https://doi.org/10.1029/2005GL022688

Gitelson, A.A., Zur, Y., Chivkunova, O.B., Merzlyak, M.N., 2002b. Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. Photochemistry and Photobiology 75, 272–281. https://doi.org/10.1562/0031-8655(2002)0750272ACCIPL2.0.CO2

Giunta, F., Motzo, R., Pruneddu, G., 2008. Has long-term selection for yield in durum wheat also induced changes in leaf and canopy traits? Field Crops Research 106, 68–76. https://doi.org/10.1016/j.fcr.2007.10.018

Gong, Y.-H., Zhang, J., Gao, J.-F., Lu, J.-Y., Wang, J.-R., 2005a. Slow Export of Photoassimilate from Stay-green Leaves during Late Grain-Filling Stage in Hybrid Winter Wheat (Triticum aestivum L.). Journal of Agronomy and Crop Science 191, 292–299. https://doi.org/10.1111/j.1439-037X.2005.00173.x

Gooding, M.J., Dimmock, J.P.R.E., France, J., Jones, S.A., 2000. Green leaf area decline of wheat flag leaves: the influence of fungicides and relationships with mean grain weight and grain yield. Annals of Applied Biology 136, 77–84. https://doi.org/10.1111/j.1744-7348.2000.tb00011.x

Granitto, P.M., Furlanello, C., Biasioli, F., Gasperi, F., 2006. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemometrics and Intelligent Laboratory Systems 83, 83–90. https://doi.org/10.1016/j.chemolab.2006.01.007

Gregersen, P.L., Culetic, A., Boschian, L., Krupinska, K., 2013. Plant senescence and crop productivity. Plant Mol Biol 82, 603–622. https://doi.org/10.1007/s11103-013-0013-8

Gregersen, P.L., Holm, P.B., Krupinska, K., 2008. Leaf senescence and nutrient remobilisation in barley and wheat. Plant Biology 10, 37–49. https://doi.org/10.1111/j.1438-8677.2008.00114.x

Grieder, C., Hund, A., Walter, A., 2015. Image based phenotyping during winter: a powerful tool to assess wheat genetic variation in growth response to temperature. Functional Plant Biol. 42, 387–396. https://doi.org/10.1071/FP14226

Gu, Y., Wylie, B.K., Howard, D.M., Phuyal, K.P., Ji, L., 2013. NDVI saturation adjustment: A new approach for improving cropland performance estimates in the Greater Platte River Basin, USA. Ecological Indicators 30, 1–6. https://doi.org/10.1016/j.ecolind.2013.01.041

Guo, Y., Gan, S.-S., 2012. Convergence and divergence in gene expression profiles induced by leaf senescence and 27 senescence-promoting hormonal, pathological and environmental stress treatments. Plant, Cell & Environment 35, 644–655. https://doi.org/10.1111/j.1365-3040.2011.02442.x

Gutierrez, M., Reynolds, M.P., Klatt, A.R., 2015. Effect of leaf and spike morphological traits on the relationship between spectral reflectance indices and yield in wheat. International Journal of Remote Sensing 36, 701–718. https://doi.org/10.1080/01431161.2014.999878

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning 46, 389–422. https://doi.org/10.1023/A:1012487302797

Guyot, G., Baret, F., Major, D.J., 1988. High Spectral Resolution: Determination of spectral shifts between the red and near infrared. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 750–760.

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. Remote Sensing of Environment 90, 337–352. https://doi.org/10.1016/j.rse.2003.12.013

Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sensing of Environment 81, 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4

Hansen, P.M., Schjoerring, J.K., 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote Sensing of Environment 86, 542–553. https://doi.org/10.1016/S0034-4257(03)00131-7

Hardisky, M.A., Daiber, F.C., Roman, C.T., Klemas, V., 1984. Remote sensing of biomass and annual net aerial primary productivity of a salt marsh. Remote Sensing of Environment 16, 91–106. https://doi.org/10.1016/0034-4257(84)90055-5

Hassan, M.A., Yang, M., Rasheed, A., Jin, X., Xia, X., Xiao, Y., He, Z., 2018. Time-Series Multispectral Indices from Unmanned Aerial Vehicle Imagery Reveal Senescence Rate in Bread Wheat. Remote Sensing 10, 809. https://doi.org/10.3390/rs10060809

Hastie, H.Z. and T., 2018. elasticnet: Elastic-Net for Sparse Estimation and Sparse PCA.

Hategekimana, A., Schneider, D., Fossati, D., Mascher, F., 2012. Performance and nitrogen efficiency of Swiss Wheat varieties from the 20th century. Agrarforschung Schweiz 3, 44–51.

Havé, M., Marmagne, A., Chardon, F., Masclaux-Daubresse, C., 2017. Nitrogen remobilization during leaf senescence: lessons from Arabidopsis to crops. J Exp Bot 68, 2513–2529. https://doi.org/10.1093/jxb/erw365

Hirel, B., Le Gouis, J., Ney, B., Gallais, A., 2007. The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Bot 58, 2369–2387. https://doi.org/10.1093/jxb/erm097

Holzkämper, A., Fossati, D., Hiltbrunner, J., Fuhrer, J., 2015. Spatial and temporal trends in agro-climatic limitations to production potentials for grain maize and winter wheat in

Switzerland. Reg Environ Change 15, 109–122. https://doi.org/10.1007/s10113-014-0627-7

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment 25, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Hund, A., Kronenberg, L., Anderegg, J., Yu, K., KU Leuven, Belgium, Walter, A., Yu, K., KU Leuven, Belgium, Walter, A., ETH Zurich, Switzerland, 2019. Non-invasive field phenotyping of cereal development, in: Ordon, F. (Ed.), Burleigh Dodds Series in Agricultural Science. Burleigh Dodds Science Publishing, pp. 249–292. https://doi.org/10.19103/AS.2019.0051.13

Hunt, E.R.J., Daughtry, C.S.T., Eitel, J.U.H., Long, D.S., 2011. Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index. Agronomy journal.

Isik, F., Holland, J., Maltecca, C., 2017. Genetic Data Analysis for Plant and Animal Breeding. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-55177-7

Jackson, P., Robertson, M., Cooper, M., Hammer, G., 1996. The role of physiological understanding in plant breeding; from a breeding perspective. Field Crops Research 49, 11–37. https://doi.org/10.1016/S0378-4290(96)01012-X

Jackson, R.D., Idso, S.B., Reginato, R.J., Pinter, P.J., Jr., 1981. Canopy temperature as a crop water stress indicator. Water Resources Research 17, 1133–1138. https://doi.org/10.1029/WR017i004p01133

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, Imaging Spectroscopy Special Issue 113, S56–S66. https://doi.org/10.1016/j.rse.2008.01.026

Jagadish, K.S.V., Kishor, K., B, P., Bahuguna, R.N., von Wirén, N., Sreenivasulu, N., 2015. Staying Alive or Going to Die During Terminal Senescence—An Enigma Surrounding Yield Stability. Front. Plant Sci. 6. https://doi.org/10.3389/fpls.2015.01070

Jiang, G.H., He, Y.Q., Xu, C.G., Li, X.H., Zhang, Q., 2004. The genetic basis of stay-green in rice analyzed in a population of doubled haploid lines derived from an indica by japonica cross. Theor Appl Genet 108, 688–698. https://doi.org/10.1007/s00122-003-1465-z

Jimenez-Berni, J.A., Deery, D.M., Rozas-Larraondo, P., Condon, A. (Tony) G., Rebetzke, G.J., James, R.A., Bovill, W.D., Furbank, R.T., Sirault, X.R.R., 2018. High Throughput Determination of Plant Height, Ground Cover, and Above-Ground Biomass in Wheat with LiDAR. Front. Plant Sci. 9. https://doi.org/10.3389/fpls.2018.00237

Joalland, S., Screpanti, C., Varella, H.V., Reuther, M., Schwind, M., Lang, C., Walter, A., Liebisch, F., 2018. Aerial and Ground Based Sensing of Tolerance to Beet Cyst Nematode in Sugar Beet. Remote Sensing 10, 787. https://doi.org/10.3390/rs10050787

Jordan, D.R., Hunt, C.H., Cruickshank, A.W., Borrell, A.K., Henzell, R.G., 2012. The Relationship Between the Stay-Green Trait and Grain Yield in Elite Sorghum Hybrids Grown in a Range of Environments. Crop Science 52, 1153–1161. https://doi.org/10.2135/cropsci2011.06.0326

Karisto, P., Hund, A., Yu, K., Anderegg, J., Walter, A., Mascher, F., McDonald, B.A., Mikaberidze, A., 2018. Ranking Quantitative Resistance to Septoria tritici Blotch in Elite Wheat Cultivars Using Automated Image Analysis. Phytopathology 108, 568–581. https://doi.org/10.1094/PHYTO-04-17-0163-R

Kaufman, Y.J., Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE Transactions on Geoscience and Remote Sensing 30, 261–270. https://doi.org/10.1109/36.134076

Kichey, T., Hirel, B., Heumez, E., Dubois, F., Le Gouis, J., 2007. In winter wheat (Triticum aestivum L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. Field Crops Research 102, 22–32. https://doi.org/10.1016/j.fcr.2007.01.002

Kipp, S., Mistele, B., Schmidhalter, U., 2014. Identification of stay-green and early senescence phenotypes in high-yielding winter wheat, and their relationship to grain yield and grain protein concentration using high-throughput phenotyping techniques. Funct. Plant Biol. 41, 227–235.

Kirchgessner, N., Liebisch, F., Yu, K., Pfeifer, J., Friedli, M., Hund, A., Walter, A., 2017. The ETH field phenotyping platform FIP: a cable-suspended multi-sensor system. Functional Plant Biol. 44, 154–168. https://doi.org/10.1071/FP16165

Kollers, S., Rodemann, B., Ling, J., Korzun, V., Ebmeyer, E., Argillier, O., Hinze, M., Plieske, J., Kulosa, D., Ganal, M.W., Röder, M.S., 2013. Whole Genome Association Mapping of Fusarium Head Blight Resistance in European Winter Wheat (Triticum aestivum L.). PLOS ONE 8, e57500. https://doi.org/10.1371/journal.pone.0057500

Kong, L., Xie, Y., Hu, L., Feng, B., Li, S., 2016. Remobilization of vegetative nitrogen to developing grain in wheat (Triticum aestivum L.). Field Crops Research 196, 134–144. https://doi.org/10.1016/j.fcr.2016.06.015

Krishnan, P., Meile, L., Plissonneau, C., Ma, X., Hartmann, F.E., Croll, D., McDonald, B.A., Sánchez-Vallet, A., 2018. Transposable element insertions shape gene regulation and melanin production in a fungal pathogen of wheat. BMC Biology 16, 78. https://doi.org/10.1186/s12915-018-0543-2

Kronenberg, L., Yates, S., Boer, M.P., Kirchgessner, N., Walter, A., Hund, A., 2019. Temperature response of wheat affects final height and the timing of key developmental stages under field conditions. bioRxiv 756700. https://doi.org/10.1101/756700

Kronenberg, L., Yu, K., Walter, A., Hund, A., 2017. Monitoring the dynamics of wheat stem elongation: genotypes differ at critical stages. Euphytica 213, 157. https://doi.org/10.1007/s10681-017-1940-2

Kuhn, M., 2008. Building Predictive Models in *R* Using the **caret** Package. Journal of Statistical Software 28. https://doi.org/10.18637/jss.v028.i05

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-6849-3

Kuhn, M., Weston, S., Keefer, C., Coulter, N., code), R.Q. (Author of imported C., code), R.R.P.L. (Copyright holder of imported C., 2018. Cubist: Rule- And Instance-Based Regression Modeling.

Kumar, R., Bishop, E., Bridges, W.C., Tharayil, N., Sekhon, R.S., 2019. Sugar partitioning and source–sink interaction are key determinants of leaf senescence in maize. Plant, Cell & Environment 42, 2597–2611. https://doi.org/10.1111/pce.13599

Kumari, M., Singh, V.P., Tripathi, R., Joshi, A.K., 2007. Variation for Staygreen Trait and its Association with Canopy Temperature Depression and Yield Traits Under Terminal Heat Stress in Wheat, in: Wheat Production in Stressed Environments. pp. 357–363.

Lancashire, P.D., Bleiholder, H., Boom, T.V.D., Langelüddeke, P., Stauss, R., Weber, E., Witzenberger, A., 1991. A uniform decimal code for growth stages of crops and weeds. Annals of Applied Biology 119, 561–601. https://doi.org/10.1111/j.1744-7348.1991.tb04895.x

Lee, E.A., Tollenaar, M., 2007. Physiological Basis of Successful Breeding Strategies for Maize Grain Yield. Crop Science 47, S-202-S-215. https://doi.org/10.2135/cropsci2007.04.0010IPBS

Lehner, B., Döll, P., Alcamo, J., Henrichs, T., Kaspar, F., 2006. Estimating the Impact of Global Change on Flood and Drought Risks in Europe: A Continental, Integrated Analysis. Climatic Change 75, 273–299. https://doi.org/10.1007/s10584-006-6338-4

Lendenmann, M.H., Croll, D., Stewart, E.L., McDonald, B.A., 2014. Quantitative Trait Locus Mapping of Melanization in the Plant Pathogenic Fungus Zymoseptoria tritici. G3: Genes, Genomes, Genetics 4, 2519–2533. https://doi.org/10.1534/g3.114.015289

Levi, L., Juan Manuel, H., Rechsteiner, S., Nuna, C., Scheuner, S., Weisflog, T., Brabant, C., Foiada, F., Städeli, C., Hund, A., 2017. Alternatives to secure wheat quality with lower supplies of nitrogen fertilizer. Cereal Technology 3, 206–216.

Li, F., Mistele, B., Hu, Y., Chen, X., Schmidhalter, U., 2014. Reflectance estimation of canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices and partial least squares regression. European Journal of Agronomy 52, 198–209. https://doi.org/10.1016/j.eja.2013.09.006

Li, X., Ingvordsen, C.H., Weiss, M., Rebetzke, G.J., Condon, A.G., James, R.A., Richards, R.A., 2019. Deeper roots associated with cooler canopies, higher normalized difference vegetation index, and greater yield in three wheat populations grown on stored soil water. J Exp Bot 70, 4963–4974. https://doi.org/10.1093/jxb/erz232

Li, X., Zhang, Y., Bao, Y., Luo, J., Jin, X., Xu, X., Song, X., Yang, G., 2014. Exploring the Best Hyperspectral Features for LAI Estimation Using Partial Least Squares Regression. Remote Sensing 6, 6221–6241. https://doi.org/10.3390/rs6076221

Liebisch, F., Kirchgessner, N., Schneider, D., Walter, A., Hund, A., 2015. Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. Plant Methods 11, 9. https://doi.org/10.1186/s13007-015-0048-8

Lim, P.O., Kim, H.J., Nam, H.G., 2007. Leaf senescence. Annu Rev Plant Biol 58, 115–136. https://doi.org/10.1146/annurev.arplant.57.032905.105316

Liu, T., Wu, W., Chen, W., Sun, C., Zhu, X., Guo, W., 2016. Automated image-processing for counting seedlings in a wheat field. Precision Agric 17, 392–406. https://doi.org/10.1007/s11119-015-9425-6

Lizana, X.C., Calderini, D.F., 2013. Yield and grain quality of wheat in response to increased temperatures at key periods for grain number and grain weight determination: Considerations for the climatic change scenarios of Chile. Journal of Agricultural Science 151, 209–221. https://doi.org/10.1017/S0021859612000639

Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate Trends and Global Crop Production Since 1980. Science 333, 616–620. https://doi.org/10.1126/science.1204531

Lopes, M.S., Reynolds, M.P., 2012. Stay-green in spring wheat can be determined by spectral reflectance measurements (normalized difference vegetation index) independently from phenology. J. Exp. Bot. 63, 3789–3798. https://doi.org/10.1093/jxb/ers071

Lopes, M.S., Reynolds, M.P., 2010. Partitioning of assimilates to deeper roots is associated with cooler canopies and increased yield under drought in wheat. Functional Plant Biol. 37, 147–156. https://doi.org/10.1071/FP09121

Louhaichi, M., Borman, M.M., Johnson, D.E., 2001. Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. Geocarto International 16, 65–70. https://doi.org/10.1080/10106040108542184

Lu, C., Lu, Q., Zhang, J., Kuang, T., 2001. Characterization of photosynthetic pigment composition, photosystem II photochemistry and thermal energy dissipation during leaf senescence of wheat plants grown in the field. J Exp Bot 52, 1805–1810. https://doi.org/10.1093/jexbot/52.362.1805

Mahlein, A.-K., 2016. Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. Plant Disease 100, 241–251. https://doi.org/10.1094/PDIS-03-15-0340-FE

Mahlein, A.-K., Oerke, E.-C., Steiner, U., Dehne, H.-W., 2012. Recent advances in sensing plant diseases for precision crop protection. Eur J Plant Pathol 133, 197–209. https://doi.org/10.1007/s10658-011-9878-z

Mahlein, A.-K., Rumpf, T., Welke, P., Dehne, H.-W., Plümer, L., Steiner, U., Oerke, E.-C., 2013. Development of spectral indices for detecting and identifying plant diseases. Remote Sensing of Environment 128, 21–30. https://doi.org/10.1016/j.rse.2012.09.019

Mahlein, A.-K., Steiner, U., Dehne, H.-W., Oerke, E.-C., 2010. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. Precision Agric 11, 413–431. https://doi.org/10.1007/s11119-010-9180-7

Makanza, R., Zaman-Allah, M., Cairns, J.E., Magorokosho, C., Tarekegne, A., Olsen, M., Prasanna, B.M., 2018. High-Throughput Phenotyping of Canopy Cover and Senescence in Maize Field Trials Using Aerial Digital Canopy Imaging. Remote Sensing 10, 330. https://doi.org/10.3390/rs10020330

Martre, P., Jamieson, P.D., Semenov, M.A., Zyskowski, R.F., Porter, J.R., Triboi, E., 2006. Modelling protein content and composition in relation to crop nitrogen dynamics for wheat. European Journal of Agronomy, Modelling Quality Traits and Their Genetic Variability for Wheat 25, 138–154. https://doi.org/10.1016/j.eja.2006.04.007

Martre, P., Quilot-Turion, B., Luquet, D., Memmah, M.-M.O.-S., Chenu, K., Debaeke, P., 2015. Chapter 14 - Model-assisted phenotyping and ideotype design, in: Sadras, V.O., Calderini, D.F. (Eds.), Crop Physiology (Second Edition). Academic Press, San Diego, pp. 349–373. https://doi.org/10.1016/B978-0-12-417104-6.00014-5

Masclaux, C., Quilleré, I., Gallais, A., Hirel, B., 2001. The challenge of remobilisation in plant nitrogen economy. A survey of physio-agronomic and molecular approaches. Annals of Applied Biology 138, 69–81. https://doi.org/10.1111/j.1744-7348.2001.tb00086.x

Mason, R.E., Singh, R.P., 2014. Considerations When Deploying Canopy Temperature to Select High Yielding Wheat Breeding Lines under Drought and Heat Stress. Agronomy 4, 191–201. https://doi.org/10.3390/agronomy4020191

McDonald, B.A., Linde, C., 2002. The population genetics of plant pathogens and breeding strategies for durable resistance. Euphytica 124, 163–180. https://doi.org/10.1023/A:1015678432355

McDonald, B.A., Mundt, C.C., 2016. How Knowledge of Pathogen Population Biology Informs Management of Septoria Tritici Blotch.

Meile, L., Croll, D., Brunner, P.C., Plissonneau, C., Hartmann, F.E., McDonald, B.A., Sánchez-Vallet, A., 2018. A fungal avirulence factor encoded in a highly plastic genomic region triggers partial resistance to septoria tritici blotch. New Phytologist 219, 1048–1061. https://doi.org/10.1111/nph.15180

Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., Rakitin, V.Y., 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiologia Plantarum 106, 135–141. https://doi.org/10.1034/j.1399-3054.1999.106119.x

MeteoSwiss, 2019. Climate Normals Zürich / Kloten. Reference period 1981–2010. Federal Department of Home Affairs FDHA, Federal Office of Meteorology and Climatology, MeteoSwiss. Available online at: http://www.meteoswiss.admin.ch.

MeteoSwiss, 2018. Hitze und Trockenheit im Sommerhalbjahr 2018 - eine klimatologische Übersicht (Fachbericht MeteoSchweiz No. 272).

Mevik, B.-H., Wehrens, R., Liland, K.H., Hiemstra, P., 2018. pls: Partial Least Squares and Principal Component Regression.

Millet, E., Pinthus, M.J., 1984. The association between grain volume and grain weight in wheat. Journal of Cereal Science 2, 31–35. https://doi.org/10.1016/S0733-5210(84)80005-3

Millet, E.J., Kruijer, W., Coupel-Ledru, A., Prado, S.A., Cabrera-Bosquet, L., Lacube, S., Charcosset, A., Welcker, C., Eeuwijk, F. van, Tardieu, F., 2019. Genomic prediction of maize yield across European environmental conditions. Nat Genet 51, 952–956. https://doi.org/10.1038/s41588-019-0414-y

Miralles, D.J., Slafer, G.A., 2007. Sink limitations to yield in wheat: how couldit be reduced? Journal of Agricultural Science 145, 139–149.

Miralles, D.J., Slafer, G.A., 1996. Grain Weight Reductions in Wheat Associated with Semidwarfism: an Analysis of Grain Weight at Different Positions Within the Spike of Near-isogenic Lines. Journal of Agronomy and Crop Science 177, 9–16. https://doi.org/10.1111/j.1439-037X.1996.tb00586.x

Mistele, B., Schmidhalter, U., 2010. Tractor-Based Quadrilateral Spectral Reflectance Measurements to Detect Biomass and Total Aerial Nitrogen in Winter Wheat. Agronomy Journal 102, 499–506. https://doi.org/10.2134/agronj2009.0282

Mistele, B., Schmidhalter, U., n.d. A COMPARISON OF SPECTRAL REFLECTANCE AND LASER- INDUCED CHOLORPHYLL FLUORESCENCE MEASUREMENTS TO DETECT DIFFERENCES IN AERIAL DRY WEIGHT AND NITROGEN UPDATE OF WHEAT 14.

Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using Deep Learning for Image-Based Plant Disease Detection. Front. Plant Sci. 7. https://doi.org/10.3389/fpls.2016.01419

Montazeaud, G., Karatoğma, H., Özturk, I., Roumet, P., Ecarnot, M., Crossa, J., Özer, E., Özdemir, F., Lopes, M.S., 2016. Predicting wheat maturity and stay–green parameters by modeling spectral reflectance measurements and their contribution to grain yield under rainfed conditions. Field Crops Research 196, 191–198. https://doi.org/10.1016/j.fcr.2016.06.021

Moreau, D., Allard, V., Martre, P., Le Gouis, J., Edwards, S., Foulkes, J., 2008. Genotypic variability associated with the dynamics of leaf senescence in wheat canopies.

Munier-Jolain, N.G., Salon, C., 2005. Are the carbon costs of seed production related to the quantitative and qualitative performance? An appraisal for legumes and other crops. Plant, Cell & Environment 28, 1388–1395. https://doi.org/10.1111/j.1365-3040.2005.01371.x

Nagler, P.L., Daughtry, C.S.T., Goward, S.N., 2000. Plant Litter and Soil Reflectance. Remote Sensing of Environment 71, 207–215. https://doi.org/10.1016/S0034-4257(99)00082-6

Naruoka, Y., Sherman, J.D., Lanning, S.P., Blake, N.K., Martin, J.M., Talbert, L.E., 2012. Genetic Analysis of Green Leaf Duration in Spring Wheat. Crop Science 52, 99. https://doi.org/10.2135/cropsci2011.05.0269

Odilbekov, F., Armonienė, R., Henriksson, T., Chawade, A., 2018. Proximal Phenotyping and Machine Learning Methods to Identify Septoria Tritici Blotch Disease Symptoms in Wheat. Front. Plant Sci. 9. https://doi.org/10.3389/fpls.2018.00685

O'Driscoll, A., Kildea, S., Doohan, F., Spink, J., Mullins, E., 2014. The wheat–Septoria conflict: a new front opening up? Trends in Plant Science 19, 602–610. https://doi.org/10.1016/j.tplants.2014.04.011

Olivares-Villegas, J.J., Reynolds, M.P., McDonald, G.K., 2007. Drought-adaptive attributes in the Seri/Babax hexaploid wheat population. Functional Plant Biol. 34, 189–203. https://doi.org/10.1071/FP06148

Orton, E.S., Deller, S., Brown, J.K.M., 2011. Mycosphaerella graminicola: from genomics to disease control. Molecular Plant Pathology 12, 413–424. https://doi.org/10.1111/j.1364-3703.2010.00688.x

Oury, F.-X., Godin, C., Mailliard, A., Chassin, A., Gardet, O., Giraud, A., Heumez, E., Morlais, J.-Y., Rolland, B., Rousset, M., Trottet, M., Charmet, G., 2012. A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France. European Journal of Agronomy 40, 28–38. https://doi.org/10.1016/j.eja.2012.02.007

Øvergaard, S.I., Isaksson, T., Korsaeth, A., 2013a. Prediction of Wheat Yield and Protein Using Remote Sensors on Plots—Part II: Improving Prediction Ability Using Data Fusion. Journal of Near Infrared Spectroscopy. https://doi.org/10.1255/jnirs.1044

Øvergaard, S.I., Isaksson, T., Korsaeth, A., 2013b. Prediction of Wheat Yield and Protein Using Remote Sensors on Plots—Part I: Assessing near Infrared Model Robustness for Year and Site Variations. Journal of Near Infrared Spectroscopy. https://doi.org/10.1255/jnirs.1042

Padfield, D., Matheson, G., 2018. nls.multstart: Robust Non-Linear Regression using AIC Scores.

Pask, A., Pietragalla, J., Mullan, D., Reynolds, M.P., 2012. Physiological breeding II : a field guide to wheat phenotyping iv, 132 pages.

Pauli, D., Chapman, S.C., Bart, R., Topp, C.N., Lawrence-Dill, C.J., Poland, J., Gore, M.A., 2016. The Quest for Understanding Phenotypic Variation via Integrated Approaches in the Field Environment. Plant Physiology 172, 622–634. https://doi.org/10.1104/pp.16.00592

Pavuluri, K., Chim, B.K., Griffey, C.A., Reiter, M.S., Balota, M., Thomason, W.E., 2015. Canopy spectral reflectance can predict grain nitrogen use efficiency in soft red winter wheat. Precision Agric 16, 405–424. https://doi.org/10.1007/s11119-014-9385-2

Penuelas, J., Baret, F., Filella, I., 1995. Semiempirical Indexes to Assess Carotenoids Chlorophyll-a Ratio from Leaf Spectral Reflectance. Photosynthetica 31, 221–230.

Peñuelas, J., Filella, I., Biel, C., Serrano, L., Savé, R., 1993. The reflectance at the 950–970 nm region as an indicator of plant water status. International Journal of Remote Sensing 14, 1887–1905. https://doi.org/10.1080/01431169308954010

Peñuelas, J., Gamon, J.A., Fredeen, A.L., Merino, J., Field, C.B., 1994. Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. Remote Sensing of Environment 48, 135–146. https://doi.org/10.1016/0034-4257(94)90136-8

Peñuelas, J., Inoue, Y., 1999. Reflectance Indices Indicative of Changes in Water and Pigment Contents of Peanut and Wheat Leaves. Photosynthetica 36, 355–360. https://doi.org/10.1023/A:1007033503276

Penuelas, J., Pinol, J., Ogaya, R., Filella, I., 1997. Estimation of plant water concentration by the reflectance Water Index WI (R900/R970). International Journal of Remote Sensing 18, 2869–2875. https://doi.org/10.1080/014311697217396

Pérez-Gianmarco, T.I., Slafer, G.A., González, F.G., 2018. Wheat pre-anthesis development as affected by photoperiod sensitivity genes (Ppd-1) under contrasting photoperiods. Functional Plant Biol. 45, 645–657. https://doi.org/10.1071/FP17195

Perich, G., Hund, A., Anderegg, J., Roth, L., Boer, M.P., Walter, A., Liebisch, F., Aasen, H., 2020. Assessment of multi-image UAV based high-throughput field phenotyping of canopy temperature. Front. Plant Sci. 11. https://doi.org/10.3389/fpls.2020.00150

Pimstein, A., Eitel, J.U.H., Long, D.S., Mufradi, I., Karnieli, A., Bonfil, D.J., 2009. A spectral index to monitor the head-emergence of wheat in semi-arid conditions. Field Crops Research 111, 218–225. https://doi.org/10.1016/j.fcr.2008.12.009

Pinto, R.S., Reynolds, M.P., Mathews, K.L., McIntyre, C.L., Olivares-Villegas, J.-J., Chapman, S.C., 2010. Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. Theor Appl Genet 121, 1001–1021. https://doi.org/10.1007/s00122-010-1351-4

Potgieter, A.B., George-Jaeggli, B., Chapman, S.C., Laws, K., Cadavid, S., A, L., Wixted, J., Watson, J., Eldridge, M., Jordan, D.R., Hammer, G.L., 2017. Multi-Spectral Imaging from an Unmanned Aerial Vehicle Enables the Assessment of Seasonal Leaf Area Dynamics of Sorghum Breeding Lines. Front. Plant Sci. 8. https://doi.org/10.3389/fpls.2017.01532

Prashar, A., Jones, H., 2014. Infra-Red Thermography as a High-Throughput Tool for Field Phenotyping. Agronomy 4, 397–417. https://doi.org/10.3390/agronomy4030397

Prey, L., Hu, Y., Schmidhalter, U., 2020. High-Throughput Field Phenotyping Traits of Grain Yield Formation and Nitrogen Use Efficiency: Optimizing the Selection of Vegetation Indices and Growth Stages. Front. Plant Sci. 10. https://doi.org/10.3389/fpls.2019.01672

Prey, L., Schmidhalter, U., 2019a. Temporal and Spectral Optimization of Vegetation Indices for Estimating Grain Nitrogen Uptake and Late-Seasonal Nitrogen Traits in Wheat. Sensors 19, 4640. https://doi.org/10.3390/s19214640

Prey, L., Schmidhalter, U., 2019b. Sensitivity of Vegetation Indices for Estimating Vegetative N Status in Winter Wheat. Sensors 19, 3712. https://doi.org/10.3390/s19173712

Pu, R., 2011. Detecting and Mapping Invasive Plant Species by Using Hyperspectral Data 447–466. https://doi.org/10.1201/b11222-27

Pu, R., Ge, S., Kelly, N.M., Gong, P., 2003. Spectral absorption features as indicators of water status in coast live oak (Quercus agrifolia) leaves. International Journal of Remote Sensing 24, 1799–1810. https://doi.org/10.1080/01431160210155965

QGIS Development Team, 2019. QGIS Geographic Information System. Open Source Geospatial Foundation Project.

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sensing of Environment 48, 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Qi, J., Marsett, R., Heilman, P., Bieden-bender, S., Moran, S., Goodrich, D., Weltz, M., 2002. RANGES improves satellite-based information and land cover assessments in southwest

United States. Eos, Transactions American Geophysical Union 83, 601–606. https://doi.org/10.1029/2002EO000411

Quinlan, 1992. Learning with continuous classes. Proceedings of Australian Joint Conference on Artificial Intelligence 343–348.

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Vienna, Austria.

Rajcan, I., Tollenaar, M., 1999a. Source : sink ratio and leaf senescence in maize: II. Nitrogen metabolism during grain filling. Field Crops Research 60, 255–265. https://doi.org/10.1016/S0378-4290(98)00143-9

Rajcan, I., Tollenaar, M., 1999b. Source : sink ratio and leaf senescence in maize: I. Dry matter accumulation and partitioning during grain filling. Field Crops Research 60, 245–253. https://doi.org/10.1016/S0378-4290(98)00142-7

Ramirez-Lopez, L., Stevens, A., 2014. prospectr: Miscellaneous functions for processing and sample selection of vis-NIR diffuse reflectance data.

Rao, N.R., Garg, P.K., Ghosh, S.K., Dadhwal, V.K., 2008. Estimation of leaf total chlorophyll and nitrogen concentrations using hyperspectral satellite imagery. The Journal of Agricultural Science 146, 65–75. https://doi.org/10.1017/S0021859607007514

Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A., 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLOS ONE 8, e66428. https://doi.org/10.1371/journal.pone.0066428

Rebetzke, G.J., Jimenez-Berni, J.A., Bovill, W.D., Deery, D.M., James, R.A., 2016. High-throughput phenotyping technologies allow accurate selection of stay-green. J Exp Bot 67, 4919–4924. https://doi.org/10.1093/jxb/erw301

Rebetzke, G.J., Rattey, A.R., Farquhar, G.D., Richards, R.A., Condon, A. (Tony) G., 2013. Genomic regions for canopy temperature and their genetic association with stomatal conductance and grain yield in wheat. Functional Plant Biol. 40, 14–33. https://doi.org/10.1071/FP12184

Reyniers, M., Walvoort, D.J.J., Baardemaaker, J.D., 2006. A linear model to predict with a multi-spectral radiometer the amount of nitrogen in winter wheat. International Journal of Remote Sensing 27, 4159–4179. https://doi.org/10.1080/01431160600791650

Reynolds, M., Foulkes, J., Furbank, R., Griffiths, S., King, J., Murchie, E., Parry, M., Slafer, G., 2012. Achieving yield gains in wheat. Plant, Cell & Environment 35, 1799–1823. https://doi.org/10.1111/j.1365-3040.2012.02588.x

Reynolds, M., Langridge, P., 2016. Physiological breeding. Current Opinion in Plant Biology, SI: 31: Physiology and metabolism 2016 31, 162–171. https://doi.org/10.1016/j.pbi.2016.04.005

Reynolds, M., Manes, Y., Izanloo, A., Langridge, P., 2009. Phenotyping approaches for physiological breeding and gene discovery in wheat. Annals of Applied Biology 155, 309–320. https://doi.org/10.1111/j.1744-7348.2009.00351.x

Reynolds, M.P., Balota, M., Delgado, M.I.B., Amani, I., Fischer, R.A., 1994. Physiological and Morphological Traits Associated With Spring Wheat Yield Under Hot, Irrigated Conditions. Functional Plant Biol. 21, 717–730. https://doi.org/10.1071/pp9940717

Rezaei, E.E., Siebert, S., Hüging, H., Ewert, F., 2018. Climate change effect on wheat phenology depends on cultivar change. Sci Rep 8, 1–10. https://doi.org/10.1038/s41598-018-23101-2

Richards, R.A., 1996. Increasing yield potential in wheat - source and sink limitations. in: Increasing Yield Potential in Wheat: Breaking the Barriers.

Rodriguez-Alvarez, M.X., Lee, D.-J., Kneib, T., Durban, M., Eilers, P., 2019. SAP: Multidimensional Generalized P-Splines Regression Models Estimation.

Rohart, F., Gautier, B., Singh, A., Cao, K.-A.L., 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Computational Biology 13, e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sensing of Environment 55, 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Rouse, J.W., Jr., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring Vegetation Systems in the Great Plains with Erts. NASA Special Publication 351, 309.

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L.G., Crossa, J., Reynolds, M., Singh, R., 2016. Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. G3: Genes, Genomes, Genetics 6, 2799–2808. https://doi.org/10.1534/g3.116.032888

Sadeghi-Tehran, P., Sabermanesh, K., Virlet, N., Hawkesford, M.J., 2017a. Automated Method to Determine Two Critical Growth Stages of Wheat: Heading and Flowering. Front. Plant Sci. 8. https://doi.org/10.3389/fpls.2017.00252

Sadeghi-Tehran, P., Virlet, N., Ampe, E.M., Reyns, P., Hawkesford, M.J., 2019. DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. Front. Plant Sci. 10. https://doi.org/10.3389/fpls.2019.01176

Sadeghi-Tehran, P., Virlet, N., Sabermanesh, K., Hawkesford, M.J., 2017b. Multi-feature machine learning model for automatic segmentation of green fractional vegetation cover for high-throughput field phenotyping. Plant Methods 13, 103. https://doi.org/10.1186/s13007-017-0253-8

Sadras, V.O., Mahadevan, M., Zwer, P.K., 2019. Stay-green associates with low water soluble carbohydrates at flowering in oat. Field Crops Research 230, 132–138. https://doi.org/10.1016/j.fcr.2018.10.007

Sanger, J.E., 1971. Quantitative Investigations of Leaf Pigments From Their Inception in Buds Through Autumn Coloration to Decomposition in Falling Leaves. Ecology 52, 1075–1089. https://doi.org/10.2307/1933816

Savin, R., Slafer, G.A., 1991. Shading effects on the yield of an Argentinian wheat cultivar. The Journal of Agricultural Science 116, 1–7. https://doi.org/10.1017/S0021859600076085

Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry 36, 1627–1639. https://doi.org/10.1021/ac60214a047

Schlemmer, M.R., Francis, D.D., Shanahan, J.F., Schepers, J.S., 2005. Remotely Measuring Chlorophyll Content in Corn Leaves with Differing Nitrogen Levels and Relative Water Content. Agronomy Journal 97, 106–112. https://doi.org/10.2134/agronj2005.0106

Schürch, C., Kronenberg, L. and Hund, A. (2018). Wheat developmental stages. Available at: https://commons.wikimedia.org/wiki/File:Wheat_developmental_stages.tif (accessed 6 July 2018).

Seelig, H.-D., Hoehn, A., Stodieck, L.S., Klaus, D.M., III, W.W.A., Emery, W.J., 2008. The assessment of leaf water content using leaf reflectance ratios in the visible, near-, and short-wave-infrared. International Journal of Remote Sensing 29, 3701–3713. https://doi.org/10.1080/01431160701772500

Serbin, G., Daughtry, C.S.T., Hunt, E.R., Brown, D.J., McCarty, G.W., 2009. Effect of Soil Spectral Properties on Remote Sensing of Crop Residue Cover. Soil Sci. Soc. Am. J. 73, 1545–1558. https://doi.org/10.2136/sssaj2008.0311

Serrano, L., Peñuelas, J., Ustin, S.L., 2002. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. Remote Sensing of Environment 81, 355–364. https://doi.org/10.1016/S0034-4257(02)00011-1

Shewry, P.R., 2009. Wheat. J Exp Bot 60, 1537–1553. https://doi.org/10.1093/jxb/erp058

Shi, H., Wang, B., Yang, P., Li, Y., Miao, F., 2016. Differences in Sugar Accumulation and Mobilization between Sequential and Non-Sequential Senescence Wheat Cultivars under Natural and Drought Conditions. PLOS ONE 11, e0166155. https://doi.org/10.1371/journal.pone.0166155

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., Muricho, G., 2013. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. Food Sec. 5, 291–317. https://doi.org/10.1007/s12571-013-0263-y

Simmonds, N.W., 1995. The relation between yield and protein in cereal grain. Journal of the Science of Food and Agriculture 67, 309–315. https://doi.org/10.1002/jsfa.2740670306

Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sensing of Environment 81, 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X

Sinclair, T.R., Wit, C.T. de, 1975. Photosynthate and Nitrogen Requirements for Seed Production by Various Crops. Science 189, 565–567. https://doi.org/10.1126/science.189.4202.565

Singh, P., Kanemasu, E.T., 1983. Leaf and Canopy Temperatures of Pearl Millet Genotypes under Irrigated and Nonirrigated Conditions1. Agronomy Journal 75, 497–501. https://doi.org/10.2134/agronj1983.00021962007500030019x

Slafer, G.A., Abeledo, L.G., Miralles, D.J., Gonzalez, F.G., Whitechurch, E.M., 2001. Photoperiod Sensitivity during Stem Elongation as an Avenue to Raise Potential Yield in Wheat, in: Bedö, Z., Láng, L. (Eds.), Wheat in a Global Environment, Developments in Plant Breeding. Springer Netherlands, pp. 487–496. https://doi.org/10.1007/978-94-017-3674-9_64

Slafer, G.A., Andrade, F.H., Feingold, S.E., 1990. Genetic improvement of bread wheat (Triticum aestivum L.) in Argentina: relationships between nitrogen and dry matter. Euphytica 50, 63–71. https://doi.org/10.1007/BF00023162

Slafer, G.A., Kantolic, A.G., Appendino, M.L., Tranquilli, G., Miralles, D.J., Savin, R., 2015. Chapter 12 - Genetic and environmental effects on crop development determining adaptation and yield, in: Sadras, V.O., Calderini, D.F. (Eds.), Crop Physiology (Second Edition). Academic Press, San Diego, pp. 285–319. https://doi.org/10.1016/B978-0-12-417104-6.00012-1

Slafer, G.A., Rawson, H.M., 1996. Responses to photoperiod change with phenophase and temperature during wheat development. Field Crops Research 46, 1–13. https://doi.org/10.1016/0378-4290(95)00081-X

Southgate, D.D., Graham, D.H., Tweeten, Luther G., 2010. The World Food Economy, 2nd Edition. ed.

Spano, G., Di Fonzo, N., Perrotta, C., Platani, C., Ronga, G., Lawlor, D.W., Napier, J.A., Shewry, P.R., 2003. Physiological characterization of 'stay green' mutants in durum wheat. J Exp Bot 54, 1415–1420. https://doi.org/10.1093/jxb/erg150

Stewart, E.L., Hagerty, C.H., Mikaberidze, A., Mundt, C.C., Zhong, Z., McDonald, B.A., 2016. An Improved Method for Measuring Quantitative Resistance to the Wheat Pathogen Zymoseptoria tritici Using High-Throughput Automated Image Analysis. Phytopathology 106, 782–788. https://doi.org/10.1094/PHYTO-01-16-0018-R

Stewart, E. l, Croll, D., Lendenmann, M.H., Sanchez-Vallet, A., Hartmann, F.E., Palma-Guerrero, J., Ma, X., McDonald, B.A., 2018. Quantitative trait locus mapping reveals complex genetic architecture of quantitative virulence in the wheat pathogen Zymoseptoria tritici. Molecular Plant Pathology 19, 201–216. https://doi.org/10.1111/mpp.12515

Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8, 25. https://doi.org/10.1186/1471-2105-8-25

Stuckens, J., Swennen, R.L., Coppin, P., Dzikiti, S., Verreynne, S., Verstraeten, W.W., 2011. EXTRACTING PHYSIOLOGICAL INFO FROM A HYPERSPECTRAL TIME SERIES OF A CITRUS ORCHARD. Acta Horticulturae 11–18. https://doi.org/10.17660/ActaHortic.2011.919.1

Sun, J., Poland, J.A., Mondal, S., Crossa, J., Juliana, P., Singh, R.P., Rutkoski, J.E., Jannink, J.-L., Crespo-Herrera, L., Velu, G., Huerta-Espino, J., Sorrells, M.E.,

2019. High-throughput phenotyping platforms enhance genomic selection for wheat grain yield across populations and cycles in early stage. Theor Appl Genet 132, 1705–1720. https://doi.org/10.1007/s00122-019-03309-0

Swissgranum, 2019. Inlandproduktion [WWW Document]. swissgranum. URL https://www.swissgranum.ch/inlandproduktion (accessed 1.30.20).

Sylvester-Bradley, R., Scott, R.K., Wright, C.E., 1990. Physiology in the production and improvement of cereals. Physiology in the production and improvement of cereals.

Szeliski, R., 2011. Structure from motion, in: Szeliski, R. (Ed.), Computer Vision: Algorithms and Applications, Texts in Computer Science. Springer, London, pp. 303–334. https://doi.org/10.1007/978-1-84882-935-0_7

Takebe, M., Yoneyama, T., Inada, K., Murakami, T., 1990. Spectral reflectance ratio of rice canopy for estimating crop nitrogen status. Plant Soil 122, 295–297. https://doi.org/10.1007/BF02851988

Taulemesse, F., Le Gouis, J., Gouache, D., Gibon, Y., Allard, V., 2016. Bread Wheat (Triticum aestivum L.) Grain Protein Concentration Is Related to Early Post-Flowering Nitrate Uptake under Putative Control of Plant Satiety Level. PLoS One 11. https://doi.org/10.1371/journal.pone.0149668

Tester, M., Langridge, P., 2010. Breeding Technologies to Increase Crop Production in a Changing World. Science 327, 818–822. https://doi.org/10.1126/science.1183700

Thomas, H., Ougham, H., 2014. The stay-green trait. J Exp Bot 65, 3889–3900. https://doi.org/10.1093/jxb/eru037

Thomas, H., Smart, C.M., 1993. Crops that stay green. Annals of Applied Biology 123, 193–219. https://doi.org/10.1111/j.1744-7348.1993.tb04086.x

Thomas, S., Behmann, J., Steier, A., Kraska, T., Muller, O., Rascher, U., Mahlein, A.-K., 2018. Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. Plant Methods 14, 45. https://doi.org/10.1186/s13007-018-0313-8

Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. PNAS 108, 20260–20264. https://doi.org/10.1073/pnas.1116437108

Torriani, S.F.F., Melichar, J.P.E., Mills, C., Pain, N., Sierotzki, H., Courbot, M., 2015. Zymoseptoria tritici: A major threat to wheat production, integrated approaches to control. Fungal Genetics and Biology, Septoria tritici blotch disease of wheat: Tools and techniques to study the pathogen Zymoseptoria tritici 79, 8–12. https://doi.org/10.1016/j.fgb.2015.04.010

Triboi, E., Triboi-Blondel, A.-M., 2002. Productivity and grain or seed composition: a new approach to an old problem—invited paper. European Journal of Agronomy 16, 163–186. https://doi.org/10.1016/S1161-0301(01)00146-0

Trnka, M., Hlavinka, P., Semenov, M.A., 2015. Adaptation options for wheat in Europe will be limited by increased adverse weather events under climate change. Journal of The Royal Society Interface 12, 20150721. https://doi.org/10.1098/rsif.2015.0721

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Uauy, C., Brevis, J.C., Dubcovsky, J., 2006. The high grain protein content gene Gpc-B1 accelerates senescence and has pleiotropic effects on protein content in wheat. J Exp Bot 57, 2785–2794. https://doi.org/10.1093/jxb/erl047

van Eeuwijk, F.A., Bustos-Korts, D., Millet, E.J., Boer, M.P., Kruijer, W., Thompson, A., Malosetti, M., Iwata, H., Quiroz, R., Kuppe, C., Muller, O., Blazakis, K.N., Yu, K., Tardieu, F., Chapman, S.C., 2018. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. Plant Science. https://doi.org/10.1016/j.plantsci.2018.06.018

van Ginkel, M., Reynolds, M.P., Trethowan, R., Hernandez, E., 2008. Complementing the Breeder's Eye with Canopy Temperature Measurements., in: International Symposium on Wheat Yield Potential: Challenges to International Wheat Breeding. CIMMYT, Mexico, pp. 134–135.

van Oosterom, E.J., Chapman, S.C., Borrell, A.K., Broad, I.J., Hammer, G.L., 2010. Functional dynamics of the nitrogen balance of sorghum. II. Grain filling period. Field Crops Research 115, 29–38. https://doi.org/10.1016/j.fcr.2009.09.019

Verma, V., Foulkes, M.J., Worland, A.J., Sylvester-Bradley, R., Caligari, P.D.S., Snape, J.W., 2004. Mapping quantitative trait loci for flag leaf senescence as a yield determinant in winter wheat under optimal and drought-stressed environments. Euphytica 135, 255–263. https://doi.org/10.1023/B:EUPH.0000013255.31618.14

Vogelmann, J.E., Rock, B.N., Moss, D.M., 1993. Red edge spectral measurements from sugar maple leaves. International Journal of Remote Sensing 14, 1563–1575. https://doi.org/10.1080/01431169308953986

Wahabzada, M., Mahlein, A.-K., Bauckhage, C., Steiner, U., Oerke, E.-C., Kersting, K., 2016. Plant Phenotyping using Probabilistic Topic Models: Uncovering the Hyperspectral Language of Plants. Scientific Reports 6, 22482. https://doi.org/10.1038/srep22482

Wahabzada, M., Mahlein, A.-K., Bauckhage, C., Steiner, U., Oerke, E.-C., Kersting, K., 2015. Metro Maps of Plant Disease Dynamics—Automated Mining of Differences Using Hyperspectral Images. PLOS ONE 10, e0116902. https://doi.org/10.1371/journal.pone.0116902

Wakie, T.T., Kumar, S., Senay, G.B., Takele, A., Lencho, A., 2016. Spatial prediction of wheat septoria leaf blotch (Septoria tritici) disease severity in Central Ethiopia. Ecological Informatics 36, 15–30. https://doi.org/10.1016/j.ecoinf.2016.09.003

Walter, A., Liebisch, F., Hund, A., 2015. Plant phenotyping: from bean weighing to image analysis. Plant Methods 11, 14. https://doi.org/10.1186/s13007-015-0056-8

Wingler, A., Masclaux-Daubresse, C., Fischer, A.M., 2009. Sugars, senescence, and ageing in plants and heterotrophic organisms. J Exp Bot 60, 1063–1066. https://doi.org/10.1093/jxb/erp067

Wingler, A., Roitsch, T., 2008. Metabolic regulation of leaf senescence: interactions of sugar signalling with biotic and abiotic stress responses. Plant Biology 10, 50–62. https://doi.org/10.1111/j.1438-8677.2008.00086.x

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

Wright, M.N., Ziegler, A., 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software 77. https://doi.org/10.18637/jss.v077.i01

Wu, C., Niu, Z., Tang, Q., Huang, W., 2008. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. Agricultural and Forest Meteorology 148, 1230–1241. https://doi.org/10.1016/j.agrformet.2008.03.005

Wu, X.-Y., Kuai, B.-K., Jia, J.-Z., Jing, H.-C., 2012. Regulation of Leaf Senescence and Crop Genetic ImprovementF. Journal of Integrative Plant Biology 54, 936–952. https://doi.org/10.1111/jipb.12005

Xie, Q., Mayes, S., Sparkes, D.L., 2016. Early anthesis and delayed but fast leaf senescence contribute to individual grain dry matter and water accumulation in wheat. Field Crops Research 187, 24–34. https://doi.org/10.1016/j.fcr.2015.12.009

Xose Rodriguez-Alvarez, M., Boer, M.P., van Eeuwijk, F.A., Eilers, P.H.C., 2018. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spatial Statistics 23, 52–71. https://doi.org/10.1016/j.spasta.2017.10.003

Xue, L., Cao, W., Luo, W., Dai, T., Zhu, Y., 2004. Monitoring Leaf Nitrogen Status in Rice with Canopy Spectral Reflectance. Agronomy Journal 96, 135. https://doi.org/10.2134/agronj2004.0135

Yang, C.-M., 2010. Assessment of the severity of bacterial leaf blight in rice using canopy hyperspectral reflectance. Precision Agric 11, 61–81. https://doi.org/10.1007/s11119-009-9122-4

Yang, J., Zhang, J., 2006. Grain filling of cereals under soil drying. New Phytologist 169, 223–236. https://doi.org/10.1111/j.1469-8137.2005.01597.x

Yendrek, C.R., Tomaz, T., Montes, C.M., Cao, Y., Morse, A.M., Brown, P.J., McIntyre, L.M., Leakey, A.D.B., Ainsworth, E.A., 2017. High-Throughput Phenotyping of Maize Leaf Physiological and Biochemical Traits Using Hyperspectral Reflectance. Plant Physiology 173, 614–626. https://doi.org/10.1104/pp.16.01447

Yu, K., Anderegg, J., Mikaberidze, A., Karisto, P., Mascher, F., McDonald, B.A., Walter, A., Hund, A., 2018. Hyperspectral Canopy Sensing of Wheat Septoria Tritici Blotch Disease. Front. Plant Sci. 9. https://doi.org/10.3389/fpls.2018.01195

Yu, K., Kirchgessner, N., Grieder, C., Walter, A., Hund, A., 2017. An image analysis pipeline for automated classification of imaging light conditions and for quantification of wheat canopy cover time series in field phenotyping. Plant Methods 13, 15. https://doi.org/10.1186/s13007-017-0168-4

Yu, K., Lenz-Wiedemann, V., Chen, X., Bareth, G., 2014. Estimating leaf chlorophyll of barley at different growth stages using spectral indices to reduce soil background and canopy structure effects. ISPRS Journal of Photogrammetry and Remote Sensing 97, 58–77. https://doi.org/10.1016/j.isprsjprs.2014.08.005

Zahedi, M., Jenner, C.F., 2003. Analysis of effects in wheat of high temperature on grain filling attributes estimated from mathematical models of grain filling. Journal of Agricultural Science 141, 203–212. https://doi.org/10.1017/S0021859603003411

Zarco-Tejada, P.J., Berjón, A., López-Lozano, R., Miller, J.R., Martín, P., Cachorro, V., González, M.R., de Frutos, A., 2005. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. Remote Sensing of Environment 99, 271–287. https://doi.org/10.1016/j.rse.2005.09.002

Zarco-Tejada, P.J., González-Dugo, V., Williams, L.E., Suárez, L., Berni, J.A.J., Goldhamer, D., Fereres, E., 2013. A PRI-based water stress index combining structural and chlorophyll effects: Assessment using diurnal narrow-band airborne imagery and the CWSI thermal index. Remote Sensing of Environment 138, 38–50. https://doi.org/10.1016/j.rse.2013.07.024

Zarco-Tejada, P.J., Rueda, C.A., Ustin, S.L., 2003. Water content estimation in vegetation with MODIS reflectance data and model inversion methods. Remote Sensing of Environment 85, 109–124. https://doi.org/10.1016/S0034-4257(02)00197-9

Zhan, J., Kema, G.H.J., Waalwijk, C., McDonald, B.A., 2002. Distribution of mating type alleles in the wheat pathogen Mycosphaerella graminicola over spatial scales from lesions to continents. Fungal Genetics and Biology 36, 128–136. https://doi.org/10.1016/S1087-1845(02)00013-0

Zhang, J., Pu, R., Huang, W., Yuan, L., Luo, J., Wang, J., 2012. Using in-situ hyperspectral data for detecting and discriminating yellow rust disease from nutrient stresses. Field Crops Research 134, 165–174. https://doi.org/10.1016/j.fcr.2012.05.011

Zheng, Q., Huang, W., Cui, X., Dong, Y., Shi, Y., Ma, H., Liu, L., 2019. Identification of Wheat Yellow Rust Using Optimal Three-Band Spectral Indices in Different Growth Stages. Sensors 19, 35. https://doi.org/10.3390/s19010035

Zhong, Z., Marcel, T.C., Hartmann, F.E., Ma, X., Plissonneau, C., Zala, M., Ducasse, A., Confais, J., Compain, J., Lapalu, N., Amselem, J., McDonald, B.A., Croll, D., Palma-Guerrero, J., 2017. A small secreted protein in Zymoseptoria tritici is responsible for avirulence on wheat cultivars carrying the Stb6 resistance gene. New Phytologist 214, 619–631. https://doi.org/10.1111/nph.14434

7 Supplementary material

Supplementary material Chapter 2

Supplementary Methods

Calculation of spectral indices and extraction of dynamics parameters

Spectral regions comprising the wavelengths from 1350 nm to 1475 nm, from 1781 nm to 1990 nm and from 2400 nm to 2500 nm were removed because of the very low signal-to-noise ratio resulting from high atmospheric absorption. Spectra were then averaged for each experimental plot and 100 published spectral indices (SI) sensitive to chlorophyll content, water content, N content and total above ground N, leaf area index, green biomass and vegetation cover, and pigment changes and plant senescence were computed according to the formulae reported in Table S1 (Figure 2.1, [1]). Values of SI were scaled to range from 0 to 10, representing the minimum and maximum values recorded during the assessment period, respectively. Scaled values of these SI were then fitted against thermal time after heading using linear interpolation and dynamics parameters were extracted as was done for visual scorings (Figure 2.1, [2]). Parametric models were not used for SI as different SI exhibited distinct temporal patterns and would have required the use of SI-specific models (*compare* Figure 2.5).

Spectral index subset selection

We then performed unsupervised subset selection to reduce multi-collinearity of the dataset using several filtering criteria (Figure 2.1, [3]). Only SI that showed a monotonous or near-monotonous decrease during the assessment period for most experimental plots were retained for further analysis. Repeatability (w^2) of the senescence dynamics parameters derived from the selected SI was then calculated based on data obtained in 2016. Only parameters with w^2 similar to those observed for the visual scorings were retained ($w^2 > 0.6$ for On_{sen}, $w^2 > 0.7$ for Mid_{sen}, $w^2 > 0.8$ for End_{sen} and $w^2 > 0.5$ for T_{sen}). Finally, highly correlated parameters were iteratively removed by calculating a correlation matrix based on three years data and discarding one of two parameters with a pair-wise correlation coefficient > 0.96. Parameters were selected preferring narrow-band SI over broad-band SI, SI with a specific physiological interpretation over more generic SI and SI developed specifically for use in wheat or barley canopies. Additionally, the smoothness of SI values over time was evaluated graphically and used as an additional selection criterion. This selection procedure resulted in a set of 83 SI-derived senescence dynamics parameters based on 51 distinct SI with pair-wise correlation coefficients $r < 0.96$.

Development and validation of full-spectrum models to infer visual senescence scorings

As an alternative to published SI, two multivariate modelling techniques were used to infer senescence scorings from spectral data (Figure 2.1, [4]) : (1) Partial Least Squares Regression (PLSR), which has been used extensively for analyzing field measured hyperspectral reflectance data (*see* Wold et al., 2001 *for details*) and (2) cubist regression, which substantially builds on the M5 model tree (Quinlan, 1992), extending it by incorporating a boosting-like procedure referred to as *committees*. The latter algorithm is more flexible since it does not require the user to specify the nature of the relationship between predictors and the response and is able to capture more

complex non-linear relationships (Kuhn and Johnson, 2013). The R packages *'pls'* V2.7.0 (Mevik *et al*., 2018) and *'Cubist'* (Kuhn *et al*., 2018) were used for the analysis. The hyperparameters of each model were tuned using 10-fold cross-validation, which was repeated 5 times in the case of PLSR, using the R package '*caret*' V6.0.80 (Kuhn, 2008). As the main interest lies on the capability of full-spectrum models to represent the entire process of senescence, average RMSE for 10 different random upsamples of the test data were calculated, with each upsample containing all possible scoring values an equal number of times, i.e. exactly the number of times of the most frequent observation. This was achieved by random sampling with replacement.

A common issue with full-spectrum models is limited applicability on data of experiments not contained in the model training process (*see e.g.* Øvergaard *et al*., 2013). We hypothesized that this problem should be less pronounced in our case, as senescence results in major changes of canopy reflectance that should be similar across years. Still, to maximize model across-year applicability, we evaluated different types of input data: (1) raw reflectance spectra, (2) smoothed reflectance spectra, (3) first derivative of smoothed reflectance spectra, (4) continuum removed smoothed reflectance spectra and (5) raw reflectance spectra limited to the spectral range between 500 nm and 700 nm, as done by Kipp *et al*. (2014). Smoothing of raw spectra and calculation of derivatives was done using the Savitzky-Golay smoothing filter (Savitzky and Golay, 1964) with a window size of 11 spectral bands and a third order polynomial. Continuum removal (Clark and Roush, 1984) was done using linear interpolation between extrema and subtraction for normalization. Spectra pre-preprocessing was done using functions of the R package '*prospectr*' V0.1.3 (Ramirez-Lopez and Stevens, 2014). Each type of input data was tested without pre-treatment and with mean-centering and scaling to unit variance of all predictors. Finally, feature selection was conducted following the procedure described in the next section to identify the most important wavelengths for each year. Model evaluation was done within experiments (years) as well as across experiments to estimate the robustness of the developed models. Validation datasets were randomly up-sampled to obtain an estimate of model performance across the whole process of senescence.

Supervised feature selection
Supervised feature selection was performed by recursive feature elimination using a nested cross-validation approach. Cubist was used as a base-learner for full-spectrum models to infer senescence scorings, while rf was used as a base-learner for selection of scoring- and SI-based senescence dynamics parameters. The '*ranger*' implementation (Wright and Ziegler, 2017) of the original rf algorithm (Breiman, 2001) was used (R package *'ranger'* V0.10.1). The *mtry* hyperparameter was tuned using 10-fold cross-validation, while variance reduction was defined as a split criterion, the minimal node size was set to five samples and the number of trees was set to 10,000, resulting in stable feature importance rankings for given resamples. Feature importance was calculated based on permutation. In order to obtain a robust ranking of features and measure of model performance, the data set was resampled 30 times into training and test sets with an 80:20 split. For each resample, a rf was trained using all features, as described above. Then, from the set of original features, those with the lowest importance values were removed iteratively in 29 decreasing steps until only one feature remained in the model. After each iteration, feature

importance was re-calculated due to remaining multi-collinearity in the feature set and performance was assessed on the corresponding test set. The results of the 30 resamples were then aggregated to obtain a performance profile over the feature subset sizes and robust feature importance rankings. For full-spectrum models, the procedure was identical. Here, the varImp() function of the R package *'caret'* was used to extract variable importance in each iteration. The returned measure is a linear combination of the usage in rule conditions and the model. For more details, we refer to Kuhn and Johnson (2013).

Supplementary Tables

Table S 1 Hyperspectral vegetation indices used in this study to track senescence dynamics.

Index	Application	Formula	Reference
ANTH	Anthocyanin	R760to800*(1/R540to560-1/R690to710)	(Gitelson *et al.*, 2006)
ARI	Anthocyanin	1/(R549to551*100)-1/(R699to701*100)	(Gitelson *et al.*, 2001)
ARVI	LAI/vegetation fraction/green biomass	(R845to885-(R845to885-R460to480+R845to885))/(R845to885+(R635to685-R460to480+R635to685))	(Kaufman and Tanre, 1992)
CAI	Plant litter / crop residues	0.5*(R2019to2021*100+R2219to2221*100)-R2099to2101*100	(Nagler *et al.*, 2000)
CARG	Car	R760to800*(1/R510to520-1/R540to560)	(Gitelson *et al.*, 2006)
CARRE	Car	R760to800*(1/R510to520-1/R690to710)	(Gitelson *et al.*, 2006)
CHLG	Chl	R760to800/R540to560	(Gitelson *et al.*, 2006)
CHLRE	Chl	R760to800/R690to720-1	(Gitelson *et al.*, 2006)
CIG	Chl / LAI	R750to800/R520to585-1	(Gitelson *et al.*, 2003)
CIRE	Chl / LAI	R750to800/R695to740-1	(Gitelson *et al.*, 2003)
CLSI	Disease discrimination	(R697to699-R569to571)/(R697to699+R569to571)-R733to735	(Mahlein *et al.*, 2013)
CNHI	Senescence	(R8455to890-R1200to1300)*(R775to805+R630to690))/(R8455to890+ R1200to1300)*(R775to805- R630to690))	(Pimstein *et al.*, 2009)
CRI1	Car	1/R506to514-1/R535to565	(Gitelson *et al.*, 2002b)
CRI2	Car	1/R506to514-1/R693to707	(Gitelson *et al.*, 2002b)
DCNI	Plant N concentration	(R717to723-R697to703)/(R697to703-R667to673)/(R717to723- R667to673+0.03)	(Chen *et al.*, 2010)
DSWI	Disease discrimination	(R795to805+R543to553)/R1654to1664+R676to686)	(Apan *et al.*, 2004)
EVI	LAI / Vegetation fraction	2.5*(R841to876-R620to670)/(R841to876+6*R620to670-7.5*R459to479+1)	(Huete *et al.*, 2002)
FII	Flowering	(R470to480-R360to370)/(R470to480+R360to370)	(Stuckens *et al.*, 2011)
GBNDVI	N concentration / LAI	(R567to579-R438to442)/(R567to579+R438to442)	(Hansen and Schjoerring, 2003)
GLI	Vegetation cover / Chl	(2*R580to610-R580to660-R400to520)/(2*R580to610+ R580to660+ R400to520)	(Louhaichi *et al.*, 2001)
GM	Chl	R841to876/R545to565 - 1	(Gitelson *et al.*, 2005)
HI	Disease discrimination	(R533to535-R697to698)/(R533to535+R697to699)-1/2*R703to705	(Mahlein *et al.*, 2013)
LCI	Chl / Senescence	(R849to851+R709to711)/(R849to851+R679to681)	(Datt, 1999)
LCI2	Chl / Senescence	(R849to851-R709to711)/(R849to851+R679to681)	(Datt, 1999)
LWI	Leaf water content / EWT	R1250to1350/R1400to1500	(Seelig *et al.*, 2008)
LWVI1	Leaf water	(R1089to1099-R888to898)/(R1089to1099+R888to898)	(Galvão *et al.*, 2005)
LWVI2	Leaf water	(R1089to1099-R1201to1209)/(R1089to1099+R1201to1209)	(Galvão *et al.*, 2005)
MCARI	LAI	{1.5*[2.5*(R797to803 - R676to673)-1.3*(R797to803 – R547to553)]}/ sqrt[(2* R797to803 + 1)^2-(6* R797to803-5*sqrt(R676to673)) - 0.5]	(Haboudane *et al.*, 2004)
MCARI/MTVI2	Leaf N concentration (SPAD)	MCARI/MTVI2	(Eitel *et al.*, 2007)

MCARI/OSAVI	Chl	MCARI/OSAVI	(Wu *et al*., 2008)
mND705	Chl	(R749to751-R704to706)/(R749to751+ R704to706-2*R444to446)	(Sims and Gamon, 2002)
MSAVI	Green cover / LAI	0.5*(2*R790to890+1-sqrt((2*R790to890+1)^2-8*(R790to890-R610to690)))	(Qi *et al*., 1994)
MSR$_{rev}$	Chl	[(R750/R705) + 1]/sqrt[(R750/R705) + 1]	(Wu *et al*., 2008)
MTCI	Chl	(R751to757-R705to713)/(R705to713-R678to684)	(Dash and Curran, 2004)
MTVI1	LAI	1.2*(1.2*(R797to803-R547to553)-2.5*(R667to673-R547to553))	(Haboudane *et al*., 2004)
MTVI2	LAI	{1.5*[1.2*(R797to803+R547to553)+2.5*(R667toR773+R547to553)]}/ sqrt[(2*R797to803+1)^2 + (6*R797to803+5*sqrt(R667toR773))+0.5]	(Haboudane *et al*., 2004)
NDLI	foliar/bulk canopy lignin (senescence) and N concentration	(log(1/R1750to1758)-log(1/ R1676to1684))/(log(1/ R1750to1758)+log(1/ R1676to1684))	(Serrano *et al*., 2002)
NDMI	Vegetation water content	(R760to900-R1550to1750)/(R760to900+R1550to1750)	(Hardisky *et al*., 1984)
NDNI	Foliar/bulk canopy N concentration	[log(1/R1506to1514) - log(1/R1676to1684)] / [log(1/ R1506to1514) + log(1/ R1676to1684)]	(Eitel *et al*., 2007)
NDRE	Crop cover / Chl / water / N stress	(R786to794-R716to724)/(R786to794+R716to724)	(Barnes *et al*., 2000)
NDREI	Chl-a / Senescence	(R749to751-R704to706)/(R749to751+R704to706)	(Gitelson and Merzlyak, 1994)
NDSVI	Residues / Senescence	(R1550to1750-R630to690)/(R1550to1750+R630to690)	(Qi *et al*., 2002)
NDTI	Tillage (residue cover) / soil plain	(R1550to1750-R2080to2350)/(R1550to1750+R2080to2350)	(Deventer *et al*., 1997)
NDVI	LAI / Vegetation fraction / Green biomass	(R799to801-R669to671)/(R799to801+R669to671)	(Rouse *et al*., 1974)
NDWI	Water content	(R856to864-R1236to1244)/(R856to864+R1236to1244)	(Gao, 1996)
NDWI1650	Water content	(R770to910 – R1550to1750)/(R770to910 + R1550to1750)	(Clay *et al*., 2006)
NDWI2130	Water content	(R841to876-R2105to2155)/(R841to876+R2105to2155)	(Chen *et al*., 2005)
NGRDI	Biomass / Water content / Chl	(R520to600 – R630to690)/(R520to600 + R630to690)	(Tucker, 1979)
NHI	Heading of wheat	(R1098to1102-R1198to1202)/(R1098to1102+R1198to1202)	(Pimstein *et al*., 2009)
NPCI	Chl/Car ratio	(R678to682-R428to432)/(R678to682+R428to432)	(Peñuelas *et al*., 1994)
OCAR	Chl	R629to631/R679to681	(Schlemmer *et al*., 2005)
OSAVI	Vegetation cover / LAI / Chl	(1+0.16)*(R797to803-R667to673)/(R797to803+R667to673+0.16)	(Rondeaux *et al*., 1996)
PBI	Chl / N	R806to814/R556to564	(Rao *et al*., 2008)
PMI	Disease discrimination	(R519to521-R583to583)/(R519to521+R583to583)+R723to725	(Mahlein *et al*., 2013)

PRInorm	Xanthophyll / Car / Canopy leaf area	$[(R570-R531)/(R570+R531)] / \{[(R800-R670) /sqrt(R800+R670)]*(R700/R670)\}$	(Zarco-Tejada *et al*., 2013)
PSND1	Chl / Car	(R799to801-R674to676)/(R799to801+R674to676)	(Blackburn, 1998)
PSND2	Chl / Car	(R799to801-R649to651)/(R799to801+R649to651)	(Blackburn, 1998)
PSND3	Chl / Car	(R799to801-R499to501)/(R799to801+R499to501)	(Blackburn, 1998)
PSND4	Chl / Car	(R799to801-R469to471)/(R799to801+ R469to471)	(Blackburn, 1998)
PSRI	Pigment changes / Senescence	(R677to679-R499to501)/R749to751	(Merzlyak *et al*., 1999)
PSSR1	Chl/Car	R799to801/R674to676	(Blackburn, 1998)
PSSR2	Chl/Car	R799to801/R649to651	(Blackburn, 1998)
PSSR3	Chl/Car	R799to801/R499to501	(Blackburn, 1998)
R1200	Leaf water content	(2*R1180to1220)/(R1090to1110+R1265to1285)	(Pu, 2011)
R760/R730	Dry matter / Total aerial N	R759to761/R729to731	(Mistele and Schmidhalter, 2010b)
R780/R550	Dry matter / Total aerial N	R779to781/R549to551	(Takebe *et al*., 1990)
R780/R700	Dry matter / Total aerial N	R779to781/R699to701	(Mistele and Schmidhalter, 2010a)
R780/R740	Dry matter / Total aerial N	R779to781/R739to741	(Mistele and Schmidhalter, 2010a)
R970/R900	Dry matter / Total aerial N	R969to971/R899to901	(Mistele and Schmidhalter, 2010b)
R975	Water content	(2*R960to990)/(R920to940 + R1090to1110)	(Pu *et al*., 2003)
REIP	Dry matter / Total aerial N	700+40*[(R669to671+R779to781)/2-R699to701]/(R739to741-R699to701)	(Guyot *et al*., 1988)
RGR	Anthocyanin / LAI	R682toR684/R509toR511	(Gamon and Surfus, 1999)
RGR2	Anthocyanin / LAI	R600to699/R500to599	(Gamon and Surfus, 1999)
RRDI~Red-edge~	Chl	(R744to746-R739to741)/(R739to741-R699to701)	(Yu *et al*., 2014)
RVI2	Leaf N accumulation (g N m^{-2})	R805to815/R655to665	(Xue *et al*., 2004)
SAVI	LAI / vegetation fraction / (green biomass)	(1.5)*(R857to863-R667to673)/(R857to863+R667to673+0.5)	(Huete, 1988)
SBRI	Disease discrimination	R569to571-R512to514)/(R569to571+R512to514)-1/2*R703to705	(Mahlein *et al*., 2013)
SGR	Green vegetation cover	$\sum$ R500to599	(Fuentes *et al*., 2001)
SINDRI	Fraction residue cover	(R2185to2225-R2235to2285)/(R2185to2225+R2235to2285)	(Serbin *et al*., 2009)
SIPI	Chl / Car / (Senescence)	R799to801-R444to446)/(R799to801-R679to681)	(Penuelas *et al*., 1995)

SIPI	Chl, Car / LUE / Senescence	(R799to801−R444to446)/(R799to801−R679to681)	(Penuelas *et al*., 1995)
SIWSI	Water stress	(R841to876-R1628to1652)/(R841to876+R1628to1652)	(Fensholt and Sandholt, 2003)
SLAIDI1	LAI	5*(R1049to1051-R1249to1251)/(R1049to1051+R1249to1251)	(Delalieux *et al*., 2008)
SLAIDI2	LAI	40*R1554to1556*(R1049to1051-R1249to1251)/(R1049to1051+R1249to1251)	(Delalieux *et al*., 2008)
SRWI	Water content	R841to876/R1230to1250	(Zarco-Tejada *et al*., 2003)
TCARI	Chl	3*((R697to703-R667to673)-0.2*(R697to703-R547to553)*(R697to703/ R667to673))	(Haboudane *et al*., 2002)
TCARI/OSAVI	Chl	TCARI/OSAVI	(Haboudane *et al*., 2002)
TGI	Chl	-0.5*(190*(R666to674-R546toR554)-120*(R666to674-R476to484))	(Hunt *et al*., 2011)
VARIgreen	Vegetation fraction	(R546to556-R620to670)/(R546to556+R620to670-R459to479)	(Gitelson *et al*., 2002a)
VI700	Vegetation fraction	(R700to710-R620to670)/(R700to710+R620to670)	(Gitelson *et al*., 2002a)
VIgreen	Vegetation fraction	(R546to556-R620to670)/(R546to556+R620to670)	(Gitelson *et al*., 2002a)
VIopt	N in crop (kg N ha^{-1})	(1+0.45)*((R760to900)*2+1)/(R630to690+0.45)	(Reyniers *et al*., 2006)
VOG1	Chl	R739to741/R719to721	(Vogelmann *et al*., 1993)
VOG2	Chl	(R733to735-R746to748)/(R714to716+R725toR727)	(Vogelmann *et al*., 1993)
VOG3	Chl	R733to735-R746to748)/(R714to716+R719to721)	(Vogelmann *et al*., 1993)
WDRVI	LAI / Vegetation fraction / (green biomass)	(0.1*R750to1000-R580to680)/(0.1*R750to1000+R580to680)	(Gitelson, 2004)
WI	Leaf water content	R896to904/R966to974	(Peñuelas *et al*., 1993)
WI/NDVI	Plant water content	(R896to904/R966to974)/((R796to804-R676to684)/(R796to804-R676to684))	(Penuelas *et al*., 1997)
YCAR	Chl	R599to601/R679to681	(Schlemmer *et al*., 2005)

Table S 2 Least squares coefficient estimates associated with the regression of grain yield (GY) or grain protein concentration (GPC) in 2016-2018 onto heading (in GDD after sowing) and midpoint of senescence (midsen, in GDD after heading) as assessed visually. In a first step, the interaction term was included in the model, but was subsequently dropped, as the analysis did not suggest the presence of a statistically significant interaction. Bold numbers highlight significant p values at the significance threshold of 0.05.

	Trait	Coefficient	Std.error	t-statistic	p-value
GY 2016	heading	-1.00E-02	1.24E-02	-0.8	0.42
	midsen	-1.45E-02	2.71E-02	-0.54	0.59
	heading:midsen	1.30E-05	1.76E-05	0.74	0.46
GY 2017	heading	-2.94E-03	1.42E-02	-0.21	0.83
	midsen	-5.51E-05	2.33E-02	0.00	1.00
	heading:midsen	2.83E-06	2.06E-05	0.14	0.89
GY 2018	heading	8.82E-03	1.09E-02	1.54	0.12
	midsen	1.78E-02	2.36E-02	1.73	0.08
	heading:midsen	-1.07E-05	1.53E-05	-1.89	0.06
GPC 2016	heading	1.68E-02	1.09E-02	1.54	0.12
	midsen	4.08E-02	2.36E-02	1.73	0.08
	heading:midsen	-2.88E-05	1.53E-05	-1.89	0.06
GPC 2017	heading	-2.40E-02	1.13E-02	-2.12	**0.03**
	midsen	-4.10E-02	1.86E-02	-2.20	**0.03**
	heading:midsen	3.16E-05	1.64E-05	1.93	0.05
GY 2016	heading	-7.92E-04	1.01E-03	-0.79	0.43
	midsen	5.37E-03	9.34E-04	5.75	**0.00**
GY 2017	heading	-9.91E-04	9.23E-04	-1.07	0.28
	midsen	3.15E-03	1.12E-03	2.82	**0.01**
GY 2018	heading	1.77E-03	7.10E-04	2.50	**0.01**
	midsen	2.86E-03	8.38E-04	3.41	**0.00**
GPC 2016	heading	-3.69E-03	8.80E-04	-4.19	**0.00**
	midsen	-3.57E-03	8.16E-04	-4.37	**0.00**
GPC 2017	heading	-2.24E-03	7.38E-04	-3.03	**0.00**
	midsen	-5.23E-03	8.93E-04	-5.86	**0.00**

Supplementary material Chapter 3

Supplementary Tables

Table S 3 Ranks of spectral-temporal features as determined by recursive feature elimination using random forest (rf) and cubist regression as base learners. Mean feature rank and standard deviation (sd) are reported based on 30 resamples of the data for the top 15 spectral-temporal features. The sensitive SVI (SVI_sen) and insensitive SVI (SVI_insen) constituting the spectral-temporal features and the dynamics parameter used are also reported, where Δ stands for the difference between the parameter values derived from the corresponding SVIs.

Spectral-temporal feature	Base Learner	Mean rank	sd	SVI_sen	SVI_insen	Parameter
SI_MCARI2-SI_SIPI_r-M_delta	rf	1.00	0.00	MCARI2	SIPI	Δ(M)
SI_DSWI-SI_780_740-M_delta	rf	6.17	3.81	DSWI	R780/R740	Δ(M)
SI_VOG1-SI_780_740-M_delta	rf	6.30	4.07	VOG1	R780/R740	Δ(M)
SI_MCARI2-SI_780_740-M_delta	rf	8.50	4.23	MCARI2	R780/R740	Δ(M)
SI_WI-SI_PRInorm_r-M_delta	rf	9.03	4.13	WI	PRInorm	Δ(M)
SI_MCARI2-SI_PRInorm_r-M_delta	rf	9.93	5.54	MCARI2	PRInorm	Δ(M)
SI_DSWI-SI_PRInorm_r-M_delta	rf	11.20	5.41	DSWI	PRInorm	Δ(M)
SI_YCAR-SI_FII_r-M_delta	rf	11.47	7.24	YCAR	FII	Δ(M)
SI_VOG1-SI_SIPI_r-M_delta	rf	12.30	5.69	VOG1	SIPI	Δ(M)
SI_DSWI-SI_GNDVI_HI-M_delta	rf	12.53	3.70	DSWI	GNDVI	Δ(M)
SI_DSWI-SI_SIPI_r-M_delta	rf	13.27	6.60	DSWI	SIPI	Δ(M)
SI_DSWI-SI_DCNI_ASD-M_delta	rf	13.77	6.97	DSWI	DCNI	Δ(M)
SI_MCARI2-SI_SR-b_ratio	rf	14.53	5.47	MCARI2	SR	ratio(b)
SI_NGRDI-SI_SIPI_r-sen15_delta	rf	14.97	4.56	NGRDI	SIPI	Δ(sen15)
SI_NDWI2130-SI_SIPI_r-sen15_delta	rf	15.00	5.02	NDWI2130	SIPI	Δ(sen15)
SI_MCARI2-SI_SIPI_r-M_delta	cubist	1.03	0.18	MCARI2	SIPI	Δ(M)
SI_NGRDI-SI_PRInorm_r-M_delta	cubist	8.80	7.46	NGRDI	PRInorm	Δ(M)
SI_DSWI-SI_780_740-M_delta	cubist	13.10	3.36	DSWI	R780/R740	Δ(M)
SI_DSWI-SI_FII_r-M_delta	cubist	13.33	4.76	DSWI	FII	Δ(M)
SI_DSWI-SI_GNDVI_HI-M_delta	cubist	13.50	4.16	DSWI	GNDVI	Δ(M)
SI_MCARI2-SI_PRInorm_r-M_delta	cubist	14.57	5.52	MCARI2	PRInorm	Δ(M)
SI_DSWI-SI_PRInorm_r-M_delta	cubist	14.67	3.52	DSWI	PRInorm	Δ(M)
SI_HI-SI_CHLRE-M_delta	cubist	14.70	3.70	HI	CHLRE	Δ(M)
SI_VI700-SI_CHLRE-M_delta	cubist	14.77	7.67	VI700	CHLRE	Δ(M)
SI_DSWI-SI_SIPI_r-M_delta	cubist	15.00	3.81	DSWI	SIPI	Δ(M)
SI_MCARI2-SI_780_740-sen15_delta	cubist	15.13	4.99	MCARI2	R780/R740	Δ(sen15)
SI_VOG1-SI_780_740-M_delta	cubist	15.40	5.76	VOG1	R780/R740	Δ(M)
SI_DSWI-SI_FII_r-sen15_delta	cubist	15.60	2.44	DSWI	FII	Δ(sen15)
SI_NDVI_nb_ASD-SI_GNDVI_HI-M_delta	cubist	15.63	5.54	NDVI	GNDVI	Δ(M)
SI_DSWI-SI_780_740-sen15_delta	cubist	15.67	1.71	DSWI	R780/R740	Δ(sen15)
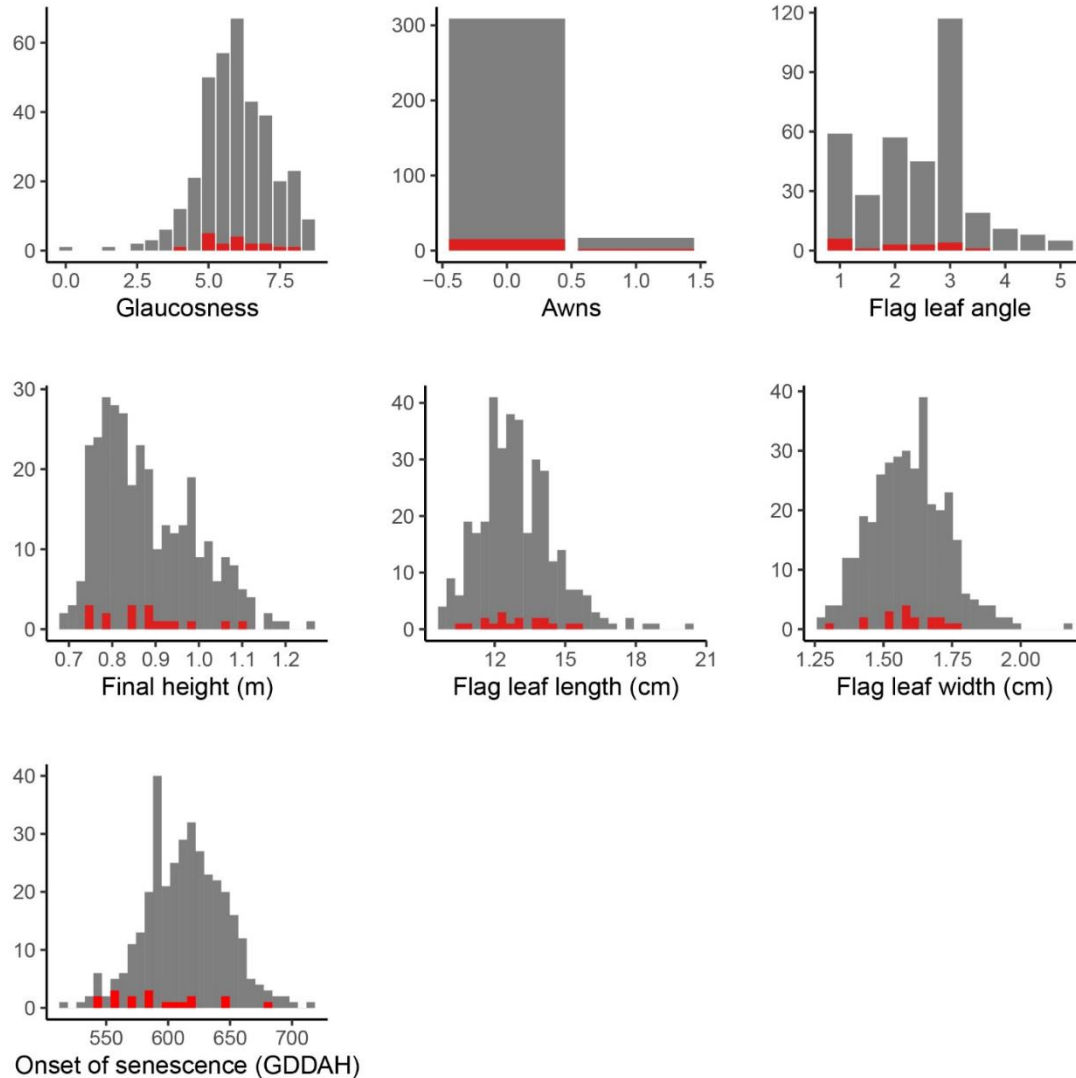
Figure S 7.1 Distribution of important morphological, phenological and canopy structural traits in the GABI wheat panel and in the subset of genotypes selected for the present study. Flag leaf glaucousness, presence or absence of awns, flag leaf angle, flag leaf length and flag leaf width were assessed in 2018, following guidelines provided by Pask *et al.* (2012). For Final height and onset of senescence, distributions of best linear unbiased estimators from experiments conducted in three consecutive years at the same location are shown. For details on methods to determine these traits, we refer to Anderegg *et al.* (2020), Kronenberg *et al.* (2017) and Pask *et al.* (2012).

Supplementary material Chapter 4

Supplementary Figures
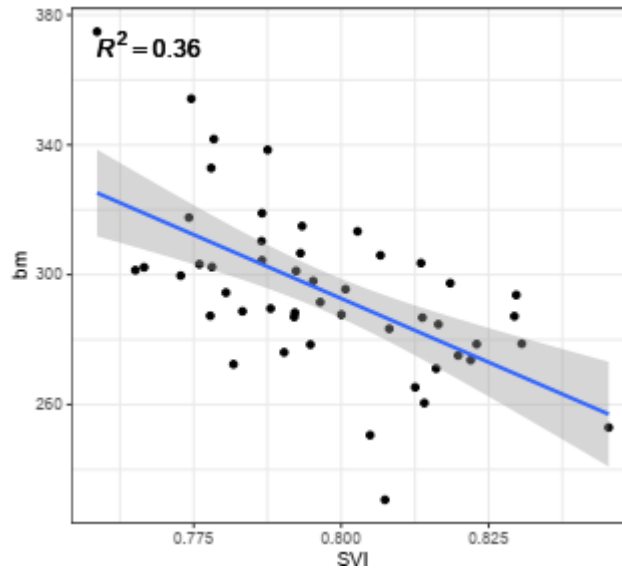
Figure S 7.2 Regression of destructively measured dry matter of the total aerial biomass (bm) on the ratio of reflectance at 1200 nm spectral vegetation index (SVI) for measurements performed on 23 June, 2018. A subset of 49 genotypes contained in the main experiment were used for to select the best-performing SVI-by-date combination. The blue line represents the least squares line of the linear regression of bm *vs.* SVI, the gray area represents the 95% confidence interval of the least squares line.
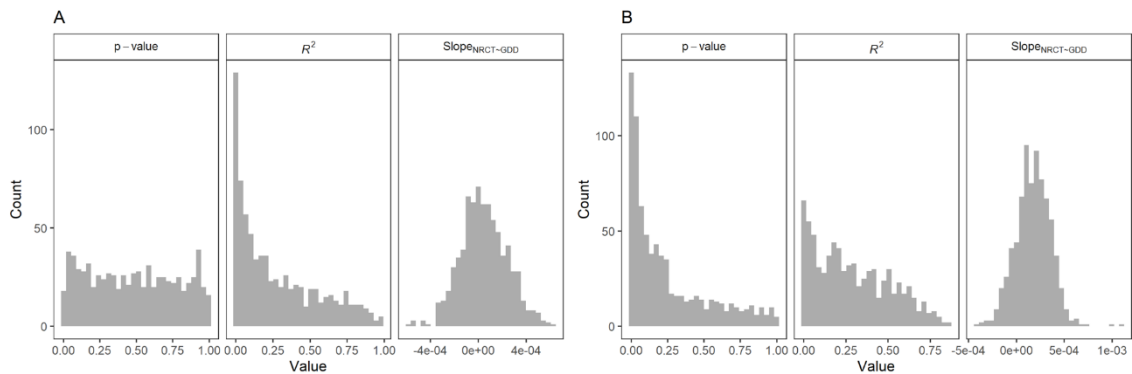
Figure S 7.3 Summary of plot based linear models fitted by ordinary least squares of NRCT vs. thermal time. Distribution of linear model slopes, R^2 and p-values. **(A)** Linear models were fitted to uncorrected NRCT; **(B)** Linear models were fitted to date-wise spatially corrected NRCT.
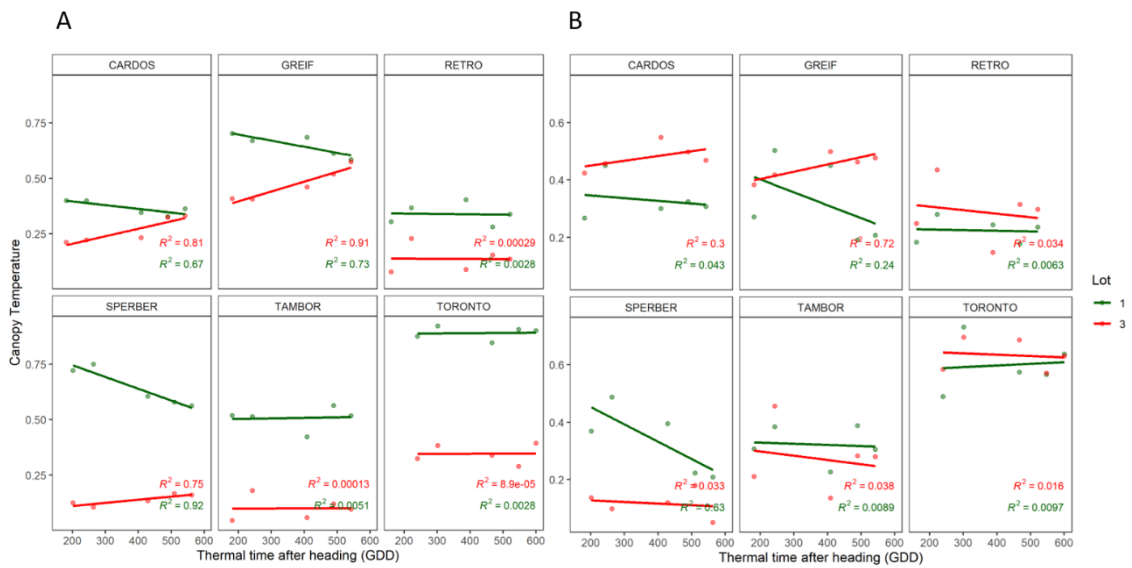
Figure S 7.4 Some examples of linear model fits. Genotypes were selected to have highest or lowest average (across two replicates) R^2 values. **(A)** Trends for uncorrected NRCT. **(B)** Trends for date-wise spatially corrected NRCT.
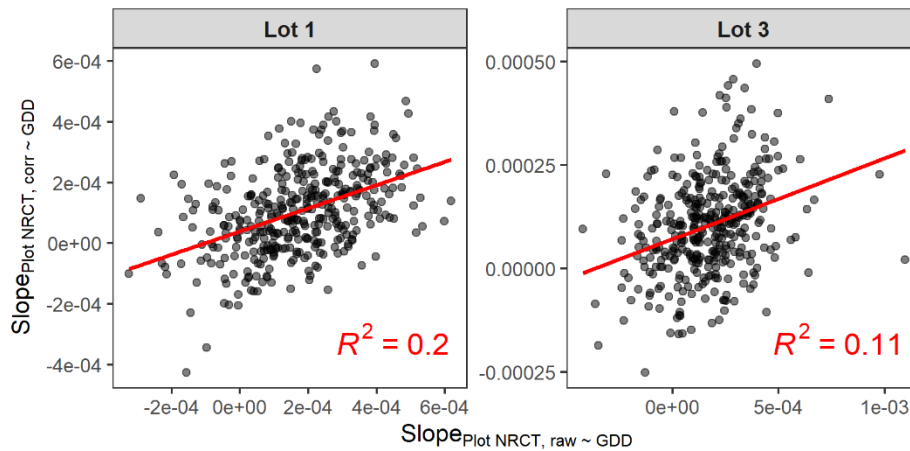
Figure S 7.5 Scatter plots of slopes fitted to date-wise spatially corrected normalized canopy temperature (CT) values (Slope$_{Plot rank, corr}$) versus spatially corrected slopes fitted to raw normalized CT values (Slope$_{Plot rank, raw}$). The red line represents the least squares regression line.
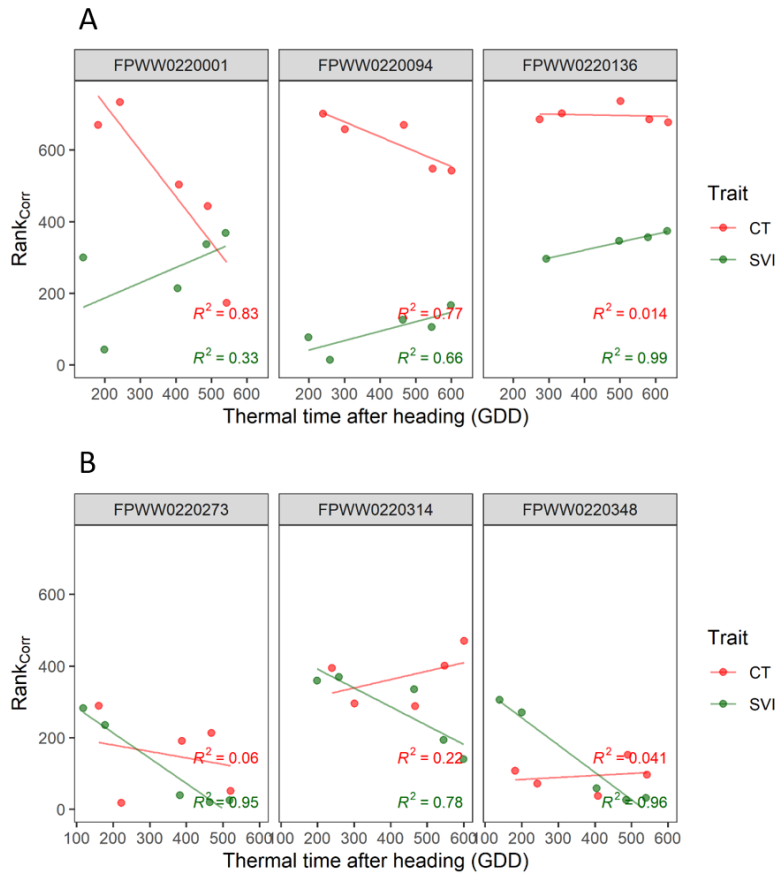
Figure S 7.6 Three extremes to both ends of the distribution shown in Figure 4.6. The NDVI was used as an indicator of canopy greenness. **(A)** Extremes at the upper end of the distribution, the slope$_{SVI}$ − slope$_{CT}$ values range from 1.10 to 1.24; **(B)** extremes at the lower end of the distribution, the slope$_{SVI}$ − slope$_{CT}$ values range from -0.92 to -1.02.
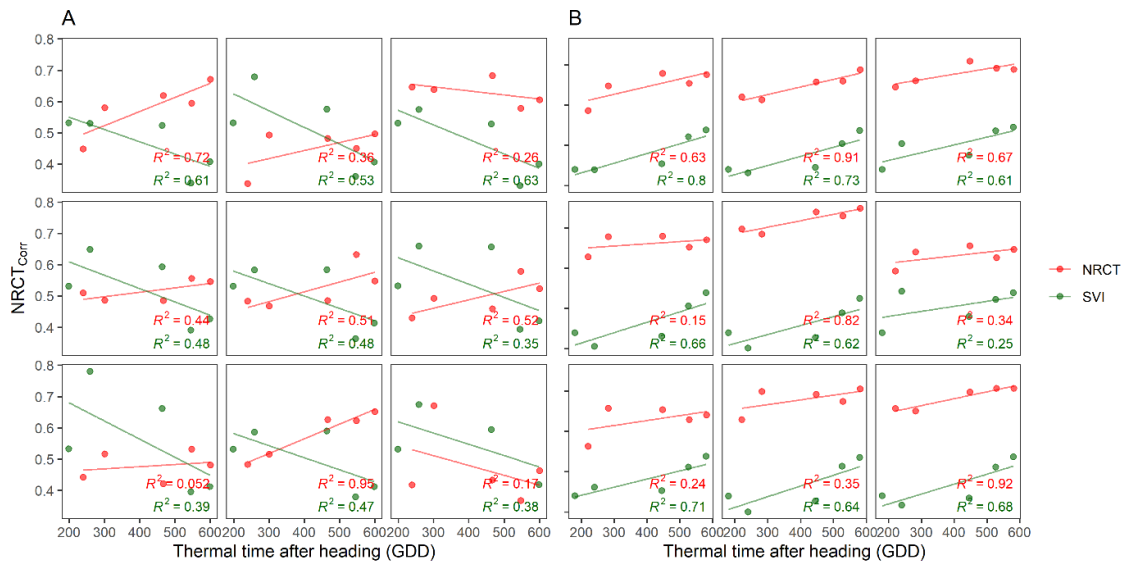
Figure S 7.7 Temporal trends in canopy temperature (CT) and the spectral vegetation index NDVI (SVI) ranks for all plots sown with the check cultivars CH CLARO **(A)** and CH NARA **(B)**. NRCT and NRSVI are date-wise spatially corrected values. Lines are least square regression lines.

133

Supplementary material: Data and Analysis Scripts

Chapter 2

Data sets generated and analyzed in this study are available from the ETH Zürich publications and research data repository (https://www.research-collection.ethz.ch).

Experimental data supporting the conclusions of this article can be downloaded from the following link: https://doi.org/10.3929/ethz-b-000365618.

All analysis scripts required to reproduce the results published in this article are publicly available. An archived version can be retrieved from: https://doi.org/10.5905/ethz-1007-227.

Development repository: https://github.com/andjonas/Andereggetal2019.

Programming language: R. License: GNU General Public License, version 3 (GPL-3.0).

Chapter 3

Data sets generated and analyzed for this study are available from the ETH Zürich publications and research data repository (https://www.research-collection.ethz.ch/)

Experimental data supporting the conclusions of this article can be downloaded from the following link: https://doi.org/10.3929/ethz-b-000370027.

All analysis scripts required to reproduce the results published in this article are publicly available. An archived version can be retrieved from: http://doi.org/10.5905/ethz-1007-228

Development repositories: https://github.com/and-jonas/Andereggetal2019b and https://github.com/and-jonas/stb_placl.

Programming language: R, Python. License: GNU General Public License, version 3 (GPL-3.0).

8 Acknowledgements

9 Curriculum Vitae

Personal Data

Name	Jonas Anderegg
Born	08 June 1989 in Wattwil, Switzerland
Citizen of	Rapperswil BE, Switzerland

Education

03/2016 – 02/2020	Doctorate, Crop Science Group, D-USYS, ETH Zürich, Switzerland
02/2014 – 03/2016	MSc in Agricultural Sciences, ETH Zürich, Switzerland
09/2010 – 02/2014	BSc in Agricultural Sciences, ETH Zürich, Switzerland
08/2004 – 06/2008	Matura, Kantonsschule Wattwil, Switzerland

Teaching experience

Regular courses	Teaching assistant (introduction to spectral reflectance measurements and spectral data analysis) in the MSc course "crop phenotyping", main lecturer PD Dr. Andreas Hund

Service

10/2008 – 03/2009	Basic military training, Wangen an der Aare, Switzerland
01/2010 – 05/2010	Civil service, UMTEC Institut für Umwelt- und Verfahrenstechnik, Rapperswil, Switzerland

Languages

German	Native
English	Good, written and spoken
French	Good, written and spoken
Italian	Good, written and spoken