Phenotyping wheat by combining ADEL-Wheat 4D structure model with proximal remote sensing measurements along the growth cycle

Liu Shouyang

THÈSE

Présentée par

Shouyang Liu

Pour obtenir le titre de

DOCTEUR DE L'UNIVERSITE D'AVIGNON ET DES PAYS DE VAUCLUSE

Spécialité : Télédétection de la Végétation

Phenotyping wheat by combining ADEL-Wheat 4D structure model with proximal remote sensing measurements along the growth cycle

Date de soutenance: 8 December 2016

Composition du jury:

Christelle Gée	INRA Agrosup-Dijon, France	Rapporteur
Jean-philippe Gastellu	CNRS, France	Rapporteur
Tao Cheng	Nanjing Agricultural University, China	Examinateur
Bruno Andrieu	INRA-AgroParisTech, France	Co-directeur de thèse
Benoit de Solan	Arvalis, Avignon, France	Examinateur
Frédéric Baret	INRA-EMMAH, France	Directeur de thèse

Table of content

1. General Introduction

1.1. Background

In the past decades, breeding and advances in cultural practices ensured the food security with a global crop production increase rate of 32 million metric tons per year (**Fig. 1**). However, the population growth and the increase of food, fuel and fiber consumption per capita, requires a production increase rate of 40 million metric tons per year to meet the demand at the global scale (Tester and Langridge 2010). In addition to this huge challenge, agriculture has to limit its environmental footprint while facing often the negative impacts of the climate change (Foley et al. 2011; Furbank and Tester 2011; Reynolds 2010; Tester and Langridge 2010). More efforts should therefore be directed to substantially increase crop yield at a higher annual rate while reducing the environmental cost (Chen et al. 2014; Foley et al. 2011).

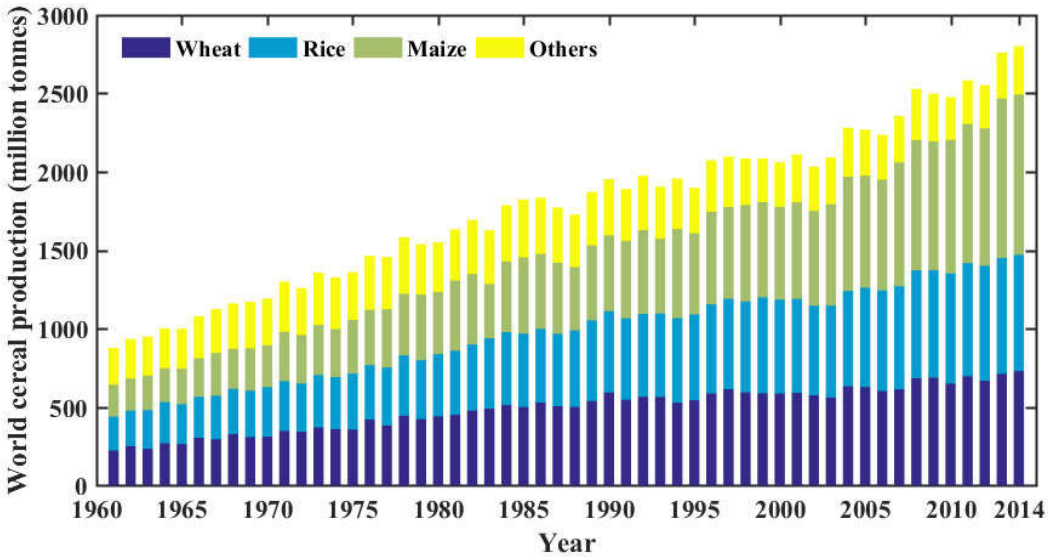

Fig. 1 World production of major crops from 1961 to 2014 (FAOSTAT 2014)

Decision making in crop management and the genetic improvement through breeding require information on the plant behavior under its given growth environment. Methods for monitoring in a quantitative way crop state variables and its functioning are highly desired. These activities are called plant phenotyping, the phenotype being defined as the expression of the genome under a given set of environmental conditions. In the past decades, phenotyping was mostly completed using visual notations, direct manual measurement and destructive measurements including yield or yield related secondary traits. More recently, new methods based on non-invasive proximal or remote sensing were developed to access relevant traits (Cabrera-Bosquet et al. 2012). This was driven by the needs for smart farming that includes precision agriculture and plant breeding for genetic improvement, resulting in significant increase of the yield of major crops (Jackson et al. 1996). However, phenotyping is still laborious, expensive and mainly based on empirical approaches (Prasanna et al. 2013). By contrast, the tremendous technological advances continuously reduce the cost and greatly increase the accuracy and throughput of genotyping, which largely contributed to describe and understand the nature of genotype by genotype (GG) and genotype by environment (GE) interactions (Cobb et al. 2013). The huge mass of genotypic information now available offers the potential to increase the rate of genetic improvement. It opens new avenues for developing more efficient methods based on molecular breeding that will replace the conventional phenotypic selection. However, attempts to implement these new methods such as marker-assisted selection (MAS), need to associate quantitative phenotypic traits of interest. However, this is currently limited by the phenotyping capability in terms of accuracy, efficiency and throughput, which corresponds to the phenotyping bottleneck of plant breeding (Cobb et al. 2013; Furbank and Tester 2011).

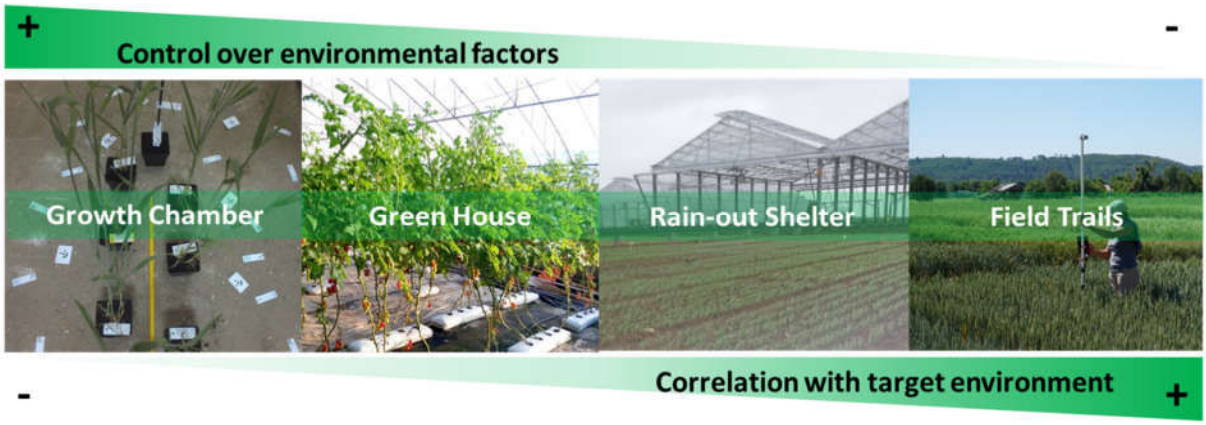

Fig. 2 Continuum of environments for drought tolerance phenotyping.

Progress in sensors, aeronautics, and high-performance computing are paving the way for the development of high-throughput phenotyping techniques to satisfy the needs from molecular biologists and breeders (Araus et al. 2015). Efforts have been continuously invested in developing and establishing phenotyping platforms by both private and public sectors (Prasanna et al. 2013) (**Fig. 2**). For instance, controlled conditions high-throughput phenotyping platforms (HTPPs) are often constructed as fully automated facilities in greenhouse or growth chamber with precise and automatic environmental control combined with noninvasive remote sensing techniques to monitor plant growth and assess the plant performances using automatic data gathering and processing (Ghanem et al. 2015). Although greenhouse HTPPs enable well control the environmental conditions and detailed precise traits to be captured during growth cycle, it requires a large investment in the construction of appropriate infrastructure and professional daily maintenance. However, the phenotypic expressions observed in such artificial growth conditions may not transpose easily to the field conditions (Araus and Cairns 2014a; Ghanem et al. 2015; Prasanna et al. 2013). Therefore, field HTPPs is a mandatory step to complete in order to make sure that the phenotype by environment interactions are pertinently described (Araus and Cairns 2014b; Ghanem et al. 2015).

Fig. 3 Different types of phenotyping systems. From left to right, it corresponds to portable instrument, semi-automatic platform (Phenotypette), automatic field robot (Phenomobile) and unmanned aerial vehicle (UAV).

Four types of vectors are shown in **Fig. 3** for field phenotyping**.** On each of those vectors, a specific set of sensors may be mounted associated with variable throughput: from medium throughput with hand held systems (30-50 plots/h) up to very high throughput (>1000 plots/h) with UAV (Unmanned Aerial Vehicle). In addition to the throughput, the spatial resolution, active or passive nature of the measurements, volume of data generated and complexity of the systems require suitable investment in terms of cost, man power and skills and data management facilities.

1.2. Why phenotyping wheat?

In 2014, wheat, including bread or common wheat and durum wheat, contributed 26.0% of the world cereal production only after maize (36.5 %) and rice (26.5%) (**Fig. 1**). Therein, bread wheat represents more than

90% of the total production. Moreover, wheat is grown on more land area than any other crops (220 million hectares) and has been the basic staple food of major civilizations in Europe, West Asia, and North Africa. It is also the most important protein source in developing countries (20%) and 2nd most important food calorie source (19%) only after rice (FAOSTAT 2014; Lumpkin 2011; Reynolds 2010). However, investment dedicated to wheat research did not match its reputation for long-term. For instance, the funding on maize research is four times greater than that in wheat research. This significantly contributes to the fact that the increase rate of maize yield appears much faster than wheat (**Fig. 1**). The situation we are facing is that wheat production did not meet the demand in 6 out of 10 years before 2011 (Tester and Langridge 2010; Wheat-Initiative 2011). Even worse, climate change is predicted to reduce wheat production in developing countries by 20% to 30% (Lumpkin 2011). Nonetheless, annual wheat yield increase must boost from the 1% to at least 1.7% to meet the demand by 2050 (Wheat-Initiative 2011).

To accelerate the wheat yield increase, several large national research programs associating public and private partners were launched during the recent years in France (BREEDWHEAT), UK (WISP), Canada (CTAG) and USA (TriticeaedCAP). At the international level, Wheat Initiative (Wheat-Initiative 2011) was created in 2011 following endorsement from the G20 Agriculture Ministries to coordinate global wheat research. Efforts will focus on coupling the wheat genetic resources with the development of efficient phenotyping methodologies to design new genome-based breeding strategies and create robust new varieties in 15 years' time (Wheat-Initiative 2011). An international network of phenotyping platforms is one of the important actions to be implemented. Additionally in the survey conducted by International Plant Phenotyping Network (IPPN) (IPPN 2016),reveals that:

- **55% researchers are not satisfied with the current phenotyping capacity.**
- **Wheat is the top plant species of interest.**
- **The largest challenge in plant phenotyping is field phenotyping.**

In this context, this work will focus on wheat phenotyping in field conditions.

1.3. Which traits to phenotype?

During the past decades, the yield increase in wheat was mainly attributed to the gradual replacement of traditional tall cultivars by dwarf and fertilizer-responsive varieties, referring to the 'Green Revolution' (Monneveux et al. 2012). Reducing height increased the proportion of carbon allocated to the grain and increased the harvest index (HI). It simultaneously reduced the risk of yield loss caused by lodging. However, the HI in most modern cultivars seems to be close to its biological maximum, i.e. 60 %. Genetic gain in yield potential have been far lower than what is required to keep pace with the increase in demand (Slafer and Araus 2007). Consequently any significant improvement in yield potential beyond the common average present rate of 0.5–1.0% per year needs to integrate advances in molecular biology and crop physiology (Slafer and Araus 2007). As noted above, molecular breeding might greatly contribute to the yield increase in the future (Reynolds et al. 2012). However, most breeders still work with the whole plant and mostly under field conditions (Reynolds et al. 2012; Slafer and Araus 2007).

Table 1 Main traits used to improve drought tolerance in wheat (Adapted from (Reynolds et al. 2001a)). I/D refers respectively to Instantaneous (state variable) or Dynamic measurements required

Traits		Physiological process associated	I/D	Reference
Morphological traits	Large seed size	Emergence, early ground cover and initial biomass	I	(Mian and Nafziger 1994)
	Long coleoptiles	Emergence from deep sowing		(Radford 1987)
	Leaf thickness and posture	Lower transpiration rate and reduced photo-inhibition		(Reynolds et al. 2000)
	Leaf rolling			(Monasterio et al. 2001)
	Specific leaf dry weight	Thinner, wider leaves, early ground cover		(Merah et al. 1999)
	Leaf color	Delayed senescence, maintenance of photosynthesis	D	(Araus et al. 1997)
	Long and thick stem internodes	Storage of carbon products	D	(Loss and Siddique 1994)

	Number of grain per spike	Spike sterility		(Hafsi et al. 2007)
	Tiller survival	Survival and recovery	D	(Loss and Siddique 1994)
	Early ground cover	Increase of water use efficiency (WUE) and radiation use efficiency (RUE)		(El Hafid et al. 1998; Richards 1996)
	Vegetation indices	Green biomass		(Royo et al. 2003)
Physiological traits	Accumulation of ABA	Reduced stomatal conductance and cell division	I	(Innes et al. 1984)
	Solute concentration in cells	Osmotic adjustment		(Morgan 1988)
	Carbon isotope discrimination	Stomatal conductance, extraction of water from soil		(Monneveux et al. 2005)
	Ash content			(Misra et al. 2006)
	Stomatal conductance	Extraction of water from soil		(Farquhar and Sharkey 1982)
	Glume pubescence	Lower transpiration rate and reduced photo-inhibition		(Trethowan et al. 1998)
	Leaf waxiness			(Richards 1996)
	Leaf pubescence			(Richards 1996)
	Fructanes in stem	Storage of carbon products		(Rawson and Evans 1971)
	Spike photosynthetic capacity	Grain filling		(Evans et al. 1972)
	Canopy temperature depression	Stomatal conductance, extraction of water from soil	D	(Reynolds et al. 2000)
	Delayed senescence	Higher RUE		(Hafsi et al. 2007)
	Earliness	Drought escape		(Monneveux et al. 2005)

Knowledge on the plant physiology allows to identify relatively simple traits that are related to yield under a wide range of field conditions (Slafer and Araus 2007). These traits are considered as secondary as compared to yield that remains the primary trait during selection despite its low heritability due to its multifactorial nature. The exploitation of secondary traits can considerably increase the efficiency of selection (Edmeades 1996; Reynolds et al. 2001a) particularly under stress conditions that prevail in most wheat-growing areas worldwide (Reynolds et al. 2012; Slafer and Araus 2007). **Table 1** lists the candidate traits of interests to improve drought tolerance in wheat over different target environments. Each trait may be associated with a particular physiological process involved in drought tolerance. However, not all these traits are easily and directly accessible from noninvasive high-throughput observations as we will see in the next section.

1.4. How to access to traits of interests?

Non-invasive methods are one of the prerequisites to achieve high-throughput in phenotyping. Imaging techniques were originally developed and extensively applied in the remote sensing domain. Imaging techniques can be grouped into four categories (**Fig. 4**) (Li et al. 2014b) depending on the spectral domain and spectral sampling, the active/passive mode of measurements, emission/reflection process:

- **RGB imaging,**
- **Multispectral/hyperspectral imaging,**
- **fluorescence imaging,**
- **thermal imaging,**
- **LiDAR.**

Imaging techniques start to be widely used for phenotyping applications to sense crops from close range (Li et al. 2014a; Minervini et al. 2015). Proximal imaging observations record the light emerging from the canopy after multiple interactions with tissues of the target canopy. The interaction process involves the configuration of the incident light, optical properties of leaves and also the spatial distribution of vegetation elements, i.e. the canopy structure. All these factors are potentially accessible from the imaging observations (Hodáňová 1985). However the success of this approach depends on our understanding of the factors affecting plant radiative properties and our ability to interpret radiance data from a wide range of ecosystems

(Ollinger 2011). Efforts have been invested to relate the spatial/spectral/directional/temporal variability of the signal measured to physiological and morphological traits (Asner 1998; Ollinger 2011).

Fig. 4 Examples of different imaging techniques for phenotyping. From left to right, they correspond to RGB camera (Sony Alpha-5000), chlorophyll fluorometers (WALZ, PAM-2500), thermal camera (DJI Zenmuse-XT), multispectral camera (Hi-Phen, Airphen-V3) and LiDAR (SICK, LM-400).

1.4.1. Physiological traits

The response of plant leaves to incident radiation, including reflection, transmission and absorption, depends on the pigments composition, surface and volume structure characteristics of the leaves (**Fig. 5)** (Jacquemoud and Baret 1990; Woolley 1971). In the ultraviolet (250 to 400 nm, UV) and photosynthetically active radiation (400 to 700 nm, PAR), the absorption of radiation is a function of changes in the spin and angular momentum of electrons, transitions between orbital states of electrons. Beyond 700 nm in the middle infrared domain vibrational-rotational modes within the polyatomic molecules explains the absorption radiation. Meanwhile when the energy of a given wavelength is absorbed, it is partly dissipated by light emission at longer wavelengths within a very short time, i.e. the fluorescence emission (Tremblay et al. 2012).

Pigments absorb light at specific wavelengths in the range of PAR for photosynthesis. They mainly include two categories: Chlorophyll *a* and *b* and accessory pigments, including carotenoids, anthocyanins and Xanthophylls. Each type of pigment can be identified by the specific pattern of wavelengths it absorbs in the visible domain, defining its absorption spectrum. Specifically, Chlorophyll *a* and *b* absorbs light in the blue-violet and red-blue spectral domains, respectively. However, none of them absorb strongly green light. As a consequence, Chlorophyll appear green. Carotenoids absorb light in the blue-green and violet region and appear red, orange or yellow to us. Investigations into pigment reflectance properties have produced a wide variety of vegetation indices, combining narrow bandwidth in the visible portion of the spectrum (Ollinger 2011). The characterization of plant pigments using hyperspectral remote sensing has been extensively reviewed by (Blackburn 2007). Chlorophylls content can be estimated by model inversion or vegetation indices, while accessory pigments are currently mainly accessible with vegetation indices (**Table 2**). Furthermore, the canopy photosynthesis status could be assessed with vegetation or fluorescence indices and model inversion approach using multi-spectral or hyperspectral observations (**Table 2**).

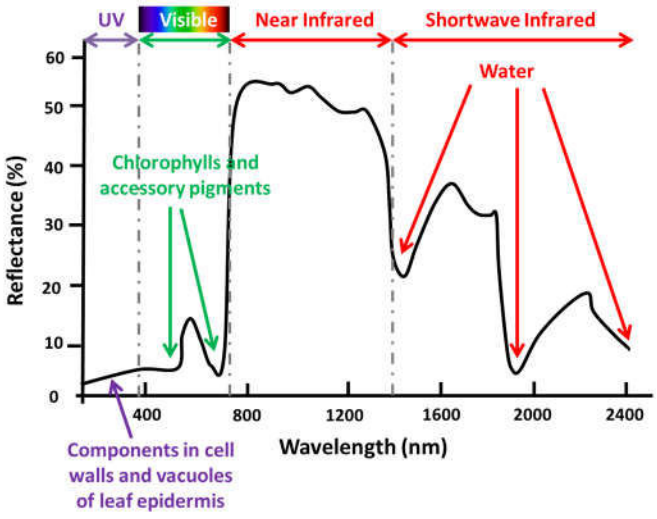

Fig. 5 Typical reflectance spectra of crop

Non-pigment leaf constituents include nitrogen mostly in the form of proteins, then lignin and cellulose, sugar or starch. In order to optimize fertilization, particular efforts have been dedicated to estimate foliar nitrogen concentration using remote sensing techniques (Ollinger 2011). Vegetation indices empirically related to nitrogen content are mainly the method currently proposed (Hansen and Schjoerring 2003b; Xue et al. 2004). However, even though most of the foliar nitrogen is associated with pigments and the proteins involved in photosynthesis, nitrogen itself is a relatively small component of leaf mass, consequently having weak impact on the reflectance spectra. Therefore, there is no strong physical basis regarding the correlation between nitrogen and vegetation indices. It deeply relies on the relationship between foliar nitrogen and photosynthetic pigments or more indirectly with canopy structure. It is still challenging to achieve reliable nitrogen estimation even though some recent work shows that fluorescence-based indices seems to be more sensitive to plant nitrogen status (Samborski et al. 2009; Tremblay et al. 2012; Zarco-Tejada 2005).

Table 2 Physiological traits accessible from imaging phenotyping techniques

Target trait		Method or Index	Equation or Model	Sensor	Reference
Leaf pigments	Chlorophyll a, b	Pigment specific normalized difference (PSND)	$\frac{R_{800}-R_{675}}{R_{800}-R_{675}}, \frac{R_{800}-R_{650}}{R_{800}-R_{650}}$	Hyperspectral /Multispectral camera	(Blackburn 1998)
		Model inversion	Prosail model		(Jacquemoud et al. 2009)
	Carotenoids	Structure-insensitive pigment index (SIPI)	$\frac{R_{800}-R_{445}}{R_{800}-R_{680}}$		(Sims and Gamon 2002)
	Anthocyanin	Red/Green (RGI)	$\frac{\sum_{i=600}^{700} R_i}{\sum_{i=500}^{600} R_i}$		(Gamon and Surfus 1999)
	Xanthophyll cycle	Photochemical reflectance index (PRI)	$\frac{R_{531}-R_{570}}{R_{531}+R_{570}}$		(Gamon et al. 1997)
Water status	Water Content Equivalent Water Thickness	Normalized different water index (NDWI)	$\frac{R_{860}-R_{1240}}{R_{860}+R_{1240}}$		(Gao 1996; Jackson et al. 2004)
		Model inversion	Prosail model		(Jacquemoud et al. 2009)
	Canopy temperature	Model inversion	SCOPE model		(Van der Tol et al.
		Crop water stress index (CWSI)	--	Thermal and	(Jackson et al. 1981;
Photosynthesis status	Photosynthetic efficiency	Photochemical reflectance index (PRI)	$\frac{R_{531}-R_{570}}{R_{531}+R_{570}}$	Hyperspectral/ Multispectral camera	(Gamon et al. 1992)
	Chlorophyll fluorescence	Fluorescence indices	--	Hyperspectral/	(Baker and
		Model inversion	FluorMOD	Hyperspectral	(Zarco-Tejada et al. 2012)

1.4.2. Morphological traits

Incident light interacts with the canopy structure at different scales. At the plant scale, the shape size and orientation of organs are key traits of interests. A number of processing tools have been developed to facilitate the extraction of traits from two dimensional RGB imagery as reviewed by Cobb et al. (2013). However, most of these methods needs to collect samples using destructive methods and to interactively process the data (**Table 3**). Moreover, three dimensional (3D) reconstruction techniques have been employed to depict explicitly and non-destructively the 3D structure. Morphological traits characterized statistically at the canopy scale can be measured on the reconstructed 3D scene. Both passive and active sensor based techniques are available to achieve the detailed 3D description of the canopy structure. However, the passive techniques are vulnerable to variations in the illumination conditions, thereby limiting their use under field grown experiments. Conversely, passive techniques including LiDAR are fully independent from the illumination conditions, therefore bringing increased repeatability of the measurements.

The dynamics of canopy height, leaf area index (LAI) or the fraction of photosynthetically actice radiation absorbed (FAPAR) or intercepted (FIPAR) by the canopy are critical traits to characterize the evolution of canopy structure (**Table 3**). Note that even considerable efforts have been made to estimate biomass using

different optical observations, as lack of sound physical basis, a reliable and accurate estimation of canopy biomass is still challenging.

Table 3 Morphological traits accessible from imaging phenotyping techniques. D and N refer to destructive and non-destructive, respectively. I and F refer to indoor or field condition.

Trait class		Method	Sensor	Note	Reference
Individual	Leaf shape	Computer vision	RGB camera	D, I	(Dornbusch and Andrieu 2010)
		3D reconstruction		N, I	(Duan et al. 2016)
	Leaf curvature				(Cabrera‐Bosquet et al. 2016)
	Reproductive organs	Computer vision		D, I	(Al-Tam et al. 2013)
Canopy	Height	3D reconstruction		N, F	(Bendig 2015)
		Point cloud	LiDAR		(Zhang and Grift 2012)
	Biomass	Empirical model	RGB camera		(Hoyos-Villegas et al. 2014)
		3D reconstruction		N, I	(Bendig 2015)
		Vegetation indices	Multispectral camera	N, F	(Hansen and Schjoerring 2003a)
		Point cloud	LiDAR		(Ehlert et al. 2009)
	Leaf area index	Gap fraction model	RGB camera		(Baret et al. 2010b)
		Radiation transfer model inversion	Multispectral camera		(Verger et al. 2014a)
		Gap fraction model inversion	LiDAR		(Zhao et al. 2015)
		Two scale model			(Béland et al. 2011; Beland et al. 2014)
	Crop density	Computer vision	RGB camera		(Shrestha and Steward 2003, 2005)
			LiDAR		(Shi et al. 2015; Shi et al. 2013)

1.5. Study objectives and outlook

Nowadays the importance of phenotyping is no more questioned. As noted above, considerable resources from both private and public sectors have been invested to construct phenotyping infrastructure. After the explosion of infrastructure investment during the last decade, phenotyping observations starts to boom. Although phenotyping in controlled conditions allows tackling specific processes, field conditions appear critical for plant phenotyping since it allows evaluating the genotypes under climate, soil and management practices that will be representative of those experienced by the crop within regular farmer's fields.

Field phenotyping techniques are still in an early development phase and cannot yet satisfy the need of breeders in terms of the throughput and the variety of traits accessible. Phenotyping under field conditions are subjected to the large uncontrolled variability of climate conditions including illumination, rain and wind. This poses a big challenge to the repeatability, reliability and robustness of traits characterization. More importantly the multi-disciplinary nature of phenotyping requires skills in mechatronics, physics of the signal, data management, image processing, ecophysiology and agronomy. Although systems are now capable to provide pertinent measurements as demonstrated in tables 2 and 3, the transformation of the sensor measurements into traits useful for the breeders is mainly based on simple, generally empirical approaches. Further, the dynamics that would provide access to the characterization of plant functioning is very rarely an explicit target. Progress in data interpretation seems still lagging far behind the needs: the identified bottleneck of genetic progress is now translated into a phenotyping data interpretation bottleneck.

The objective of this work is to contribute to the advancement of the interpretation of field phenotyping measurements with emphasis on the exploitation of the repeated observations along the growth cycle to access functional traits. Further, the use of models, either statistical, physically or eco-physiologically based models, will be introduced to exploit as much available information as possible.

The study is mainly decomposed into three main chapters that are sketched in **Fig. 6**:

- ***Plant density and sowing pattern estimation*** from high resolution imagery. These first observations will allow initiating the interpretation for the following ones. As a matter of facts, the knowledge of plant density and spatial distribution of plants impact strongly the fate of the crop along its whole growth cycle.
- ***Estimating wheat Green area index (GAI) from ground-based LiDAR*** measurement through 3D ADEL-Wheat model; In this chapter, we will demonstrate how the use of a realistic 3D model of canopy architecture, the Adel-Wheat model, will improve the instantaneous estimates of GAI as compared to more classical methods based on very simple assumptions on canopy architecture. This will be achieved at the stage when GAI reaches its maximum value, which is known to pose specific problems related to the limited sensitivity of the signal measured to changes in GAI.
- ***Deriving ADEL-Wheat parameters from the monitoring of the green fractions up to stem elongation.*** The repeated RGB camera measurements provide estimates of the green fraction that are used to tune the ADEL-Wheat parameters sequentially. Focus will be put on the identifiability of the model parameters. Parameters that can be retrieved with a reasonable accuracy may be considered as traits, or used within a second step to derive higher order variables (such as GAI dynamics) of interest for the breeder.

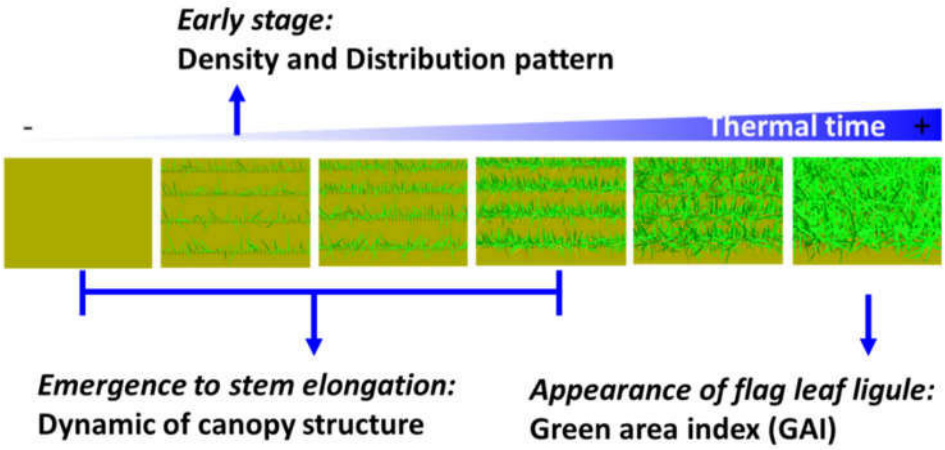

Fig.6 Diagram of three main chapters in this thesis

Reference

Al-Tam, F., Adam, H., Anjos, A.d., Lorieux, M., Larmande, P., Ghesquière, A., Jouannic, S., & Shahbazkia, H.R. (2013). P-TRAP: a Panicle Trait Phenotyping tool. *BMC Plant Biology, 13*, 122

Araus, J., Amaro, T., Zuhair, Y., & Nachit, M. (1997). Effect of leaf structure and water status on carbon isotope discrimination in field-grown durum wheat. *Plant, Cell & Environment, 20*, 1484-1494

Araus, J.L., & Cairns, J.E. (2014a). Field high-throughput phenotyping: the new crop breeding frontier. *TRENDS in Plant Science, 19*, 52-61

Araus, J.L., & Cairns, J.E. (2014b). Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Science, 19*, 52-61

Araus, J.L., Elazab, A., Vergara, O., Cabrera-Bosquet, L., Serret, M.D., Zaman-Allah, M., & Cairns, J.E. (2015). New Technologies for Phenotyping. *Phenomics* (pp. 1-14): Springer

Asner, G.P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment, 64*, 234-253

Baker, N.R., & Rosenqvist, E. (2004). Applications of chlorophyll fluorescence can improve crop production strategies: an examination of future possibilities. *Journal of Experimental Botany, 55*, 1607-1621

Baret, F., de Solan, B., Lopez-Lozano, R., Ma, K., & Weiss, M. (2010). GAI estimates of row crops from downward looking digital photos taken perpendicular to rows at 57.5° zenith angle: Theoretical considerations based on 3D architecture models and application to wheat crops. *Agricultural and Forest Meteorology, 150*, 1393-1401

Beland, M., Baldocchi, D.D., Widlowski, J.L., Fournier, R.A., & Verstraete, M.M. (2014). On seeing the wood from the leaves and the role of voxel size in determining leaf area distribution of forests with terrestrial LiDAR. *Agricultural and forest meteorology, 184*, 82-97

Béland, M., Widlowski, J.-L., Fournier, R.A., Côté, J.-F., & Verstraete, M.M. (2011). Estimating leaf area distribution in savanna trees from terrestrial LiDAR measurements. *Agricultural and forest meteorology, 151*, 1252-1266

Bendig, J.V. (2015). Unmanned aerial vehicles (UAVs) for multi-temporal crop surface modelling. A new method for plant height and biomass estimation based on RGB-imaging. In: Universität zu Köln

Blackburn, G.A. (1998). Quantifying chlorophylls and caroteniods at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sensing of Environment, 66*, 273-285

Blackburn, G.A. (2007). Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany, 58*, 855-867

Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M.D., & Luis Araus, J. (2012). High-throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding ConvergeF: High-throughput Phenotyping and Genomic Selection. *Journal of Integrative Plant Biology, 54*, 312-320

Cabrera-Bosquet, L., Fournier, C., Brichet, N., Welcker, C., Suard, B., & Tardieu, F. (2016). High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytologist*

Chen, X., Cui, Z., Fan, M., Vitousek, P., Zhao, M., Ma, W., Wang, Z., Zhang, W., Yan, X., Yang, J., Deng, X., Gao, Q., Zhang, Q., Guo, S., Ren, J., Li, S., Ye, Y., Wang, Z., Huang, J., Tang, Q., Sun, Y., Peng, X., Zhang, J., He, M., Zhu, Y., Xue, J., Wang, G., Wu, L., An, N., Wu, L., Ma, L., Zhang, W., & Zhang, F. (2014). Producing more grain with lower environmental costs. *Nature, 514*, 486-489

Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics, 126*, 867-887

Dornbusch, T., & Andrieu, B. (2010). Lamina2Shape—An image processing tool for an explicit description of lamina shape tested on winter wheat (Triticum aestivum L.). *Computers and Electronics in Agriculture, 70*, 217-224

Duan, T., Chapman, S., Holland, E., Rebetzke, G., Guo, Y., & Zheng, B. (2016). Dynamic quantification of canopy structure to characterize early plant vigour in wheat genotypes. *Journal of experimental botany, 67*, 4523-4534

Edmeades, G.O. (1996). Developing Drought and Low N-tolerant Maize: Proceedings of a Symposium, March 25-29, 1996, CIMMYT, El Batán, Mexico. CIMMYT

Ehlert, D., Adamek, R., & Horn, H.-J. (2009). Laser rangefinder-based measuring of crop biomass under field conditions. *Precision Agriculture, 10*, 395-408

El Hafid, R., Smith, D.H., Karrou, M., & Samir, K. (1998). Morphological attributes associated with early-season drought tolerance in spring durum wheat in a Mediterranean environment. *Euphytica, 101*, 273-282

Evans, L., Bingham, J., Jackson, P., & SUTHERLAND, J. (1972). Effect of awns and drought on the supply of photosynthate and its distribution within wheat ears. *Annals of Applied Biology, 70*, 67-76

FAOSTAT (2014). In. http://faostat.fao.org/: Food and Agriculture Organization of the United Nations

Farquhar, G.D., & Sharkey, T.D. (1982). Stomatal conductance and photosynthesis. *Annual review of plant physiology, 33*, 317-345

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., & Zaks, D.P.M. (2011). Solutions for a cultivated planet. *Nature, 478*, 337-342

Furbank, R.T., & Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science, 16*, 635-644

Gamon, J., Penuelas, J., & Field, C. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment, 41*, 35-44

Gamon, J., Serrano, L., & Surfus, J. (1997). The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia, 112*, 492-501

Gamon, J., & Surfus, J. (1999). Assessing leaf pigment content and activity with a reflectometer. *New Phytologist, 143*, 105-117

Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*, 257-266

Ghanem, M.E., Marrou, H., & Sinclair, T.R. (2015). Physiological phenotyping of plants for crop improvement. *Trends in Plant Science, 20*, 139-144

Hafsi, M., Akhter, J., & Monneveux, P. (2007). Leaf senescence and carbon isotope discrimination in durum wheat (Triticum durum Desf.) under severe drought conditions. *Cereal Research Communications, 35*, 71-80

Hansen, P., & Schjoerring, J. (2003a). 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

Hansen, P.M., & Schjoerring, J.K. (2003b). 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

Hodáňová, D. (1985). Leaf optical properties. *Photosynthesis during leaf development* (pp. 107-127): Springer

Hoyos-Villegas, V., Houx, J., Singh, S., & Fritschi, F. (2014). Ground-based digital imaging as a tool to assess soybean growth and yield. *Crop Science, 54*, 1756-1768

Innes, P., Blackwell, R., & Quarrie, S. (1984). Some effects of genetic variation in drought-induced abscisic acid accumulation on the yield and water use of spring wheat. *The Journal of Agricultural Science, 102*, 341-351

IPPN (2016). Plant Phenotyping Survey 2016. In: IPPN

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

Jackson, R.D., Idso, S., Reginato, R., & Pinter, P. (1981). Canopy temperature as a crop water stress indicator. *Water Resources Research, 17*, 1133-1138

Jackson, T.J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., Doriaswamy, P., & Hunt, E.R. (2004). Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. *Remote Sensing of Environment, 92*, 475-482

Jacquemoud, S., & Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment, 34*, 75-91

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., & Ustin, S.L. (2009). PROSPECT+ SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment, 113*, S56-S66

Leinonen, I., & Jones, H.G. (2004). Combining thermal and visible imagery for estimating canopy temperature and identifying plant stress. *Journal of experimental botany, 55*, 1423-1431

Li, L., Zhang, Q., & Huang, D. (2014a). A review of imaging techniques for plant phenotyping. *Sensors, 14*, 20078-20111

Li, L., Zhang, Q., & Huang, D. (2014b). A Review of Imaging Techniques for Plant Phenotyping. *Sensors, 14*, 20078-20111

Loss, S.P., & Siddique, K. (1994). Morphological and physiological traits associated with wheat yield increases in Mediterranean environments. *Advances in agronomy, 52*, 229-276

Lumpkin, T.A. (2011). Wheat-global alliance for improving food security and the livelihoods of the resource-poor in the developing world. *National Agricultural Research Institute*

Merah, O., Deléens, E., & Monneveux, P. (1999). Grain yield, carbon isotope discrimination, mineral and silicon content in durum wheat under different precipitation regimes. *Physiologia Plantarum, 107*, 387-394

Mian, M.A.R., & Nafziger, E.D. (1994). Seed Size and Water Potential Effects on Germination and Seedling Growth of Winter Wheat. *Crop Science, 34*, 169-171

Minervini, M., Scharr, H., & Tsaftaris, S.A. (2015). Image Analysis: The New Bottleneck in Plant Phenotyping [Applications Corner]. *IEEE Signal Processing Magazine, 32*, 126-131

Misra, S., Randive, R., Rao, V., Sheshshayee, M., Serraj, R., & Monneveux, P. (2006). Relationship between carbon isotope discrimination, ash content and grain yield in wheat in the Peninsular Zone of India. *Journal of Agronomy and Crop Science, 192*, 352-362

Monasterio, J.O., Reynolds, M., & McNab, A. (2001). *Application of physiology in wheat breeding*. Cimmyt

Monneveux, P., Jing, R., & Misra, S.C. (2012). Phenotyping for drought adaptation in wheat using physiological traits

Monneveux, P., Reynolds, M.P., Trethowan, R., González-Santoyo, H., Peña, R.J., & Zapata, F. (2005). Relationship between grain yield and carbon isotope discrimination in bread wheat under four water regimes. *European Journal of Agronomy, 22*, 231-242

Morgan, J. (1988). The use of coleoptile responses to water stress to differentiate wheat genotypes for osmoregulation, growth and yield. *Annals of Botany, 62*, 193-198

Ollinger, S. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist, 189*, 375-394

Prasanna, B.M., Araus, J.L., Crossa, J., Cairns, J.E., Palacios, N., Das, B., & Magorokosho, C. (2013). High-Throughput and Precision Phenotyping for Cereal Breeding Programs. In P.K. Gupta, & R.K. Varshney (Eds.), *Cereal Genomics II* (pp. 341-374). Dordrecht: Springer Netherlands

Radford, B. (1987). Effect of constant and fluctuating temperature regimes and seed source on the coleoptile length of tall and semidwarf wheats. *Animal Production Science, 27*, 113-117

Rawson, H., & Evans, L. (1971). The contribution of stem reserves to grain development in a range of wheat cultivars of different height. *Crop and Pasture Science, 22*, 851-863

Reynolds, M., Calderini, D., Condon, A., & Rajaram, S. (2001). Physiological basis of yield gains in wheat associated with the LR19 translocation from Agropyron elongatum. *Wheat in a Global Environment* (pp. 345-351): Springer

Reynolds, M., Pask, A., & Mullan, D. (2012). Physiological breeding I: interdisciplinary approaches to improve crop adaptation. CIMMYT

Reynolds, M., Skovmand, B., Trethowan, R., & Pfeiffer, W. (2000). Evaluating a conceptual model for drought tolerance. *Molecular approaches for genetic improvement of cereals for stable production in water-limited environments*, 49-53

Reynolds, M.P. (2010). Climate Change and Crop Production. CABI

Richards, R. (1996). Defining selection criteria to improve yield under drought. *Plant growth regulation, 20*, 157-166

Royo, C., Aparicio, N., Villegas, D., Casadesus, J., Monneveux, P., & Araus, J. (2003). Usefulness of spectral reflectance indices as durum wheat yield predictors under contrasting Mediterranean conditions. *International Journal of Remote Sensing, 24*, 4403-4419

Samborski, S.M., Tremblay, N., & Fallon, E. (2009). Strategies to make use of plant sensors-based diagnostic information for nitrogen recommendations. *Agronomy Journal, 101*, 800-816

Shi, Y., Wang, N., Taylor, R., & Raun, W. (2015). Improvement of a ground-LiDAR-based corn plant population and spacing measurement system. *Computers and Electronics in Agriculture, 112*, 92-101

Shi, Y., Wang, N., Taylor, R.K., Raun, W.R., & Hardin, J.A. (2013). Automatic corn plant location and spacing measurement using laser line-scan technique. *Precision Agriculture, 14*, 478-494

Shrestha, D.S., & Steward, B.L. (2003). Automatic corn plant population measurement using machine vision. *Transactions of the ASAE, 46*, 559

Shrestha, D.S., & Steward, B.L. (2005). Shape and size analysis of corn plant canopies for plant population and spacing sensing. *Applied engineering in agriculture, 21*, 295-303

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

Slafer, G., & Araus, J. (2007). Physiological traits for improving wheat yield under a wide range of conditions. *Scale and Complexity in Plant Systems Research: Gene-Plant-Crop Relations, 21*, 147-156

Tester, M., & Langridge, P. (2010). Breeding Technologies to Increase Crop Production in a Changing World. *Science, 327*, 818-822

Tremblay, N., Wang, Z., & Cerovic, Z.G. (2012). Sensing crop nitrogen status with fluorescence indicators. A review. *Agronomy for sustainable development, 32*, 451-464

Trethowan, R., Reynolds, M., Skovmand, B., & van Ginkel, M. (1998). The effect of glume pubescence on floret temperature in wheat. In, *Agronomy Abstracts* (p. 161)

Van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., & Su, Z. (2009). An integrated model of soil-canopy spectral radiance observations, photosynthesis, fluorescence, temperature and energy balance. *Biogeosciences Disc, 6*, 6025-6075

Verger, A., Vigneau, N., Chéron, C., Gilliot, J.-M., Comar, A., & Baret, F. (2014). Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sensing of Environment*

Wheat-Initiative (2011). Document submitted to the G20 Agricultural Minsters. In

Woolley, J.T. (1971). Reflectance and transmittance of light by leaves. *Plant Physiology, 47*, 656-662

Xue, L., Cao, W., Luo, W., Dai, T., & Zhu, Y. (2004). Monitoring Leaf Nitrogen Status in Rice with Canopy Spectral Reflectance. *Agronomy Journal, 96*, 135-142

Zarco-Tejada, P. (2005). Development of a Vegetation Fluorescence Canopy Model. *Paris, ESA Scientific and Technical Publications Branch, ESTEC*

Zarco-Tejada, P.J., González-Dugo, V., & Berni, J.A. (2012). Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. *Remote Sensing of Environment, 117*, 322-337

Zhang, L., & Grift, T.E. (2012). A LIDAR-based crop height measurement system for Miscanthus giganteus. *Computers and Electronics in Agriculture, 85*, 70-76

Zhao, K., García, M., Liu, S., Guo, Q., Chen, G., Zhang, X., Zhou, Y., & Meng, X. (2015). Terrestrial lidar remote sensing of forests: Maximum likelihood estimates of canopy profile, leaf area index, and leaf angle distribution. *Agricultural and forest meteorology, 209–210*, 100-113

2. Plant density and sowing pattern estimation from high resolution imagery

Plant density along with the distribution pattern of individual plants on the row is required to describe the competition between neighboring plants. It allows also initializing the simulations of the 3D ADEL-Wheat model that will be used in the following chapters. The objectives of this chapter are to develop a high-throughput method able to provide an estimation of the plant density and the sowing pattern. This chapter is split therefore into 3 main sections that correspond to journal articles:

- ***Article 1: Estimation of plant density from high resolution RGB imagery over wheat crops,*** submitted to Frontiers in Plant Science [1]. This article presents a method of plant density estimation from high spatial resolution images. The images used in this article were acquired using a light platform called "phenotypette". The method was recently adapted to images acquired from UAVs to increase the throughput[2].
- ***Article 2: Modeling the spatial distribution of plants on the row for wheat crops: Consequences on the green fraction at the canopy level***, submitted to Computers & Electronic in Agriculture[3]. The sowing pattern of the plants on the row was described by statistical models for the spacing of plants along the row and distance to the row center. This description of the plant spacing is supported by the previous observations made on the RGB images of the precise plant positions. It will be later used to investigate the optimal sampling strategy for plant density estimation as well as to provide a more realistic representation of the actual sowing patterns when simulating 3D scenes of wheat crops in the 2 following chapters. It has been used also here to evaluate the influence of the sowing pattern on the green fraction that can be easily measured with phenotyping techniques.
- ***Article 3: Optimal sampling to estimate plant density and quantify plant spacing heterogeneity along the row in wheat crops*** submitted to Plant Methods[4]***.*** In this last article, the statistical models developed in the previous paper is used to define the optimal sampling density estimation to reach a given uncertainty level. The optimal sampling strategy could be later used when estimating the sowing density based on high resolution images as proposed in the first two articles, or more classically with manual counting in the field. Further, application to quantify the spatial heterogeneity of the plant spacing is also addressed. This may be important to characterize the sowing pattern and its possible influence of the fate of the crop.

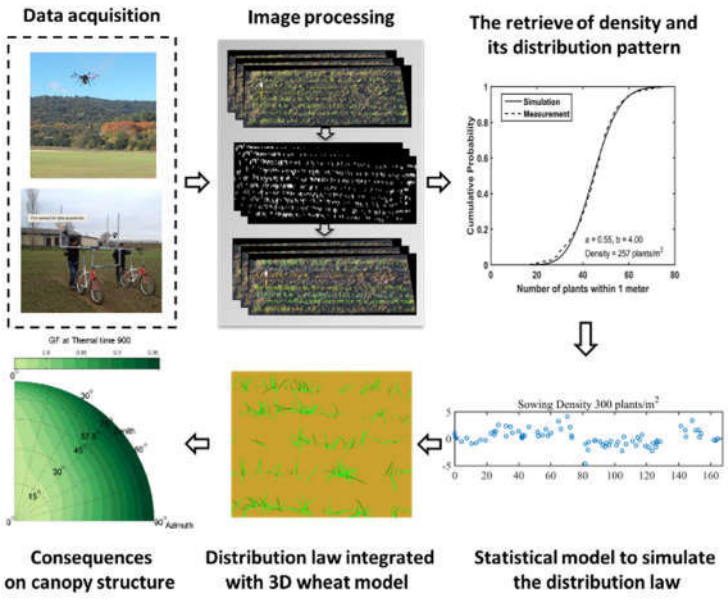

Figure abstract of Chapter 2

[1] Liu, S., Baret, F., Andrieu, B., & Burger, P. (2016). Estimation of plant density from high resolution RGB imagery over wheat crops. Frontiers in Plant Science, Submitted in September 2016

[2] Jin, X., Liu, S., Baret, F., Hemerlé, M., & Comar, A. (2016). Estimates of plant density from images acquired from UAV over wheat crops at emergence. Remote sensing of the environment, submitted

[3] Liu, S., Baret, F., Andrieu, B., Abichou, M., Allard, D., de Solan, B., & Burger, P. (2016). Modeling the spatial distribution of plants on the row for wheat crops: Consequences on the green fraction at the canopy level. Computers & Electronic in Agriculture, Submitted in October 2016

[4] Liu, S., Baret, F., Allard, D., Jin, X., Andrieu, B., Burger, P., Hemmerlé, M., & Comar, A. (2016). Optimal sampling to estimate plant density and quantify plant spacing heterogeneity along the row in wheat crops. Plant Methods

2.1. Estimation of plant density from high resolution RGB imagery over wheat crops

Shouyang Liu [a*], Fred Baret [a], Bruno Andrieu [b], Philippe Burger [c] and Matthieu Hemmerlé [d]

[a] *EMMAH, INRA, Avignon, France*

[b] *INRA-AgroParisTech, INRA, Thiverval-Grignon, France*

[c] *AGIR, INRA, Castanet Tolosan, France*

[d] *Hi-Phen, Avignon, France*

Abstract

Crop density is a key agronomical trait used to manage wheat crops and estimate yield. Visual counting of plants in the field is currently the most common method used. It is tedious and time consuming. The main objective of this work is to develop a machine vision based method to automate the density survey of wheat at early stages. The RGB images taken with a high resolution RGB camera are classified to identify the green pixels corresponding to the plants. Crop rows are extracted and the connected components (objects) are identified. A neural network is then trained to estimate the number of plants in the objects using the object features. The method was evaluated over 3 experiments showing contrasted conditions with sowing densities ranging from 100 to 600 seeds·m^{-2}. Results demonstrate that the density is accurately estimated with an average relative error of 12%.

The density estimation pipeline presented in this work enables an efficient and accurate prediction of wheat density at early stages. In the future, more sophisticated vectors, such as automatic rover or an unmanned aerial vehicle platform could be exploited to facilitate the acquisition of images in the field. It will further improve the throughput of the density estimation.

Keywords

Plant density, RGB imagery, Neural network, Wheat, Recursive feature elimination, Hough transform

1 Introduction

Wheat is one of the main crops cultivated around the world with sowing density usually ranging from 150 to 400 seed·m^{-2}. Plant population density may significantly impact the competition among plants as well as with weeds and consequently affect the effective utilization of available resources including light, water and nutrients (Olsen et al. 2006a; Shrestha and Steward 2003). Crop density appears therefore as one of the important variables that drive the potential yield. This explains why this information is often used for the management of cultural practices (Godwin and Miller 2003). Plant population density is still investigated most of the time by visually counting the plants in the field over samples corresponding either to a quadrat or to a segment. This is achieved at the stage when the majority of plants have just emerged and before the beginning of tillering (Norman 1995) which happens few days to few weeks after emergence. This method is time and labor intensive and may be prone to human error.

Some efforts have been dedicated to the development of high-throughput methods for quantifying plant density. This was mainly applied to maize using either capacitive sensors during the harvest (Li et al. 2009; Nichols 2000) or optical sensors including 2D cameras (Shrestha and Steward 2003, 2005; Tang and Tian 2008b, d) and range sensors (Jin and Tang 2009b; Nakarmi and Tang 2012a; Nakarmi and Tang 2014; Shi et al. 2015; Shi et al. 2013). However, quantifying the population density of maize is much simpler than that of wheat since maize plants are normally bigger, with larger plant spacing along the row and more evenly distributed. In wheat crops, leaves between neighboring plants overlap rapidly, and tillers will also appear, making the plant identification very difficult when they have more than three leaves, even using visual counting in the field. Most studies during these early stages report results derived from estimates of the vegetation fraction coverage measured using high resolution imagery (Guo et al. 2013) or based on vegetation indices computed with either multispectral (Sankaran et al. 2015) or hyperspectral (Liu et al. 2008) reflectance measurements. However, none of these investigations specifically addressed the estimation of plant density. Advances in digital photography providing very high resolution images, combined with the development of computer vision systems, offer new opportunities to develop a non-destructive high-throughput method for plant density estimation.

The objective of this study is to develop a system based on high resolution imagery that measures wheat plant population density at early stages. We collected pictures over a wide range of experiment conditions. Then we built the pipeline to process the images. Finally, the method was evaluated with emphasis on its accuracy and the corresponding domain of validity.

2 Material and Methods

2.1 Field experiments and measurements

Three experiments were conducted in 2014 in France (**Table 1**): Avignon, Toulouse and Paris. In Avignon, four sowing densities were sampled (100, 200, 300 and 400 seeds·m^{-2}) with the same "Apache" cultivar. In Toulouse, five sowing densities were sampled (100, 200, 300, 400 and 600 seeds·m^{-2}) with "Apache" and "Caphorn" cultivars. In Paris, two plots were sampled, corresponding to different cultivars with a single sowing density of 150 seeds·m^{-2}. All measurements were taken around 1.5 Haun stage, when most plants already emerged. A total of 16 plots were therefore available over the 3 experiments under contrasted conditions in terms of soil, climate, cultivars and sowing densities.

Table 1 Characteristics of the three experimental sites

Sites	Latitude	Longitude	Cultivars	Sowing density (seeds·m^{-2})	Illumination conditions	Cameras	Resolution	Focal length	Spatial resolution (mm)
Toulouse	43.5°N	1.5°E	Apache	100, 200, 300, 400, 600	Diffuse	Sigma SD14	2640 by 1760	50 mm	0.23
			Caphorn	100, 200, 300, 400, 600					
Paris	48.8°N	1.9°E	Premio	150	Flash	NIKON D5200	4496 by 3000		0.16
			Attlass	150					

Avignon	43.9°N	4.8°E	Apache	100, 200, 300, 400	Direct	Sigma SD14	4608 by 3072		0.13

In Toulouse and Avignon, images were acquired using an RGB camera fixed on a light moving platform, termed Phenotypette (**Fig. 1**). The platform was driven manually at about 0.5 $m \cdot s^{-1}$. For each treatment, at least 10 pictures were collected at a frequency allowing significant overlap. For Paris experiment, the camera was mounted on a monopod to take two pictures with no overlap. In all the cases, the camera was oriented at 45° inclination perpendicular to the row direction and was pointing at the center row from a distance of 1.5m and with spatial resolution around 0.2 mm (**Table 1**).

Fig. 1 The image acquisition in the field with the Phenotypette

2.2 Image processing

The image processing includes several steps sketched on **Fig. 2** and explained in detail in the following paragraphs.

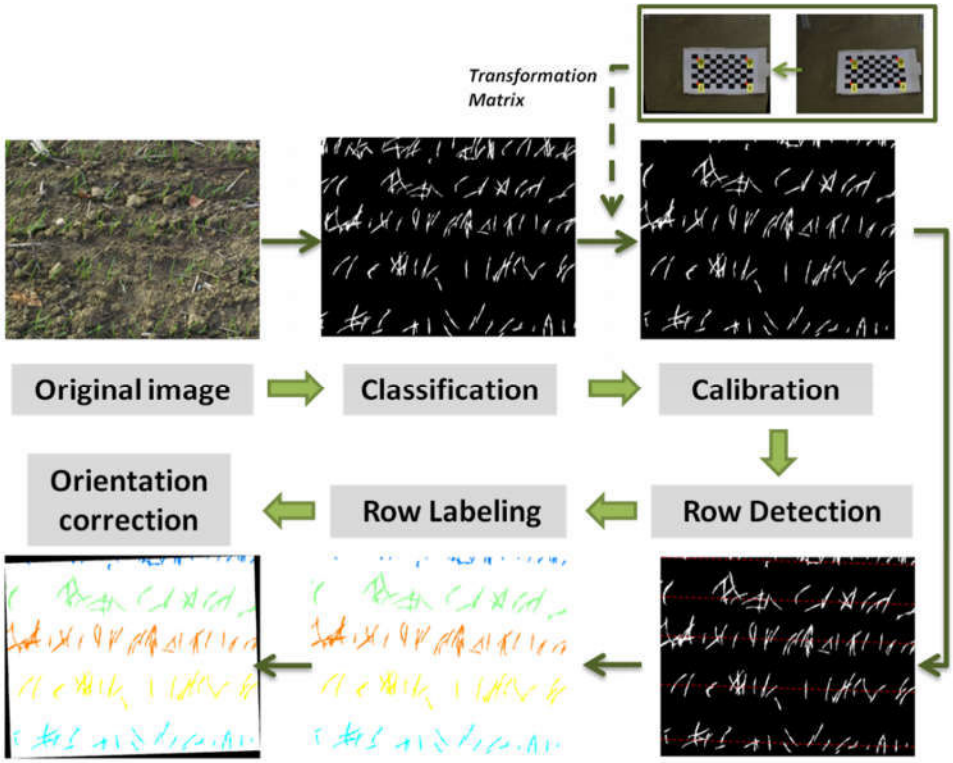

Fig. 2 The methodology involving image processing and feature extraction

2.2.1 Classification of green elements

The images collected display green pixels corresponding to the emerged wheat plants, and brown pixels corresponding to the soil background. The RGB color space was firstly transformed into Lab, consequently

enhancing its sensitivity to variations in greenness (Philipp and Rath 2002). The Otsu automatic thresholding method (Otsu 1975) was then applied to channel 'a' to separate the green from the background pixels. Results show that the proposed method performs well under the contrasted illumination conditions experienced (**Fig. 2**). Further, this approach provides a better identification of the green pixels (results not presented for the sake of brevity) as compared to the use of supervised methods (Guo et al. 2013) based on indices such as the excess green (Woebbecke et al. 1995) or more sophisticated indices proposed by Meyer and Neto (2008).

2.2.2 Geometric transformation

The perspective effect creates a variation of the spatial resolution within the image. A transformation was therefore applied to remap the image into an orthoimage where the spatial resolution remains constant. The transformation matrix was calibrated using an image of a chessboard for each camera setup (**Fig. 2**). The chessboard covered the portion of the image that was later used for plant counting. The corners of the squares in the chessboard were identified and the transformation matrix was computed (**Fig. 2**). The transformation matrix was finally applied to the whole image. This geometric transformation is fully automatic once the actual dimension of the squares of the chessboard is provided.

2.2.3 Row identification and orientation

The plant density measurement for row crops such as wheat is achieved by counting plants over a number of row segments of given length. Row identification is therefore a mandatory step as sketched in **Fig. 2**. Row identification methods have been explored intensively mostly for the automation of robot navigation in field (Vidović et al. 2016). Montalvo et al. (2012) reviewed the existing methods and found that the Hough transform (Slaughter et al. 2008) is one of the most common and reliable methods. It mainly involves computing the co-distribution of the length (ρ) and orientation (θ) of the segments defined by two green pixels (**Fig. 3a**). The Hough transform detects dominant lines even in the presence of noise or discontinuous rows. The noise could include objects between rows such as weeds or misclassified background pixels such as stones (Marchant 1996; Rovira-Más et al. 2005). Although the Hough transform is computationally demanding, its application on edge points of the green objects decreases this constraint. Hence, the 'Canny Edge Detector' (Canny 1986) was consequently used to detect edges prior to the application of the Hough transform. The Hough transform was conducted with orientation $-90^{o} < \theta < 90^{o}$ with 0.1^{o} angular steps and a radius $-3000 < \rho < 3000$ pixels with 1 pixel steps (**Fig. 3a**).

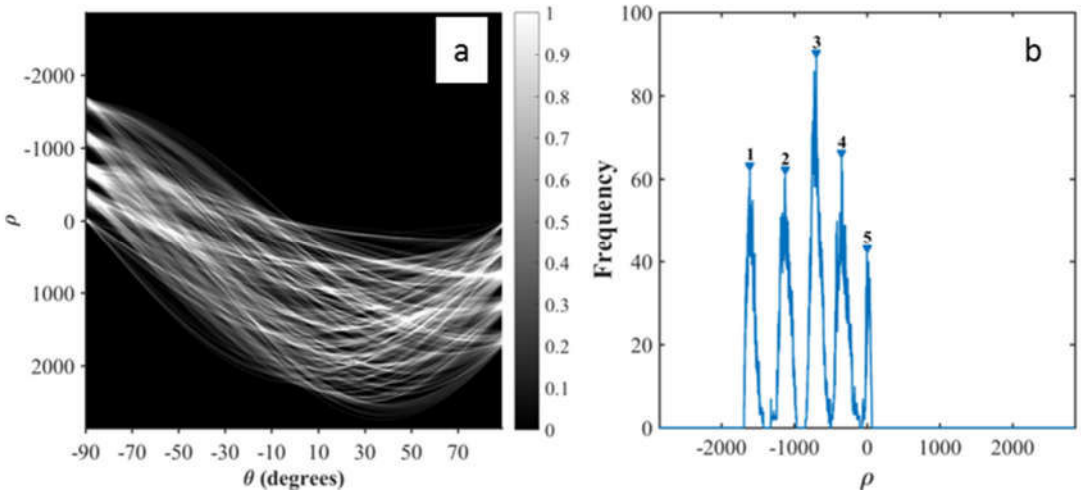

Fig. 3 Hough transform to detect rows. a) Hough transform. b) Identification of the peaks of ρ corresponding to the rows.

Five main components show up in the image (**Fig. 3a**), corresponding to the five rows of the original image (**Fig. 2**). As all rows are expected to be roughly parallel, their orientation could be inferred as the θ value, θ_{row} (where $\theta_{row} = 90°$ corresponds to the horizontal orientation on the images on **Fig. 2**), that maximizes the variance of ρ. The positions of the rows are derived from the peaks of frequency for $\theta = \theta_{row}$ (**Fig. 3b**). Five lines on **Fig. 2** highlight the center of each row. Because of the uncertainty in the orientation of the

Phenotypette along the row, the row line drawn on the images are not exactly horizontal. This is illustrated in **Fig. 3b** where $\theta_{row} = -88.2°$. The images were therefore rotated according to θ_{row} so that the rows are strictly horizontal in the displayed images.

Object identification and feature extraction

The objects corresponding to connected pixels with the connectivity of eight neighbors were identified on the binary image. Each object was associated to the closest row line and characterized by ten main features (the top 10 features in **Table 5**). Three additional features were derived from skeletonization of the object: the length, number of branch and end points of the skeleton (the last 3 features in **Table 5**).

Table 4 The 13 features extracted for each of the connected object

#	Name	Meaning	Unit
F1	Area	Number of pixels of the connected component (object)	Pixel
F2	FilledArea	Number of pixels of the object with all holes filled	Pixel
F3	ConvexArea	Number of pixels within the associated convex hull	Pixel
F4	Solidity	Ratio of number of pixels in the region to that of the convex hull	Scalar
F5	Extent	Ratio of number of pixels in the region to that of the bounding box	Scalar
F6	EquivDiameter	Diameter of a circle with the same area as the region	Pixel
F7	MajorAxisLength	Length of the major axis of the ellipse equivalent to the region.	Pixel
F8	MinorAxisLength	Length of the minor axis of the ellipse equivalent to the region.	Pixel
F9	Eccentricity	Eccentricity of the equivalent ellipse to the region	Scalar
F10	Orientation	Orientation of the major axis of the equivalent ellipse	Degree
F11	LengthSkelet	Number of pixels of the skeleton	Pixel
F12	NumEnd	Number of end points of the skeleton	Scalar
F13	NumBranch	Number of branch points of the skeleton	Scalar

2.3 Estimation of the number of plants contained in each object

Machine learning methods were used to estimate the number of plants contained in each object from the values of their 13 associated features (**Table 5**). Artificial Neural Networks (ANNs) have been recognized as one of the most versatile and powerful method to relate a set of variables to one or more variables. ANNs are interconnected neurons characterized by a transfer function. They combine the input values (the features of the object) to best match the output values (number of plants in our case) over a training database. Training process requires first to define the network architecture (the number of hidden layers and nodes per layer and the type of transfer function of each neuron). Then the synaptic weights and biases are tuned to get a good agreement between the number of plants per object estimated from the object's features and the corresponding number of plants per object in the training database. A one-layer feed-forward network with n_{node} tangent sigmoid hidden nodes and one linear node was used. The number of hidden nodes was varied between $1 \leq n_{node} \leq 10$ to select the best architecture. The weights and biases were initialized randomly. The training was achieved independently over each situation, considering 90% of the whole data set containing for each site a total of the 606 (Toulouse), 347 (Paris) and 476 (Avignon) objects were identified. The remaining 10% of each site was used to evaluate the performance of the training. Note that the estimates of number of plants per object were continuous, i.e. representing actually the average probability of getting a discrete number of plants.

A compact, parsimonious and non-redundant subset of features should contribute to speed up the learning process and improve the generalization of predictive models (Kuhn and Johnson 2013; Louppe 2014; Tuv et al. 2009). Guyon et al. (2002) proposed Recursive Feature Elimination (RFE) to select the optimal subtest of features. Specific to ANN, the combinations of the absolute values of the weights were used firstly to rank the importance of predictors (features) (Gevrey et al. 2003; Olden and Jackson 2002). For the subset including

n features, RFE presumes that the subset of the top n features outperforms the other possible combinations (Granitto et al. 2006; Guyon et al. 2002). Then 13 iterations corresponding to the 13 features are needed to select the optimal subset. The optimal subset of n features is defined as the minimum number of features providing an $RMSE_n$ lower than 1.02 $RMSE_{best}$, where $RMSE_{best}$ is the minimum RMSE over 13 iterations. To minimize possible overfitting of the training dataset, a cross-validation scheme was used (Seni and Elder 2010) with the training data set including 90% of the cases and the test data set containing the remaining 10%. The process was repeated 5 times with a random drawing of the training and test data sets for each trial.

3 Results

3.1 Number of plants per object and object feature selection

The number of plants per object resulted in a consistently right-skewed distribution over the three experimental sites (**Fig. 4**): more than 80% of the objects contained 2 plants or less. Approximately 50% of the objects contained single plants in Toulouse and Avignon experiments. Conversely, about 60% objects contained a single plant in the Paris experiment because of its lower plant density (**Fig. 4**): the probability of overlap between plants is obviously lesser than that for higher plant densities. Note that 10% to 20% of the objects were classified as null, i.e. containing no plants. This corresponds to errors in separating plants from the background: objects such as straw residues, stone or weeds may show colors difficult to separate in the classification step. Further, due to the variability of the illumination conditions, plants may be misclassified into two disconnected objects. In this case, the larger part is considered as a plant while the smaller remaining part is considered as non-plant, i.e. set to 0.

Most of the 13 features described in **Table 5** are closely related as illustrated by the plot-matrix of the Toulouse experiment (**Fig. 5**). Correlations are particularly high between the 4 area related features (F1, F2, F3, F6), between the skeleton derived features (F11, F12, F13), and between the area and skeleton related features. Similar correlations were observed over the Paris and Avignon experiments. These strong relationships indicate the presence of redundancy between the 13 features, which may confuse the training of ANN. However, this could be partly overcome by the RFE feature selection algorithm.

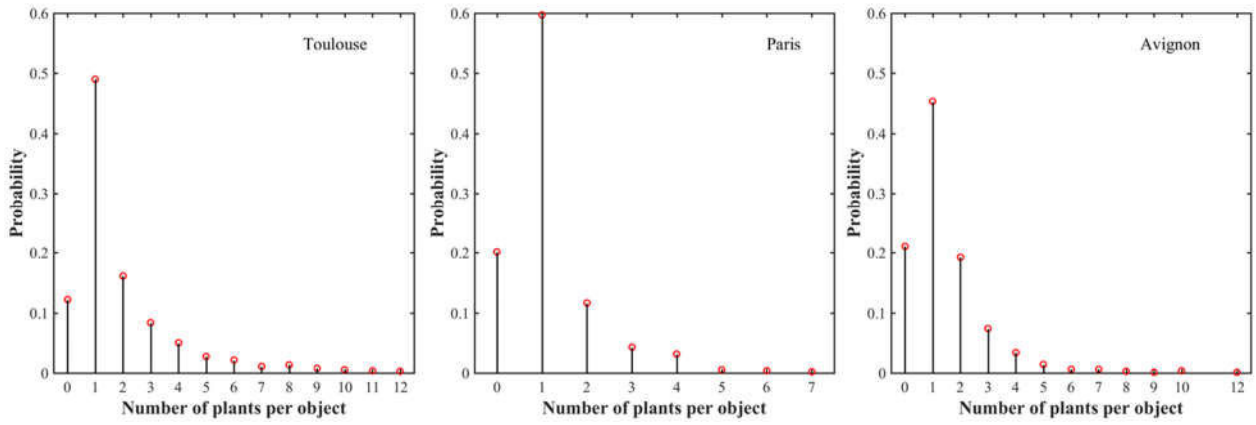

Fig. 4 Number of plants per object three experiments

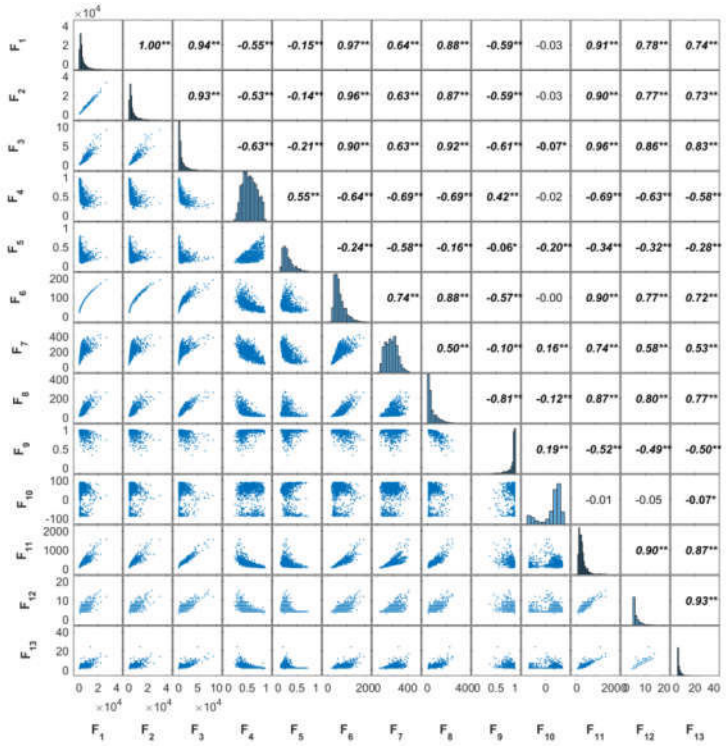

Fig. 5 The correlation among the 13 objects' features over Toulouse experiment (: 0.01, *: 0.05). The abbreviation of features refers to names in Table 5.**

The estimation performance of the number of plants per objects was firstly evaluated using the RMSE (**Fig. 6**). For the three sites, the RMSE decreases rapidly with the number of features used, particularly for the first 4 features (**Fig. 6**). The improvement in estimation performance is then relatively small with the inclusion of remaining features. The number of features required according to our criterion varies from 10 (Toulouse) to 4 (Avignon). A more detailed inspection of the main features used across the three experiments (**Table 6** and **Table 7**) shows the importance of the area related features (F1, F2, F3, F4 and F6) despite their high correlation (**Fig. 5**). The length of the skeleton (F11) also appears important particularly for the Avignon site, while the orientation and extent do not help much (**Table 7**).

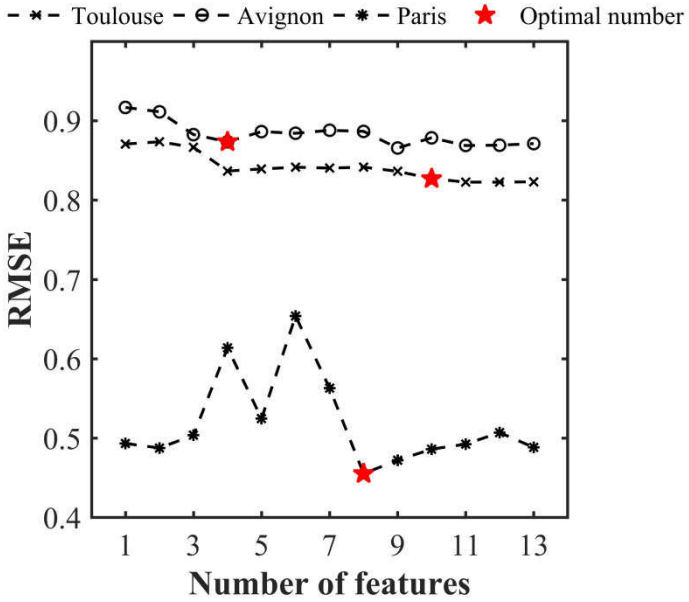

Fig. 6 RMSE associated to the estimates of the number of plants per object as a function of the number of features used. The RMSE was evaluated over the test data set for each individual experiment.

Table 5 Performance of the estimation of the number of plants per object over three experiments

Sites	Training size	n_{node}	Number of Features	R^2	RMSE	Bias
Toulouse	606	2	10	0.83	0.83	0.28
Paris	347	2	8	0.79	0.47	0.077
Avignon	476	2	4	0.61	0.87	0.45

Table 6 Features selected and the corresponding rank over three experiments

#	Features	Toulouse	Paris	Avignon
F_1	Area	2	1	3
F_2	FilledArea	1	3	
F_3	ConvexArea	4	4	2
F_4	Solidity	10	7	
F_5	Extent			
F_6	EquivDiameter	3	2	4
F_7	MajorAxisLength			
F_8	MinorAxisLength	5	5	
F_9	Eccentricity	8	8	
F_{10}	Orientation			
F_{11}	LengthSkelet	6	6	1
F_{12}	NumEnd	7		
F_{13}	NumBranch	9		

As expected, the model performs the best for the Paris experiment (**Fig. 7**) where the situation is simpler because of the low density inducing limited overlap between plants (**Fig. 4**). Conversely, the larger number of null objects (**Fig. 4**) corresponding to misclassified objects or split plants in the Avignon experiment, explains the degraded performance (**Fig. 7**). The bias in the estimation of the number of plants per object appears relatively small, except for the Avignon site. Attention should be paid on the bias since the application of the neural network on a larger number of objects is not likely to improve the estimation of the total number of plants. The bias is mostly due to difficulties associated to the misclassified objects. Note that the estimation performance degraded for the larger number of plants per objects (**Fig. 7**) as a consequence of more ambiguities and smaller samples used in the training process.

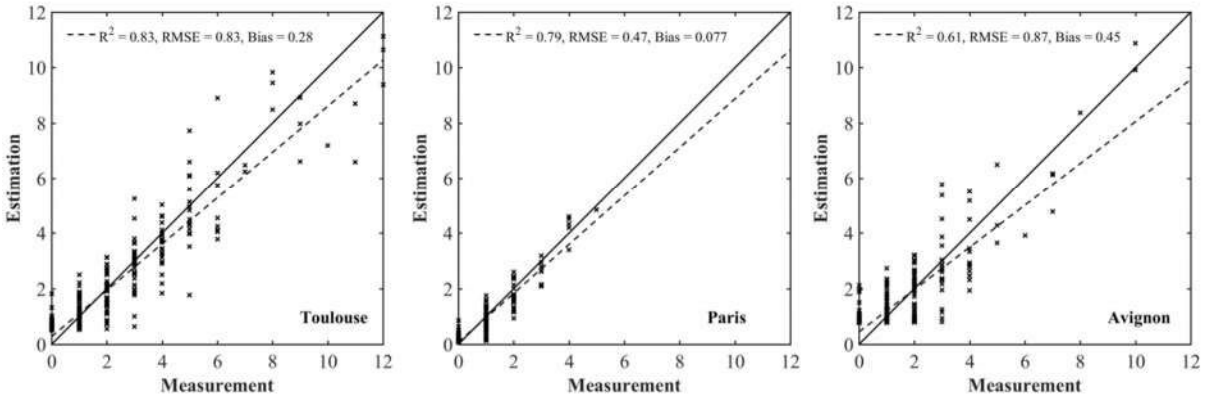

Fig. 7 Comparison between the estimated number of plants per object with the value measured over the test dataset for each individual experiment

3.2 Performance of the method for plant density estimation

The estimates of plant density were computed by summing the number of plants in all the objects extracted from the row segments identified in the images, divided by the segment area (product of the segment length

and the row spacing). The reference density was computed from the visually identified plants. Results show a good agreement between observations and predictions over sowing densities ranging from 100 to 600 seeds·m^{-2} (**Fig. 8**). The performances slightly degrade for densities higher than 350 plants·m^{-2}. This may be explained by the difficulty to handle more complex situation when plant spacing decreases, with a higher probability of plant overlap. Also note the slight overestimation for the low densities in the Avignon site, mainly attributed to the bias introduced in the estimation of the number of plants per object due to the classification problem already outlined.

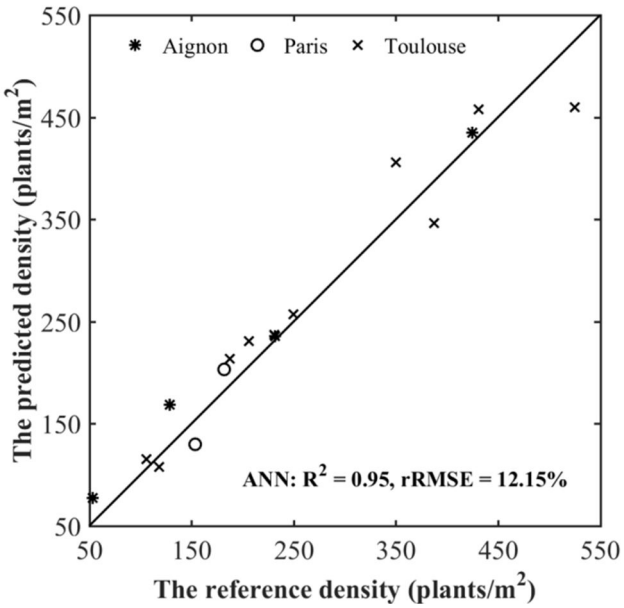

Fig. 2 Performance of density estimation over the three experiments

4 Discussion and conclusions

The method proposed in this study relies on the ability to identify plants or group of plants from RGB images. Image classification is a thus a critical step driving the accuracy of the plant density estimation. Wheat plants at emergence have a relatively simple structure and color. The image quality is obviously very important, including the optimal spatial resolution that should be better than 0.4 mm as advised by (Jin et al. 2016). Further, the image quality should not be compromised by undesirable effects due to image compression algorithms. As a consequence, when the resolution is between 0.2 to 0.5 mm, it would be preferable to record images in raw format to preserve its quality. A known and fixed white balance should be applied to make the series of images comparable in terms of color. Finally, the view direction is chosen to increase the plant cross section by taking images inclined at around 45° zenith angle in a compass direction perpendicular to the row orientation. Note that too inclined views may result in large overlap of plants from adjacent rows which will pose problems for row (and plant) identification.

Plants were separated from the background based on their green color. A unique unsupervised method based on the Lab transform on which automatic thresholding is applied was used with success across a range of illumination conditions. However, the method should be tested under a much larger range of illumination and soil conditions before ensuring that it is actually applicable in all scenarios. Additionally, attention should be paid to weeds that are generally green. Fortunately, weeds were well controlled in our experiments. Although this is also generally observed during emergence, weed detection algorithm could be integrated in the pipeline in case of significant infestation. Weeds may be identified by their position relative to the row (Woebbecke et al. 1995). However, for the particular observational configuration proposed (45° perpendicular to the row), the application of these simple algorithms are likely to fail. Additional (vertical) images should be taken, or more refined methods based on the color (Gée et al. 2008) or shape (Swain et al. 2011) should be implemented.

Once the binary images are computed from the original RGB ones, objects containing uncertain number of plants can be easily identified. An ANN method was used in this study to estimate the number of plants from

the 13 features of each object. Alternative machine learning techniques could have been used. However, additional work demonstrating the superiority of ANN as compared to random forest (Breiman 2001), multilinear regression (Tabachnick et al. 2001) and generalized linear model (Lopatin et al. 2016) is not presented in this study for the sake of brevity. The RFE algorithm used to select the minimum subset of features for the best estimation of number of plants per object (Granitto et al. 2006) resulted in 5 to 10 features depending on the data set considered. The features selected are mainly related to the object area and the length of the corresponding skeleton. Conversely, object orientation and extent appear not to be of pertinent importance. The RFE framework employed here partly accounts for the strong co-dependency between the 13 features considered. The selection process could probably be improved using a recursive scheme similar to the one employed in stepwise regression, or a transformation of the space of the input features.

The wheat population density was estimated with an average of 12% error. The error increases with the population density because of the increase of overlap between plants creating larger objects, hence making it more difficult to associate accurately the number of plants they contain. Likewise, the degradation of performances can be observed when plants are well developed. A timely observation just between Haun stage 1.5 to 2 corresponding to 1.5 to 2 phyllochron after emergence appears optimal: plants are enough developed to be well identified while the overlap between plants is minimized because of the low number of leaves (between 1 and 2) and their relatively erect orientation. However, in case of heterogeneous emergence, it is frequent to observe a delay of about 1 phyllochron (Hokmalipour 2011; Jamieson et al. 2008) between the first and the last plant emerged. Observation between Haun stage 1.5 and 2 can thus ensure that the majority has emerged. Since the phyllochron varies between 63 °C·d to 150 °C·d (McMaster and Wilhem 1995), the optimal time window of 0.5 phyllochron (between Haun stage 1.5 to 2) can last about 4 to 8 days under an average 10°C air temperature. This short optimal time window for taking the images is thus a strong constraint when operationally deploying the proposed method.

The success of the method relies heavily on the estimation of the number of plants per object. The machine learning technique used in this study was trained independently for each experiment. This provides the best performances because it takes into account the actual variability of single plant structure that depends on its development stage at the time of observation, on the genotypic variability as well as on possible influence of the environmental conditions, especially wind. Operational deployment of the method therefore requires the model to be re-calibrated over each new experimental site. However, a single training encompassing all the possible situations may be envisioned in near future. This requires a large enough training data set representing the variability of genotypes, development stage and environmental conditions. This single training data base could also include other cereal crops species similar to wheat at emergence such as barley, triticale or oat.

Several vectors could be used to take the images, depending mostly on the size of the experiment and the resources available. A monopod and a light rolling platform, the Phenotypette, were used in our study. More sophisticated vectors with higher throughput could be envisioned in the next step, based either on a semi-automatic (Comar et al. 2012a) or fully automatic rover (de Solan et al. 2015) or a UAV platform as recently demonstrated by (Jin et al. 2016).

References

Breiman, L. (2001). Random forests. *J Mach Learn, 45*

Canny, J. (1986). A Computational Approach to Edge Detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-8*, 679-698

Comar, A., Burger, P., de Solan, B., Baret, F., Daumard, F., & Hanocq, J.-F. (2012). A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: description and first results. *Functional Plant Biology, 39*, 914-924

de Solan, B., Baret, F., Thomas, S., Beauchêne, K., Comar, A., Fournier, A., Weiss, M., & Gouache, D. (2015). PHENOMOBILE V: A fully automated high throughput phenotyping system. In Poster (Ed.), *Eucarpia*. Montpellier (France)

Gée, C., Bossu, J., Jones, G., & Truchetet, F. (2008). Crop/weed discrimination in perspective agronomic images. *Computers and Electronics in Agriculture, 60*, 49-59

Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling, 160*, 249-264

Godwin, R., & Miller, P. (2003). A review of the technologies for mapping within-field variability. *Biosystems Engineering, 84*, 393-407

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

Guo, W., Rage, U.K., & Ninomiya, S. (2013). Illumination invariant segmentation of vegetation for time series wheat images based on decision tree model. *Computers and Electronics in Agriculture, 96*, 58-66

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning, 46*, 389-422

Hokmalipour, S. (2011). The Study of Phyllochron and Leaf Appearance Rate in Three Cultivar of Maize (Zea mays L.) At Nitrogen Fertilizer Levels. *World Applied Sciences Journal, 12*, 850–856

Jamieson, P., Brooking, I., Zyskowski, R., & Munro, C. (2008). The vexatious problem of the variation of the phyllochron in wheat. *Field Crops Research, 108*, 163-168

Jin, J., & Tang, L. (2009). Corn plant sensing using real‐time stereo vision. *Journal of Field Robotics, 26*, 591-608

Jin, X., Liu, S., Baret, F., Hemerlé, M., & Comar, A. (2016). Estimates of plant density from images acquired from UAV over wheat crops at emergence. *Remote sensing of the environment, submitted*

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer

Li, H., Worley, S., & Wilkerson, J. (2009). Design and optimization of a biomass proximity sensor. *Transactions of the ASABE, 52*, 1441-1452

Liu, J., Miller, J.R., Haboudane, D., Pattey, E., & Hochheim, K. (2008). Crop fraction estimation from casi hyperspectral data using linear spectral unmixing and vegetation indices. *Canadian Journal of Remote Sensing, 34*, S124-S138

Lopatin, J., Dolos, K., Hernández, H., Galleguillos, M., & Fassnacht, F. (2016). Comparing Generalized Linear Models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. *Remote Sensing of Environment, 173*, 200-210

Louppe, G. (2014). Understanding random forests: From theory to practice. *arXiv preprint arXiv:1407.7502*

Marchant, J. (1996). Tracking of row structure in three crops using image analysis. *Computers and Electronics in Agriculture, 15*, 161-179

McMaster, G.S., & Wilhem, W.W. (1995). Accuracy of equations predicting the phyllochron of wheat. *Crop Science, 35*, 30–36

Meyer, G.E., & Neto, J.C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture, 63*, 282-293

Montalvo, M., Pajares, G., Guerrero, J.M., Romeo, J., Guijarro, M., Ribeiro, A., Ruz, J.J., & Cruz, J. (2012). Automatic detection of crop rows in maize fields with high weeds pressure. *Expert Systems with Applications, 39*, 11889-11897

Nakarmi, A., & Tang, L. (2012). Automatic inter-plant spacing sensing at early growth stages using a 3D vision sensor. *Computers and Electronics in Agriculture, 82*, 23-31

Nakarmi, A.D., & Tang, L. (2014). Within-row spacing sensing of maize plants using 3D computer vision. *Biosystems Engineering, 125*, 54-64

Nichols, S.W. (2000). Method and apparatus for counting crops. In: Google Patents

Norman, D.W. (1995). *The farming systems approach to development and appropriate technology generation*. Food & Agriculture Org.

Olden, J.D., & Jackson, D.A. (2002). Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling, 154*, 135-150

Olsen, J., Kristensen, L., & Weiner, J. (2006). Influence of sowing density and spatial pattern of spring wheat (Triticum aestivum) on the suppression of different weed species. *Weed Biology and Management, 6*, 165-173

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica, 11*, 23-27

Philipp, I., & Rath, T. (2002). Improving plant discrimination in image processing by use of different colour space transformations. *Computers and Electronics in Agriculture, 35*, 1-15

Rovira-Más, F., Zhang, Q., Reid, J., & Will, J. (2005). Hough-transform-based vision algorithm for crop row detection of an automated agricultural vehicle. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 219*, 999-1010

Sankaran, S., Khot, L.R., & Carter, A.H. (2015). Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. *Computers and electronics in agriculture, 118*, 372-379

Seni, G., & Elder, J.F. (2010). Ensemble methods in data mining: improving accuracy through combining predictions. *Synthesis Lectures on Data Mining and Knowledge Discovery, 2*, 1-126

Shi, Y., Wang, N., Taylor, R., & Raun, W. (2015). Improvement of a ground-LiDAR-based corn plant population and spacing measurement system. *Computers and Electronics in Agriculture, 112*, 92-101

Shi, Y., Wang, N., Taylor, R.K., Raun, W.R., & Hardin, J.A. (2013). Automatic corn plant location and spacing measurement using laser line-scan technique. *Precision Agriculture, 14*, 478-494

Shrestha, D.S., & Steward, B.L. (2003). Automatic corn plant population measurement using machine vision. *Transactions of the ASAE, 46*, 559

Shrestha, D.S., & Steward, B.L. (2005). Shape and size analysis of corn plant canopies for plant population and spacing sensing. *Applied engineering in agriculture, 21*, 295-303

Slaughter, D., Giles, D., & Downey, D. (2008). Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture, 61*, 63-78

Swain, K.C., Nørremark, M., Jørgensen, R.N., Midtiby, H.S., & Green, O. (2011). Weed identification using an automated active shape matching (AASM) technique. *Biosystems Engineering, 110*, 450-457

Tabachnick, B.G., Fidell, L.S., & Osterlind, S.J. (2001). Using multivariate statistics

Tang, L., & Tian, L.F. (2008a). Plant Identification in Mosaicked Crop Row Images for Automatic Emerged Corn Plant Spacing Measurement. *Transactions of the ASABE, 51*, 2181-2191

Tang, L., & Tian, L.F. (2008b). Real-time crop row image reconstruction for automatic emerged corn plant spacing measurement. *Transactions of the ASABE, 51*, 1079-1087

Tuv, E., Borisov, A., Runger, G., & Torkkola, K. (2009). Feature selection with ensembles, artificial variables, and redundancy elimination. *The Journal of Machine Learning Research, 10*, 1341-1366

Vidović, I., Cupec, R., & Hocenski, Ž. (2016). Crop Row Detection by Global Energy Minimization. *Pattern recognition*

Woebbecke, D., Meyer, G., Von Bargen, K., & Mortensen, D. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE, 38*, 259-269

2.2. Modeling the spatial distribution of plants on the row for wheat crops: Consequences on the green fraction at the canopy level

Shouyang Liu[a, *], Frédéric Baret[a], Bruno Andrieu[b], Mariem Abichou[b], Denis Allard[c], Benoit de Solan[a, d], Philippe Burger[e]

[a] *INRA EMMAH, UMR 1114 Domaine Saint-Paul, Site Agroparc, 84914 Avignon Cedex 9, France*

[b] *INRA-AgroParisTech, UMR 1091 EGC, 78850 Thiverval-Grignon, France*

[c] *INRA BioSP, UMR 1114 Domaine Saint-Paul, Site Agroparc, 84914 Avignon Cedex 9, France*

[d] *ARVALIS-Institut du végétal, Station expérimentale, 91720 Boigneville, France*

[e] *INRA, UMR 1248 AGIR, Chemin de Borde Rouge, BP52627, 31326 Castanet Tolosan Cedex, France*

Abstract

This work investigates the spatial distribution of wheat plants that drives the competition between plants and weeds with consequences on crop productivity. A set of RGB images were taken from nadir on a total 14 plots showing a range of sowing density, cultivars and environmental conditions. The coordinates of the plants were extracted from RGB images. Results show that the distance between-plants along the row follow a gamma distribution law, with no dependency between the distances. Conversely, the positions of the plants across rows follow a Gaussian distribution with parameters relatively stable over the 14 cases investigated. Further, the positions across the row were strongly interdependent. A statistical model was thus proposed to simulate the possible plant distribution pattern and evaluate its impact on canopy structure emerging properties such as the green fraction (GF) that drives the light interception efficiency. The architecture of individual plants was described at several growth stages using the 3D Adel-Wheat model. Simulations showed that the effects on the GF of the plant distribution pattern were generally small. However, for the intermediate development stages, large zenith angles and directions parallel to the row, the deviations across the row of plant position increased the GF by more than 0.1. These results were obtained with a wheat architecture model that does not account for the capacity of plants to adapt to their local environment. Further developments of plant architecture models should include possible plasticity mechanisms.

Keywords

Plant distribution pattern, Green fraction, FSPM, Wheat, 3D model

1 Introduction

Plants adapt to the environment through adjusting physiological functions and structure (Vos et al. 2010). The plant spatial distribution is obviously initialized at the sowing. The seed drill is pulled by the tractor and the sowing is driven by a combination of gears synchronized with the ground speed of the drill. The variation of seed spacing along the row can arise in seed metering, release, flight from the mechanism to the soil surface and soil covering and pressing (Wilson 1980).The limitations of the seed metering system can cause misses and bunches from place to place (Karayel et al. 2006). In addition, several factors may contribute to the non-uniformity of seed inter-spacing along the row and deviations from the row direction including variations of the state of the soil surface, changes in the driving speed and vibrations or deviations of the drill from the pre-defined row direction. Further, the geminating process is driven by the seed vigor and seedbed temperature and moisture conditions(Bradford 2002; Orzolek and Daum 1984) which impact the rate and the dynamics of emergence (Bradford 2002; Rowse and Finch-Savage 2003). All these factors interactively contribute to the final non-uniformity of plant distribution, including the non-even distribution of inter-plant distance along the row and the deviation across the row.

The crowding stress coming from the competition with neighboring plants is an important factor that determines the inter-plant variability and the final performance at the stand scale (Pagano et al. 2007). The yield of a single plant in a stand is reduced by the presence of competing neighbors (Duncan 1984), including weed plants (interspecific competition) as well as crop plants (intraspecific competition) (Evers and Bastiaans 2016). For a given environment, the yield reduction depends on the intensity of the competition for yield-correlated resources (Duncan 1984). Agronomists have found that evenly spaced stands have generally higher potential yield than unevenly spaced stands (Liu et al. 2004). However recent work on maize showed mixed results on the effect on maize grain yield of crowding stress experienced by non-uniformly spaced plants (Liu et al. 2004; Pagano et al. 2007). Nevertheless for wheat, increasing crop density and spatial uniformity consistently strengthen crop competitiveness across species, resulting in suppression of weed growth and a reduction in yield loss (Olsen et al. 2006a; Olsen et al. 2005; Weiner et al. 2001). However, the previous results were mainly derived from the comparison of crop yield or biomass under the uniform pattern achieved by precise seed drill systems and the so-called 'standard' sowing condition. However, the standard sowing pattern may vary greatly among experimental conditions. It is therefore necessary to describe more rigorously the actual plant distribution achieved under standard seeding conditions. In addition, most of the results cited previously focused on the within-row non-uniformity while only little attention has been paid on the deviation across the row.

Apart from the experimental approach, modelling may also provide very useful outputs. Most current crop functioning models are based either on the 'big leaf' concept or are considering the behavior of the canopy as resulting from the functioning of identical plants under a fixed spatial distribution pattern. Therefore, these models do not allow describing the possible effect of the plant spatial distribution (Deen et al. 2003; Evers and Bastiaans 2016). Functional-structural models (FSPMs) are the emerging and potential tools to describe canopy response across environments, especially when within-field variability should be taken into account (Vos et al. 2010). FSPMs combine a structural module that represents the 3D structure of the plant with physiological modules describing the main processes including photosynthesis, transpiration, respiration and resource allocation that drive the dynamics under specific environmental conditions. As a matter of facts, the spatial arrangement of plants and organs affects the micro-climate within the canopy with consequences on plant functioning and also interactions with other organisms such as pathogens (Costes et al. 2012). Therefore FSPMs should ideally be driven by physical variables at the organ scale that describe the phylloclimate and its spatial distribution within canopy space along the growth cycle (Chelle 2005). For instance, Evers and Bastiaans (2016) developed a FSPM to simulate the crop-weed competition for light. It includes modules to describe plant architecture and the distribution pattern of the crop-weed canopy, and is capable of simulating the dynamics of individual plants while competing for resources in three-dimensional (3D) space. However, the pattern of plant spatial distribution is not well characterized even for the state of the art 3D canopy models. Very simple assumptions on the plant spatial distribution along rows are generally used (Garin et al. 2014; Gigot et al. 2014; Robert et al. 2008): plants are often aligned exactly along the row with strictly equal spacing. Sometimes, random deviations are added subjectively through

randomly moving (Feng et al. 2014) or removing plants (López-Lozano et al. 2007) from their original systematic positions. However, these plant spatial distribution patterns were not supported by actual field observations. Further, the consequences of the sowing pattern along the row on the stand level properties and canopy functioning was not yet addressed.

The objectives of this study are to document the actual plant distribution pattern based on field measurements at several locations. A statistical description of the distribution of plants along and across the row will be tentatively proposed. Then, the consequences of the plant distribution pattern will be evaluated at the stand scale using an in silico experiment. However, although modelling efforts have been made to improve our understanding on the competition mechanisms at the plant level (Berger et al. 2008), the plasticity of plants is still largely unknown. The impact of the plant spatial distribution pattern will therefore be evaluated on a canopy level emerging property, the green fraction (GF), assuming that plants do not adapt to their neighboring environment. The green fraction in the sun direction corresponds to the light fraction intercepted by green elements which drives photosynthesis and transpiration processes. When all the leaves are green, GF is closely related to the gap fraction, Po: GF=1-Po (Baret et al. 1993a; Baret et al. 1993b). GF is also very important for interpreting remote-sensing observations achieved either from the sensors aboard satellite or vectors closer to the ground level (Comar et al. 2012b; Verger et al. 2014b) for precision agriculture or phenotyping applications. The ADEL-Wheat 3D model (Fournier et al. 2003) will be used here for simulating wheat stands with a range of plant spatial distribution patterns and derive the corresponding GF under a set of directions.

2 Materials and methods

2.1 Field experiment

The experiment was conducted at three locations in France, Avignon (43.917°N; 4.879°E), Toulouse (43.531°N; 1.500°E) and Grignon (48.849°N; 1.917°E) in 2014. In all the cases, a mechanical seed drill was used, which represents the standard practice for wheat crops. In Grignon, five plots were sampled, corresponding to different cultivars with a single sowing density 170 plt/m². In Toulouse, five sowing densities were sampled (100, 200, 300, 400 and 600 plt/m²) with the same “Apache” cultivar. In Avignon, four sowing densities were sampled (100, 200, 300 and 400 plt/m²) also with the same “Apache” cultivar. The row spacing was all set as 17.5 cm over three experimental sites. All measurements were taken at around 1.5 Haun stage, when most plants already emerged and when it is easy to identify individual plants visually. A total of 14 plots are therefore available over the 3 experiments showing contrasted conditions in terms of soil, climate, cultivars, sowing density and sowing machine.

2.2 Extracting plant coordinates from digital photos

The methods used to measure the plant distribution along the row are relatively tedious (Van der Heijden et al. 2007). Electromagnetic digitizer describes plant structure from a sample of coordinates of plant elements within the canopy volume (Ma et al. 2007). However, it is very low throughput and is not adapted to conduct intensive field measurements. Alternatively, algorithms have been developed to measure the inter-plant spacing along the row for maize crops from top-view RGB (Red Green Blue) images (Tang and Tian 2008b, d). Improvements were then proposed by using 3D (Three dimensional) sensors (Jin and Tang 2009b; Nakarmi and Tang 2012a; Nakarmi and Tang 2014). However, these algorithms were only validated on maize crops that show relatively simple plant architecture with generally fixed plant spacing along the row.

In this work, plant coordinates along the row were measured using high resolution RGB images. A Sigma SD14 RGB camera with a resolution of 4608 by 3072 pixels was installed on a light moving platform (**Fig. 1a**). The camera was oriented at 45° inclination perpendicular to the row direction and was focused on the central row from a distance of about 1.5 m (**Fig. 1a**). The 50 mm focal length allowed sampling about 0.9 m of the row. Images were recorded along the row with at least 30% overlap to stitch the images together and describe the row over few meters length (Brown and Lowe 2006). For each setup, one picture was taken over a chessboard panel placed on the soil surface to calibrate the image: the transformation matrix derived from the chessboard image was applied to all the images acquired within the same setup. It allows removing perspective effects and scaling the pixels projected on soil surface. Finally, the coordinates of the plant are

interactively retrieved from the photos displayed on the screen. It corresponds to the position of the intersection between the bottom of the plant and the soil surface. To ensure that the horizontal direction of the image (**Fig. 1b**) is actually parallel to the row direction, the center row line was computed by fitting a simple linear regression on the plant coordinates. The deviation from the horizontal direction was corrected by applying the corresponding rotation. The plant coordinates along and across the row correspond respectively to the X and Y axes. The n^{th} plant position was notated $Plant_n$ $(\boldsymbol{x_n}, \boldsymbol{y_n})$, and $\Delta\boldsymbol{x_n}$ denotes to the plant spacing, i.e. the distance between $Plant_n$ and $Plant_{n-1}$ along X axis and $\Delta\boldsymbol{y_n}$ denotes to the deviation from the row position along Y axis (**Fig. 1b**). The uncertainties in plant position measurements have been evaluated to be around 1 to 2 mm mainly corresponding to the accuracy of the identification of the bottom of the plants on the image.

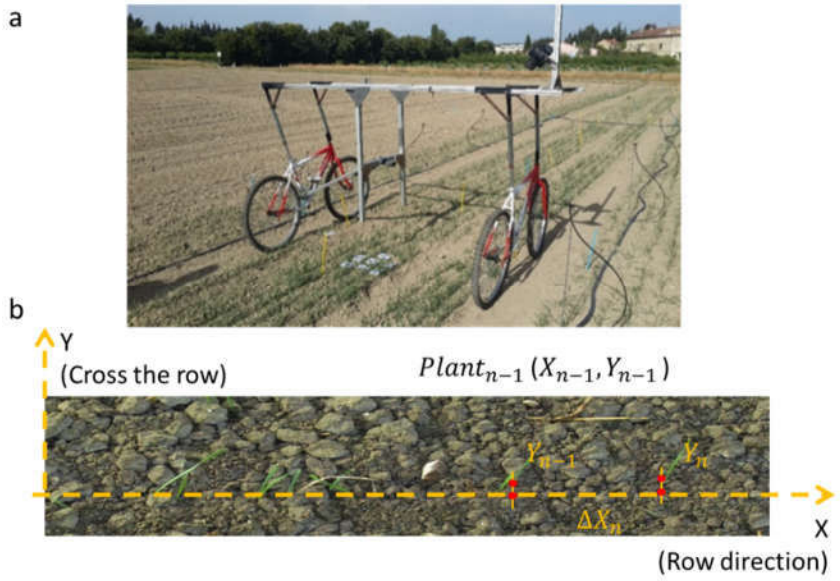

Fig. 1 a) the moving platform used to take the images in the field; b) notation for plant coordinates

2.3 Simulations of wheat canopy architecture and green fraction computation

3D virtual wheat canopies were generated running ADEL-Wheat. The ADEL-Wheat model simulates the architecture of single plants and then arranges them to pre-defined positions along the rows. As a matter of fact, the main assumption made here is that there is no interaction between neighbor plants: they are supposed to grow independently. All the details on the model and the other assumptions and parameters can be found in Abichou et al. (2013b) and Fournier et al. (2003). The model is implemented under the open source OpenAlea platform (Pradal et al. 2008). ADEL-Wheat was already validated across various conditions (Dornbusch et al. 2011b; Evers et al. 2005a). The parameters used to describe the plants were derived from Abichou et al. (2013b). The plants were simulated at three different development stages corresponding to 100 °C · day (around Haun stage 1.25), 500 °C · day (around Haun stage 4.9) and 900 °C · day (around Haun stage 8.6). Focus was put on the early stages when senescence is negligible and interactions between plants are likely to be limited in agreement with the main assumption made when simulating the canopy structure. For the later stages, plants would adapt their architecture according to their local environment. The simulated scenes correspond to a 2 m x 2 m square area. Three densities were considered: 200 plt/m², 300 plt/m² and 400 plt/m². The row spacing was set to 17.5 cm. Four sowing patterns along the row were considered (**Fig. 2**): 1) Pattern 'XY' refers to the actual case with coordinates extracted from the measurements; 2) Pattern 'X' only keeps the variation of inter-plant distance along the X direction and aligns plants on the row; 3) Conversely, pattern 'Y' maintains the deviation in the Y direction and sets equal plant spacing; 4) Finally pattern 'Regular' aligns the plants along the row with equal plant spacing.

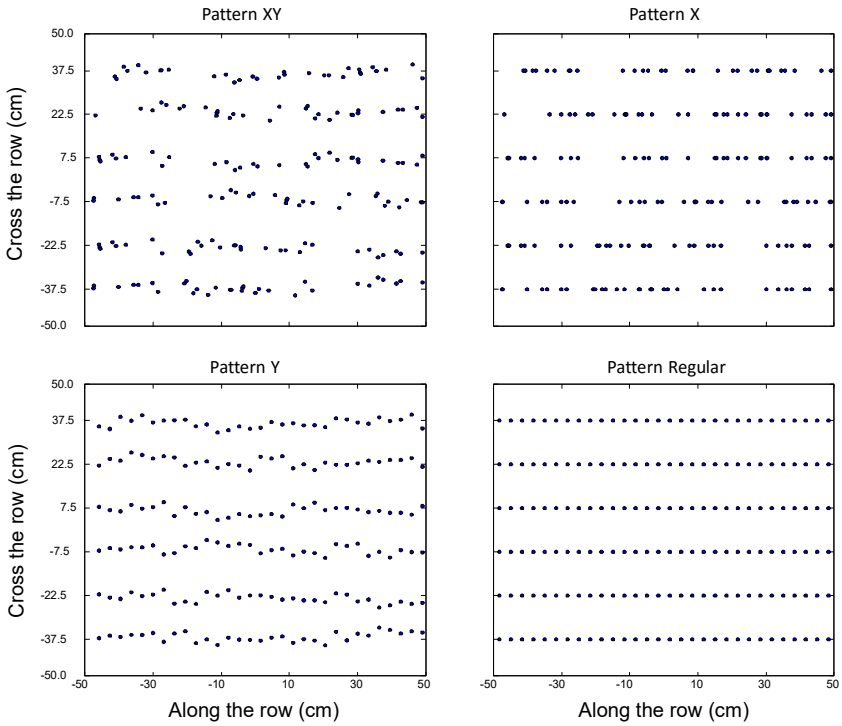

Fig. 2 The four patterns considered illustrated over a square meter canopy with 200 plt/m².

To achieve green fraction measurements, the images were rendered with Povray (Persistence of Vision Pty. Ltd. http://www.povray.org/) on the 3D model at resolution of 2000 by 2000 pixels from different camera configurations, including three azimuth angles (0°, 45° and 90°), and six zenith angles (0°, 15°, 30°, 45°, 57.5° and 75°) (**Fig. 3**). An orthographic projection was used to avoid perspective effects. The rendering process was replicated six times, corresponding to six locations along the central row. This allows avoiding possible border effects. Fig. 3 shows examples of rendered canopies. The green fraction (GF) is easily evaluated by counting the fraction of green pixels in the scene.

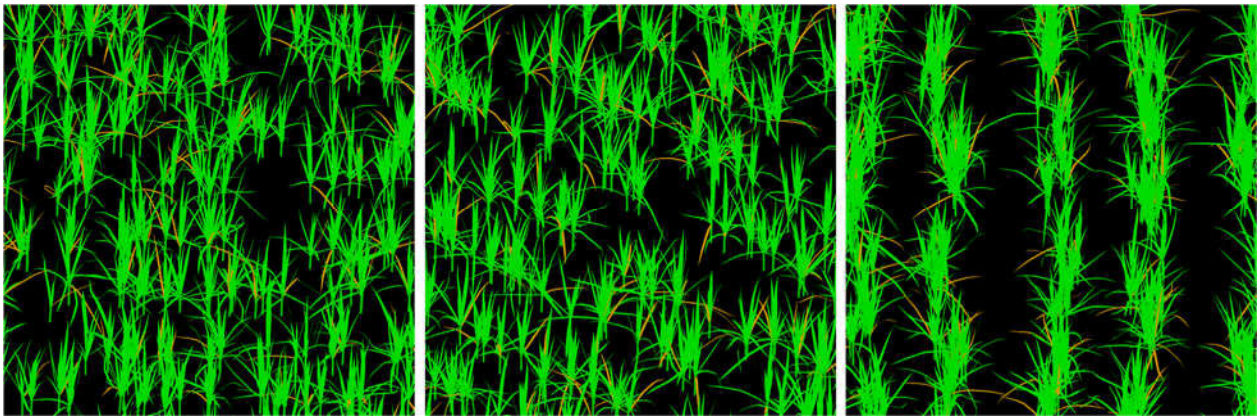

Fig. 3 Canopies simulated at 500 ℃ · day with 200 plt/m² rendered from 57.5° view zenith angle under different azimuth angles: perpendicular to the row (left), 45° to the row (middle) and parallel to the row (right). The plants are distributed along the row according to the actual pattern measured. The size of this scene is 2 m by 2 m.

3 Results

3.1 Modeling the plant distribution on the row

The spatial distribution of plants can be described by the variability of plant spacing along the row (Δx_i) and that of the deviation across the row (Δy_i) as well as the row spacing (r). In this study, the row spacing will be considered fixed and perfectly known. As both Δx_i and Δy_i can be considered as stochastic variables, knowledge about possible spatial dependency is mandatory to model the actual plant distribution.

3.1.1 Distribution of plant spacing along the row

The autocorrelation technique was used to test if Δx_{n-m} and Δx_n are linearly correlated when the lag m varies. Results observed over the Toulouse experiment show (**Fig. 4)** that the autocorrelation coefficient of inter-plant distance is not significant at 95% confidence interval. Similar results are observed for the Grignon and Avignon experiments. It is therefore concluded that there is no spatial correlation at the significance level of 0.05 between the positions of plants along the row direction: each observation Δx_i is considered as one independent realization of the random variable ΔX_i.

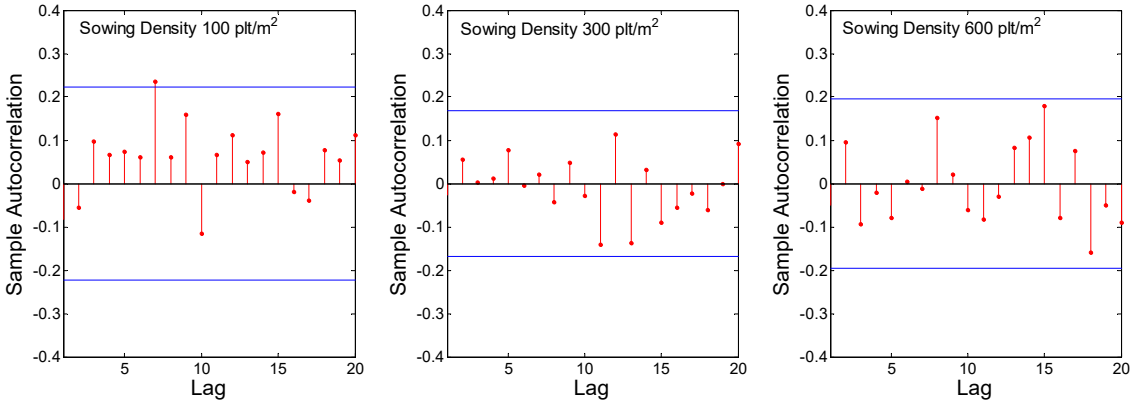

Fig. 4 The autocorrelation of plant spacing along the row direction illustrated with the three plant densities as observed over the Toulouse experiment. Lag denotes the shift in number of plants along the row direction. The upper and lower horizontal line represent the bound of 95% confidence interval around 0. Samples of 80 continuous plants were considered for each plant density. Lags 1 to 20 are only presented.

Distribution of plant spacing (**Fig. 5**) is right-skewed. As expected, the median and mean of the distribution of the plant spacing decrease when plant density increases (**Table 1**). The variance of plant spacing expressed as a coefficient of variation (CV) is relatively stable across the 14 cases, with CV≈0.95.

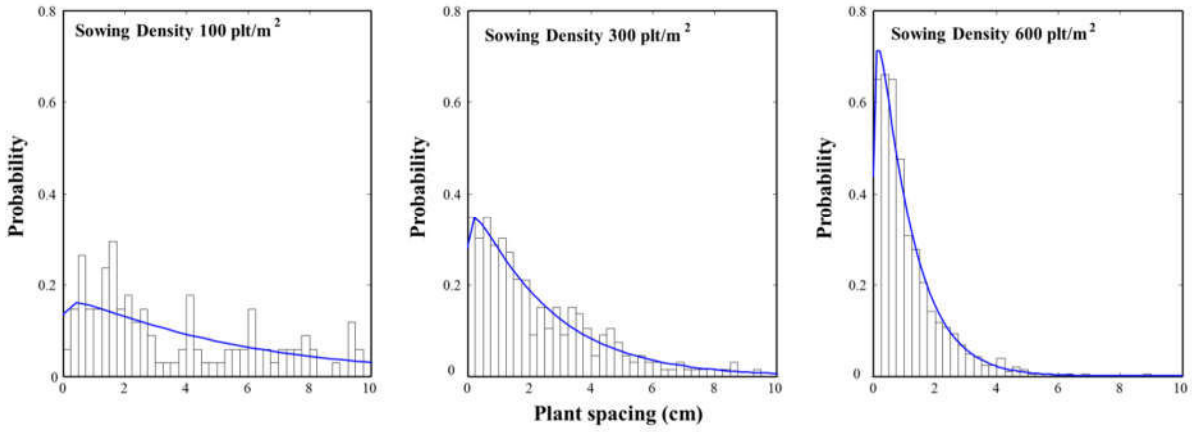

Fig. 5 Histograms of plant spacing and the fitted gamma distribution (solid line) for three densities from the Toulouse experiment.

Table 1 Characteristics of the fitted gamma distribution of the plant spacing (Δx) from 14 treatments. The Sowing density indicated is the targeted one, different from the actual one.

Experimental sites	Target Sowing Density (plt/m²)	Actual Density (plt/m²)	Cultivar	a	b	Mean (cm)	Median (cm)	CV
Avignon	100	58	Apache	1.14	6.38	7.27	5.29	0.94
	200	112	Apache	1.25	4.04	5.06	3.8	0.89
	300	252	Apache	0.99	2.53	2.51	1.74	1.00
	400	448	Apache	0.96	1.5	1.43	0.98	1.03
Toulouse	100	103	Apache	1.07	5.01	5.36	3.81	0.97
	200	208	Apache	1.39	1.95	2.72	2.1	0.85

	300	252	Apache	1.08	2.27	2.45	1.75	0.96
	400	416	Apache	1.24	1.37	1.7	1.27	0.90
	600	650	Apache	1.16	0.96	1.11	0.81	0.93
Grignon	150	125	Premio	1.12	3.37	3.76	2.72	0.95
	150	160	Attlass	1.13	2.48	2.81	2.04	0.94
	150	156	Flamenko	1.11	3.3	3.68	2.65	0.95
	150	122	Midas	1.24	3.03	3.75	2.81	0.90
	150	168	Koréli	1.15	2.89	3.32	2.42	0.93

Chi-square tests were conducted to evaluate the pertinence of the gamma distribution to describe the actual variability of plant spacing:

$$f(\Delta x|a,b) = \frac{1}{b^a \Gamma(a)} \Delta x^{a-1} e^{\frac{-\Delta x}{b}} \quad \Delta x, a, b \in \mathrm{R}^+ \text{ (1)}$$

Where a and b represent respectively the shape and scale parameters. The Chi-square test computed at the 5% significance level demonstrates the pertinence of the gamma function to describe the actual distribution of plant spacing for all the 14 cases considered. The shape parameter, a, is quite stable varying from 0.96 to 1.39 (**Table 1**), while the scale parameter, b, shows wider range of variation, between 0.96 and 6.38 and decreases generally with the plant density.

3.1.2 Distribution of plant deviation from the row

The plant positions seem organized along a wave-like pattern centered on the row line (**Fig. 6**). This indicates spatial correlation between the cross row deviation (Y) and the along row position (X) that was confirmed by the experimental variogram. The exponential variogram model (Chilès and Delfiner 2009) matches well the observations as demonstrated in **Fig. 6** and proves the significant spatial dependency of the deviation from the row direction (Δy) as a function of the position along the row (x). The exponential model writes:

$$\gamma(x) = \big(\gamma(\infty) - \gamma(0)\big)\left(1 - e^{\frac{-x}{range}}\right) + \gamma(0) \qquad (2)$$

With $\gamma(\infty)$ and $\gamma(0)$ being respectively the sill and nugget values. Results show that the nugget is negligible and that higher plant density decreases the range and reduces the sill (**Fig. 6**).

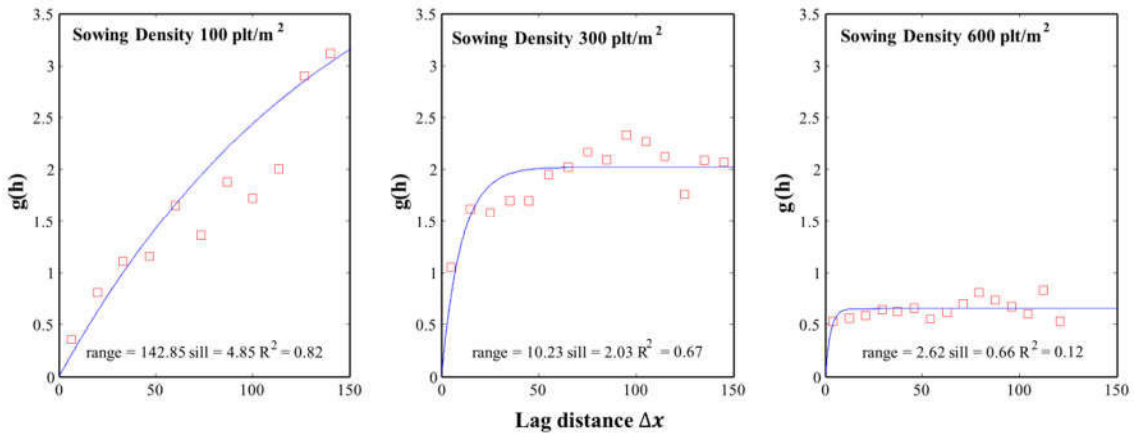

Fig. 6 Variogram of the deviation across the row direction as a function of the lag distance along the row direction. The red square symbols correspond to the observed variance, and the blue curve to the best fit of the exponential model.

The distribution of the deviation from the row of plant position shows a typical bell-shape almost symmetric around zero for all the plots (**Fig. 7**). It was expected that all samples in the same experimental site belong to the same distribution law since a variation in plant density should only affect the plant spacing. Chi-square test at confidence level of 95% confirmed that the deviation follows a Gaussian distribution except for the cases from the Avignon experiment. Closer inspection of the values of the distribution shows that parameters are very similar between locations (**Table 2**). This indicates that the 14 cases may be assumed to be samples coming from the same population. Note that the average location (μ) is always close to 0.00 by construction since the quantity considered is the deviation, $\Delta \boldsymbol{y}$, from the row direction defined by the best linear fit across the plant position points of the row.

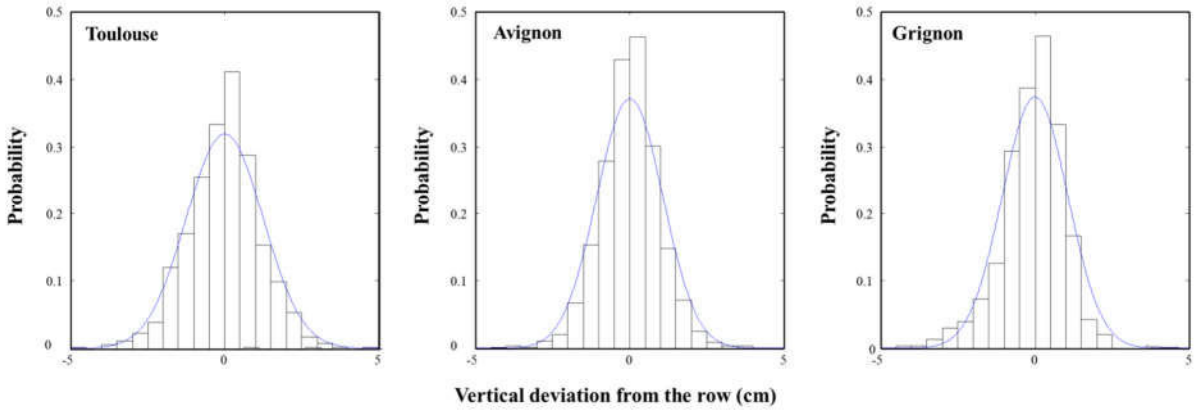

Fig. 7 Histograms of deviation across the row and the fitted Gaussian distribution curve for three experimental sites where all densities were merged.

Table 2 parameters and properties of fitted Gaussian distribution to characterize the deviation distance (ΔY) from the row direction over the three sites considered.

Experimental sites	μ	σ	Mean(cm)	σ (cm)
Avignon	0.00	1.25	0.00	1.25
Grignon	0.00	1.08	0.00	1.08
Toulouse	0.00	1.07	0.00	1.07

3.2 Simulating plant positions

The model made of the gamma distribution that describes the plant spacing on the row, the Gaussian model describing the distribution of the deviation of plant position from the row and the exponential variogram model representing the dependency of the deviation from the row requires 6 parameters as outlined in **Table 3**.

Table 3 The parameters of the model describing the plant position for a row

Symbol	Unit	Meaning	Model used
a	-	Shape parameter	gamma model for distribution along X direction
b	cm^{-1}	Scale parameter	
$range$	cm	Range parameter	Exponential variogram model for Y positions
$\gamma(\infty)$	cm^2	Sill parameter	
σ	cm	Dispersion parameter	Gaussian model for Y positions
ν	-	Shape parameter	

The simulation of plant positions starts by generating the vector, $Plants\_X$, of X positions along the row for n plants. It is achieved by randomly drawing n realizations of Δx from the fitted gamma distribution (**Eq. 1**):

$$Plants\_X = (x_1, \dots, x_n) \qquad (3)$$

Note that according to our results, no spatial dependency was taken into account for the X direction. Then the Gaussian distribution was employed to compute the corresponding vector of across row deviation, $Plants\_Y$:

$$Plants\_Y = (y_1, \dots, y_n) = LW \qquad (4)$$

With $W = (w_1, \dots, w_n)$ being a vector of n elements randomly sampled from a Gaussian distribution $\aleph(0,1)$. L is the term that accounts for the interdependency of the Y positions. It is computed from $\varepsilon = L * L^T$ with ε being the covariance matrix:

$$\varepsilon = cov\big((y_1, \dots, y_n), (y_1, \dots, y_n)\big) = \begin{bmatrix} cov(y_1 - y_1) & \cdots & cov(y_1 - y_n) \\ \vdots & \ddots & \vdots \\ cov(y_n - y_1) & \cdots & cov(y_n - y_n) \end{bmatrix} \qquad (5)$$

For $1 < i, j < n,\ cov(y_i - y_j)$ represents the covariance at the inter-distance $(x_i - x_j)$, which is computed with the exponential variogram model (**Eq. 2**).

To illustrate the resulting simulations, the parameters experimentally adjusted over three contrasted plant densities observed in the Toulouse experiment were used. Results (**Fig. 8**) show similar wave-like pattern as observed in the field. The non-uniformity of plants spacing and spatial correlation are well represented. Further tests have shown that the parameters of the distributions adjusted over the simulations provide values consistent with those used as inputs in the forward model, partly validating the plant position simulation algorithm.

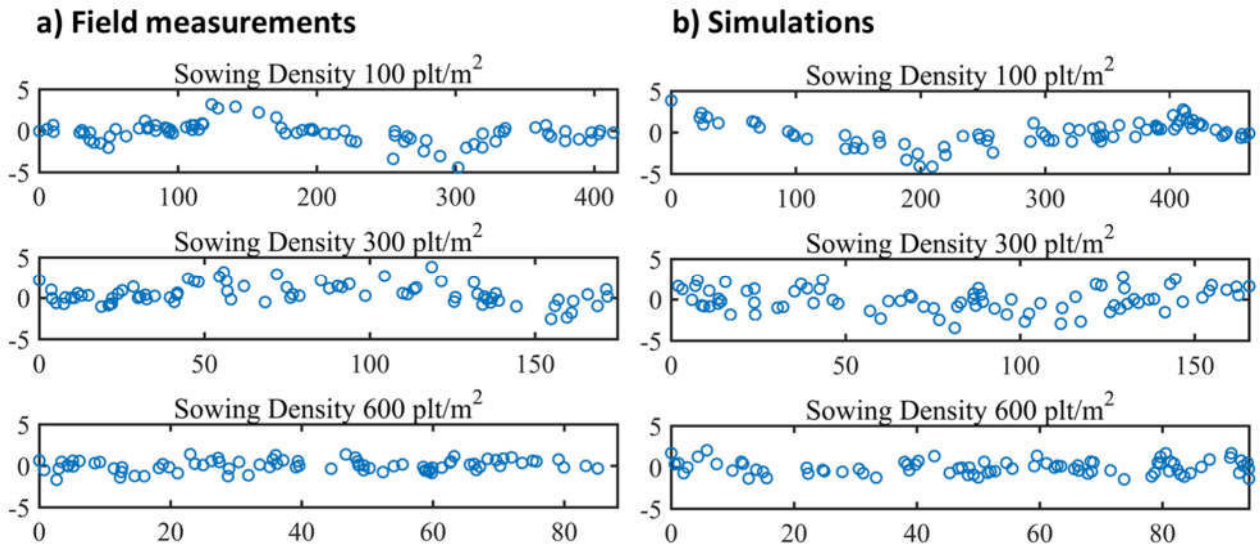

Fig. 8 a) actual plant distribution extracted from the images; b) simulated pattern with distribution laws. The three densities presented here are corresponding to the Toulouse site.

3.3 Consequences on the green fraction (GF)

The Povray rendering over the ADEL-Wheat simulations applied over the several sowing patterns were used to simulate the corresponding GF. Results (**Fig. 9**) show that GF varies as expected as a function of azimuthal (ϕ) and zenithal (θ) angles. GF increases as a function of the zenith angle because of the increasing optical depth. It also increases with azimuth angle because of the leaf clumping observed over the row direction. However, for zenith angle close to nadir ($0°\leq\theta\leq15°$), the azimuthal variation is limited. The azimuthal effects are the maximum in the $0°\leq\phi\leq45°$ range. For azimuths between $45°\leq\phi\leq 90°$, the azimuthal effects are then limited. These main patterns of GF variation as a function of zenith and azimuth directions were observed over all densities and sowing patterns.

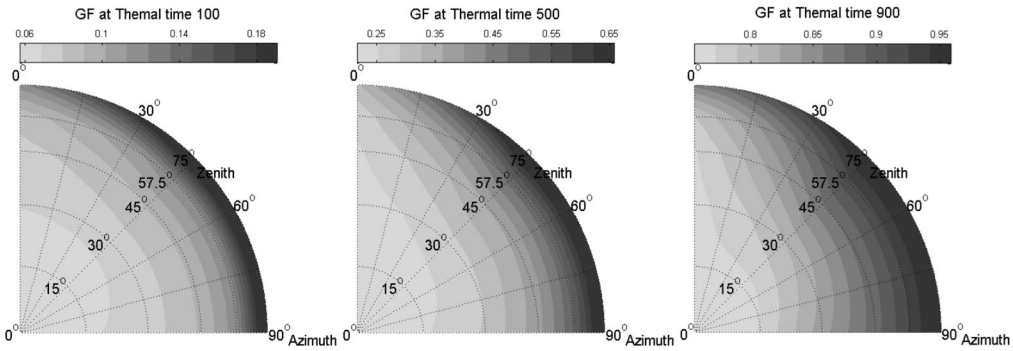

Fig. 9 Green Fraction at three development stages corresponding to thermal time 100, 500 and 900 $°C \cdot day$. Simulations were achieved over the actual sowing pattern (XY) with density 300 plt/m2. Azimuth ϕ=0° corresponds to the row direction. Note that the gray scale is adapted for each development stage displayed. Because of the symmetry in azimuth, only the first quadrant ($0°\leq\phi\leq90$) is presented.

The difference of GF values observed between the actual (XY) sowing pattern and the other considered cases (X, Y and Regular as described in **Fig. 2**) were investigated to document possible effects. Results (**Fig. 10**) show that the GF of patterns X and Regular for azimuth parallel to the row ($\phi\approx0°$) is systematically lower than the

actual one (XY). Conversely, pattern Y is very close to the actual case (XY). This is mainly explained by the difference in Y positions for which the plants aligned on the row for X and Regular patterns induce large leaf clumping on the row ($\phi \approx 0°$), therefore decreasing the GF. The effect increases as a function of the zenith angle because of the increase of the casting from one plant to the next one. For the later development stages, the effect is reduced because of the saturation (GF≈1.0). Conversely, for the earlier development stages, the effect is more pronounced for the higher plant densities, the clumping effect being enhanced in these conditions. The maximum effect can reach a difference larger than 0.1 for intermediate development stages and large zenith angles.

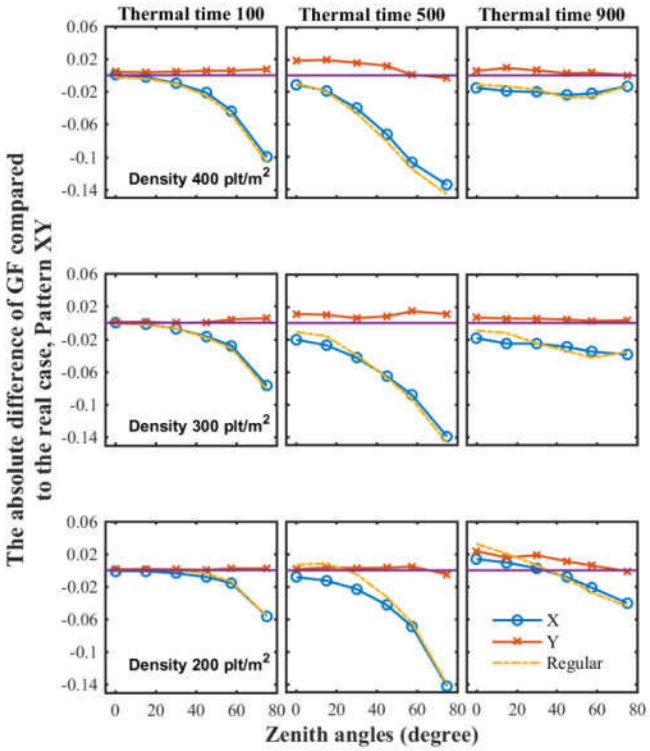

Fig. 10 Difference (absolute value) in GF between the actual pattern XY and the three other patterns (X, Y and Regular). Results of simulations for azimuth parallel (ϕ=0°), zenith angles zenith angles between 0 and 75° (X axis) and several plant densities (from top to bottom: 400, 300 and 200 plt/m²), and thermal time (from left to right: 100, 500 and 900 °C · day).

In the azimuthal direction perpendicular to the row ($\phi \approx 90°$) differences due to the sowing pattern are less pronounced (**Fig. 3**). They are almost negligible for the early stages (Thermal time 100 °C · day). For the intermediate development stage, an increase of GF as compared to the XY actual pattern is observed for the Y and in a lesser way R patterns. This is due to the regularity along the row direction introduced with patterns Y and R that minimizes the clumping between plants and thus increases the GF. The effect is more pronounced for the medium zenith angles since saturation (GF≈1.0) is likely to reduce the effect. The same applies for the late development stage for which the variability in GF according to the sowing pattern is limited. Note that some artifacts due to limited sampling (only 2 square meters were simulated) may explain the effects observed for thermal time=500 °C · day with 400 plt/m², and thermal time 900 °C · day with 200 plt/m².

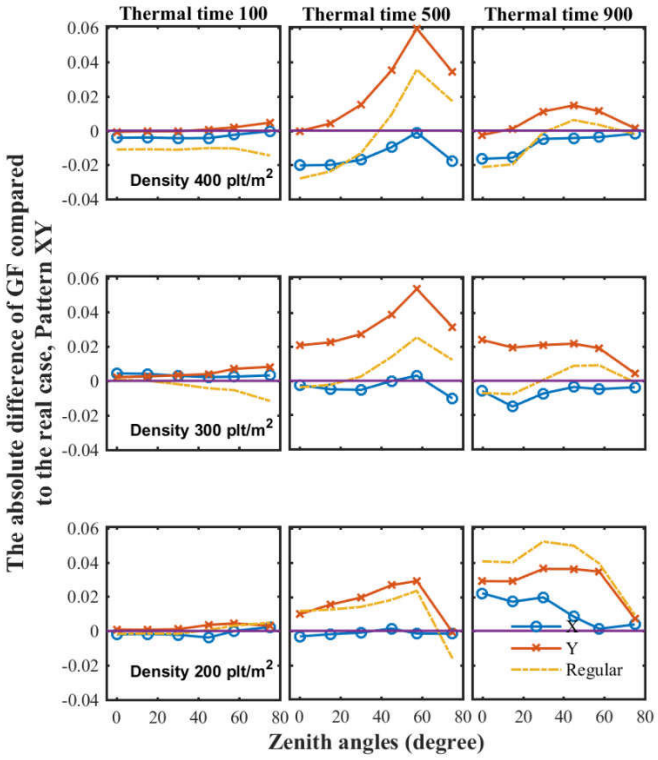

Fig. 3 Idem as in Fig. but for GF perpendicular to the row (ϕ=90°).

4 Discussion

This work investigates the plant spatial distribution pattern across a wide range of conditions in the case of wheat crops. This may be key information either to describe experimentally the possible impact of stand heterogeneity on canopy performances as well as for models simulating the canopy functioning and able to ideally 'grow' each plant depending on its local climate conditions, or at least to account for the bulk possible structure effect at the canopy scale.

The measurements of the sowing pattern were made using high resolution RGB cameras from which the location of the plants was measured around Haun stage 1.5 with an uncertainty around 1 to 2mm. The distribution of the distance between consecutive plants along the row direction was well described by a gamma distribution law with a scale and a shape parameters. Since the shape parameter was found to be generally close to 1.0 across conditions, obviously the variation of scale parameter mainly depends on the sowing density, thus the plant spacing as the row spacing was considered as known. Although the gamma model was providing good performances for our cases, Weibull and Generalized Exponential model could have been alternatives to effectively fit the right-skewed distribution for other crops or different types of seed drills (Balasooriya and Abeysinghe 1994; Gupta and Kundu 2003, 2004). The positions along the row direction of consecutive plants were proved to be independent from the neighboring plants. In the direction perpendicular to the row, plants distribute symmetrically on both sides of the row line according to a Gaussian distribution law with about the same parameters for three experimental sites. This is probably due to the same type of seed drills used over each site. However, significant spatial correlations were found between the positions across the row with characteristics varying across the sites. Based on these findings, a model with 6 input parameters was proposed to describe the actual distribution of row-planted crops. The distribution of the plant position proposed in this study is based on a purely statistical approach. The equipment and sowing condition should be further considered to predict the parameters of the statistical model that was presented. Ultimately, accounting for all these factors may lead to the development of a more mechanistic description of the plant position pattern. Coupling with FSPMs, it will consequently generate a comprehensive mechanism model in the simulation of plant competition. It could be also combined with a weed spatial distribution model to evaluate the competition for resources between species.

This can be used to design optimal plant distribution patterns and weed management systems (Colbach et al. 2014; Evers and Bastiaans 2016).

This study reveals that the plant distribution pattern impacts generally marginally the green fraction and thus the light interception efficiency. For the early development stages, no impact is expected for directions perpendicular to the row direction since plants' positions can be considered as independent for the several sowing patterns considered here. For the late development stages, the saturation of the canopy observed for the higher GAI values prevents the sowing pattern to impact the light interception at the stand scale. However, additional investigations should be carried out to verify this assertion at the organ level. At the canopy scale, the most important feature that impacts the light interception by more than 0.10 for intermediate developments stages is the deviation of plants positions across the row direction when the sun is parallel to the rows and with large zenith angles: under these conditions, the deviation of plant position from the row reduces the clumpiness of leaves as compared to stands with plants exactly aligned on the row. This will decrease the competition for light (and probably other resources in the soil) between the plants of the crop while increasing the competition with weeds. Although the deviation across the row appears to be a positive factor, most seed drilling systems do not actually control this feature. Investigations should therefore focus on this aspect by designing sowing systems providing a better control of the distribution of the seeds across the row, with hopefully some possible negative effects on weed development.

The simulations of canopy architecture for the several sowing patterns considered were achieved using Adel-Wheat model coupled with the spatial distribution pattern model developed in this study. However, like most FSPMS, the possible adaptation of individual plant architecture to their local environment, i.e. their plasticity, is currently ignored (Chelle 2005; Dornbusch et al. 2011a). The impact of plant plasticity on canopy level properties should therefore be investigated, including possible mechanisms for shade avoidance responses to neighboring plants as proposed by (Evers and Bastiaans 2016). As a consequence, it may modify our current conclusions drawn on the effect of the sowing pattern at later stages that was not accounting for plant plasticity. Nevertheless, the plant spatial distribution model proposed here may be very useful as it provides the starting point for FSPM simulations from which the stand will develop differently depending on the assumptions on plant position heterogeneity on the row.

Reference

Abichou, M. et al., 2013. Re-parametrisation of Adel-wheat allows reducing the experimental effort to simulate the 3D development of winter wheat, Proceedings of the 7th International Conference on Functional-Structural Plant Models, pp. 304–306.

Balasooriya, U. and Abeysinghe, T., 1994. Selecting between gamma and Weibull distributions—an approach based on predictions of order statistics. Journal of Applied Statistics, 21: 17-27.

Baret, F., Andrieu, B., Folmer, J., Hanocq, J. and Sarrouy, C., 1993a. Gap fraction measurement from hemispherical infrared photography and its use to evaluate PAR interception efficiency. Crop structure and light microclimate: 359-371.

Baret, F., Andrieu, B. and Steven, M., 1993b. Gap frequency and canopy architecture of sugar beet and wheat crops. Agricultural and Forest Meteorology, 65(3): 261-279.

Berger, U., Piou, C., Schiffers, K. and Grimm, V., 2008. Competition among plants: concepts, individual-based modelling approaches, and a proposal for a future research strategy. Perspectives in Plant Ecology, Evolution and Systematics, 9(3): 121-135.

Bradford, K.J., 2002. Applications of hydrothermal time to quantifying and modeling seed germination and dormancy. Weed Science, 50: 248–260.

Brown, M. and Lowe, D.G., 2006. Automatic Panoramic Image Stitching using Invariant Features. International Journal of Computer Vision, 74: 59-73.

Chelle, M., 2005. Phylloclimate or the climate perceived by individual plant organs: what is it? How to model it? What for? The New Phytologist, 166: 781-790.

Chilès, J.-P. and Delfiner, P., 2009. Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sons, 718 pp.

Colbach, N., Collard, A., Guyot, S.H., Mézière, D. and Munier-Jolain, N., 2014. Assessing innovative sowing patterns for integrated weed management with a 3D crop: weed competition model. European Journal of Agronomy, 53: 74-89.

Comar, A. et al., 2012. A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: description and first results. Journal of Functional Biology.

Costes, E., Lauri, P.E., Simon, S. and Andrieu, B., 2012. Plant architecture, its diversity and manipulation in agronomic conditions, in relation with pest and pathogen attacks. European Journal of Plant Pathology, 135: 455-470.

Deen, W. et al., 2003. An evaluation of four crop: weed competition models using a common data set. Weed Research, 43(2): 116-129.

Dornbusch, T. et al., 2011a. Plasticity of winter wheat modulated by sowing date, plant population density and nitrogen fertilisation: Dimensions and size of leaf blades, sheaths and internodes in relation to their position on a stem. Field Crops Research, 121: 116-124.

Dornbusch, T., Watt, J., Baccar, R., Fournier, C. and Andrieu, B., 2011b. A comparative analysis of leaf shape of wheat, barley and maize using an empirical shape model. Annals of Botany, 107: 865-873.

Duncan, W., 1984. A theory to explain the relationship between corn population and grain yield. Crop Science, 24(6): 1141-1145.

Evers, J.B. and Bastiaans, L., 2016. Quantifying the effect of crop spatial arrangement on weed suppression using functional-structural plant modelling. Journal of Plant Research: 1-13.

Evers, J.B. et al., 2005. Towards a generic architectural model of tillering in Gramineae, as exemplified by spring wheat (Triticum aestivum). New Phytologist, 166: 801–812.

Feng, L. et al., 2014. Comparing an empirical crop model with a functional structural plant model to account for individual variability. European Journal of Agronomy, 53: 16-27.

Fournier, C., Andrieu, B., Ljutovac, S. and Saint-Jean, S., 2003. ADEL-wheat: a 3D architectural model of wheat development. Proceedings of the 2003 Plant Growth Modeling, Simulation, Visualization, and Applications: 54–63.

Garin, G. et al., 2014. A modelling framework to simulate foliar fungal epidemics using functional-structural plant models. Annals of Botany, 114: 795-812.

Gigot, C., Vallavieille-Pope, C.d., Huber, L. and Saint-Jean, S., 2014. Using virtual 3-D plant architecture to assess fungal pathogen splash dispersal in heterogeneous canopies: a case study with cultivar mixtures and a non-specialized disease causal agent. Annals of Botany: mcu098.

Gupta, R.D. and Kundu, D., 2003. Discriminating between Weibull and generalized exponential distributions. Computational Statistics & Data Analysis, 43: 179-196.

Gupta, R.D. and Kundu, D., 2004. Discriminating between gamma and generalized exponential distributions. Journal of Statistical Computation & Simulation, 74: 107–121.

Jin, J. and Tang, L., 2009. Corn plant sensing using real-time stereo vision. Journal of Field Robotics, 26(6-7): 591-608.

Karayel, D., Wiesehoff, M., Özmerzi, A. and Müller, J., 2006. Laboratory measurement of seed drill seed spacing and velocity of fall of seeds using high-speed camera system. Computers and Electronics in Agriculture, 50: 89-96.

Liu, W., Tollenaar, M., Stewart, G. and Deen, W., 2004. Within-row plant spacing variability does not affect corn yield. Agronomy Journal, 96(1): 275-280.

López-Lozano, R., Baret, F., Chelle, M., Rochdi, N. and España, M., 2007. Sensitivity of gap fraction to maize architectural characteristics based on 4D model simulations. Agricultural and Forest Meteorology, 143: 217-229.

Ma, Y. et al., 2007. Parameter Optimization and Field Validation of the Functional-Structural Model GREENLAB for Maize at Different Population Densities. Annals of Botany, 101: 1185-1194.

Nakarmi, A. and Tang, L., 2012. Automatic inter-plant spacing sensing at early growth stages using a 3D vision sensor. Computers and electronics in agriculture, 82: 23-31.

Nakarmi, A.D. and Tang, L., 2014. Within-row spacing sensing of maize plants using 3D computer vision. Biosystems Engineering, 125: 54-64.

Olsen, J., Kristensen, L. and Weiner, J., 2006. Influence of sowing density and spatial pattern of spring wheat (Triticum aestivum) on the suppression of different weed species. Weed Biology and Management, 6(3): 165-173.

Olsen, J., Kristensen, L., Weiner, J. and Griepentrog, H.W., 2005. Increased density and spatial uniformity increase weed suppression by spring wheat. Weed Research, 45(4): 316-321.

Orzolek, M.D. and Daum, D.R., 1984. EFFECT OF PLANTING EQUIPMENT AND TECHNIQUES ON SEED GERMINATION AND EMERGENCE: A REVIEW. Journal of Seed Technology, 9: 99-113.

Pagano, E., Cela, S., Maddonni, G. and Otegui, M., 2007. Intra-specific competition in maize: Ear development, flowering dynamics and kernel set of early-established plant hierarchies. Field crops research, 102(3): 198-209.

Pradal, C., Dufour-Kowalski, S., Boudon, F., Fournier, C. and Godin, C., 2008. OpenAlea: a visual programming and component-based software platform for plant modelling. Functional Plant Biology, 35: 751-760.

Robert, C., Fournier, C., Andrieu, B. and Ney, B., 2008. Coupling a 3D virtual wheat (*Triticum aestivum*) plant model with a *Septoria tritici* epidemic model (Septo3D): a new approach to investigate plant–pathogen interactions linked to canopy architecture. Functional Plant Biology, 35: 997.

Rowse, H. and Finch-Savage, W., 2003. Hydrothermal threshold models can describe the germination response of carrot (Daucus carota) and onion (Allium cepa) seed populations across both sub- and supra-optimal temperatures. New Phytologist, 158: 101-108.

Tang, L. and Tian, L.F., 2008a. Plant Identification in Mosaicked Crop Row Images for Automatic Emerged Corn Plant Spacing Measurement. Transactions of the Asabe, 51(6): 2181-2191.

Tang, L. and Tian, L.F., 2008b. Real-time crop row image reconstruction for automatic emerged corn plant spacing measurement. Transactions of the Asabe, 51(3): 1079-1087.

Van der Heijden, G., De Visser, P.H.B. and Heuvelink, E., 2007. Measurements for functional-structural plant models. Frontis, 22: 13–25.

Verger, A., Vigneau, N., Chéron, C., Gilliot, J.M. and Baret, F., 2014. Green area index from unmanned aerial system over wheat and rapeseed crops. Remote Sensing of Environment, 152: 654-664.

Vos, J. et al., 2010. Functional–structural plant modelling: a new versatile tool in crop science. Journal of Experimental Botany, 61: 2101-2115.

Weiner, J., Griepentrog, H.W. and Kristensen, L., 2001. Suppression of weeds by spring wheat Triticumaestivum increases with crop density and spatial uniformity. Journal of Applied Ecology, 38(4): 784-790.

Wilson, J.M., 1980. The effect of release errors and the release point on the design of precision seed drills. Journal of Agricultural Engineering Research, 25: 407-419.

2.3. A method to estimate plant density and plant spacing heterogeneity: application to wheat crops

Shouyang Liu[a]*, Fred Baret[a], Denis Allard[b], Xiuliang Jin[a], Bruno Andrieu[c], Philippe Burger[d], Matthieu Hemmerlé[e] and Alexis Comar[e]

[a] *UMR EMMAH 1114, INRA, UAPV, 84914, Avignon, France*

[b] *UMR BioSP, INRA, UAPV, 84914, Avignon, France*

[c] *UMR ECOSYS, INRA, AgroParisTech, Université Paris-Saclay, 78850, Thiverval-Grignon, France*

[d] *UMR AGIR, INRA, INPT, 31326, Toulouse, France*

[e] *Hi-Phen, 84914, Avignon, France*

Abstract

Background: Plant density and its non-uniformity are key traits of interests to characterize the canopy at the emergence stage because they drive the competition among plants as well as with weeds.

Results: This study focuses on the sampling strategy requires estimating the plant density and heterogeneity along the row in wheat crops. Three experiments were conducted in 2014 resulting in 14 cases across varied sowing density, cultivars and environmental conditions. The coordinates of the plant along the row were measured over RGB high resolution images taken from the ground. Results show that the spacing between consecutive plants along the row direction are independent and follow a gamma distribution under the varied conditions experienced. A gamma count model was then derived to define the optimal sample size required to estimate plant density at a given precision. Results suggest that measuring the length of segments containing 90 plants will achieve a precision better than 10%, independent of the plant density. This approach appears more efficient than using fixed length segments on which the number of plants are counted: the optimal length for a given precision on the density estimation will depend on the actual plant density. The gamma count model parameters may also be used to quantify the heterogeneity of plant spacing along the row by exploiting the variability between replicated samples. Results show that to achieve a 10% precision on the estimates of the 2 parameters of the model, the measurements over elementary samples corresponding to the spacing between 2 consecutive plants should be repeated about 200 times.

Conclusions: This method allows optimizing the sampling strategy and also offering quantitative assessment of plant spacing heterogeneity.

Keywords

Wheat, Gamma-count model, Density, RGB imagery, Sampling strategy, Plant spacing heterogeneity

1 Background

Plant density at emergence is governed by the sowing density and the emergence rate. For a given plant density, the uniformity of plant distribution at emergence may significantly impact the competition among plants as well as with weeds (Olsen et al. 2006b; Olsen and Weiner 2007). Plant density and uniformity is therefore a key factor explaining production, although a number of species are able to compensate for low plant densities by a comparatively significant development of individual plants during the growth cycle. For wheat crops which are largely cultivated over the globe, tillering is one of the main mechanisms used by the plant to adapt its development to the available resources that are partly controlled by the number of tillers per unit area. The tillering coefficient therefore appears as an important trait to be measured. It is usually computed as the ratio of the number of tillers per unit area divided by the plant density (Reynolds et al. 2001b). Plant density is therefore one of the first variables measured commonly in most agronomical trials.

Crops are generally sown in rows approximately evenly spaced by seedling devices. Precision seedling systems mostly used for crops with plants spaced on the row by more than few centimeters (e.g. maize, sunflower or soybean) distribute seeds relatively evenly along the row. Conversely, for most crops with short distances among plants on the row, e.g. wheat, barley or canola, seeds are distributed non-evenly along the row. This can be attributed both to the mechanisms that free, at a variable frequency, the seed from the seed tank, and the trajectory of the seed that may also vary in the pipe that drives it from the seed tank to the soil. Further, once reaching the soil, the seed may also move with the soil displaced by the sowing elements penetrating the soil surface. Finally, some seeds may abort or some young plants may die because of pests or too extreme local environmental conditions (excess or deficit of moisture, low temperature etc.). The population density and its non-uniformity are therefore recognized as key traits of interests to characterize the canopy at the emergence stage. However, very little work documents the plant distribution pattern along the row, which is partly explained by the lack of dedicated device for accurate plant position measurement (Van der Heijden et al. 2007). Electromagnetic digitizers are very low throughput and not well adapted to such field measurements (Ma et al. 2007). Alternatively, algorithms have been developed to measure the inter-plant spacing along the row for maize crops from top-view RGB (Red Green Blue) images (Tang and Tian 2008a; Tang and Tian 2008c). Improvements were then proposed by using three dimensional sensors (Jin and Tang 2009a; Nakarmi and Tang 2012b; Nakarmi and Tang 2014). However, these algorithms were only validated on maize crops that show relatively simple plant architecture with generally fixed inter-plant spacing along the row.

Wheat is sown at relatively high plant density (150-400 seeds·m^{-2}) with sowing machines providing non-evenly distributed seeds along the row. Manual field counting in wheat crops is still extensively employed as the reference method. Measurements of plant population density should be completed when the majority of plants have just emerged and before the beginning of tillering when individual plants start to be difficult to be identified. Plants are counted over elementary samples corresponding either to a quadrat or to a segment (Norman 1995). The elementary samples need to be replicated in the plot to provide a more representative value (Marshall 1996). For wheat crops, (Reynolds et al. 2001b) suggested that at least a total of 3 m of rows (0.5 m segment length repeated 6 times) should be counted, while (Gate 1995) proposed to sample a total of 6 m (segments made of 2 consecutive rows by one meter repeated 3 times in the plot). (Whaley et al. 2000) proposed to repeat at least 4 times the counting in 0.25 m² quadrats corresponding roughly to a total of 6.7 m length of rows (assuming the rows spaced by 0.15 m). In this case, quadrats may be considered as a set of consecutive row segments with the same length when the quadrat is oriented parallel to the row direction or with variable lengths when the quadrat is oriented differently. Although these recommendations are simple and easy to apply, they may not correspond to an optimal sampling designed to target a given precision level. They may either provide low precision if under sampled or correspond to a waste of human resources in the opposite case.

The sample size required to reach a given precision of the plant density will depend on the population density and the heterogeneity of plant positions along the row that may be described by the distribution of the distances between consecutive plants. The distribution of the distance between consecutive plants provides

not only access to the plant density at the canopy level (Marshall 1996), but also to its local variation that may impact the development of neighboring plants as discussed earlier.

The objective of this study is to propose an optimal sampling method for plant density estimation and to quantify the heterogeneity of plant spacing along the row. For this purpose, a model is first developed to describe the distribution of the plants along the row. The model is then calibrated over a number of ground experiments. Further the model is used to compare several plant counting strategies and to evaluate the optimal sampling size to reach a given precision. Additionally, the model can also be exploited to design a method for quantifying the non-uniformity of plant distribution.

2 Methods

2.1 Field experiment

Three sites in France were selected in 2014 (**Table 1**): Avignon, Toulouse and Grignon. A mechanical seed drill was used in the three sites, which represents the standard practice for wheat crops. In Grignon, five plots were sampled, corresponding to different cultivars with a single sowing density. In Toulouse, five sowing densities were sampled with the same "Apache" cultivar. In Avignon, four sowing densities were sampled also with the same "Apache" cultivar. All measurements were taken at around 1.5 Haun stage (Haun 1973), when most plants had already emerged and was easy to identify individual plants visually. For wheat in France, this stage is reached approximately 10 to 14 days after the germination (Reynolds et al. 2001b). A total of 14 plots are thus available over the 3 sites showing contrasted conditions in terms of soil, climate, cultivars, sowing density and sowing machine, with however a fixed row spacing of 17.5 cm.

Table 1 The experiment design in 2014 over the three sites

Sites	Latitude	Longitude	Cultivar	Density (seeds·m^{-2})
Toulouse	43.5°N	1.5°E	Apache	100, 200, 300, 400, 600
Grignon	48.8°N	1.9°E	Premio; Attlass; Flamenko; Midas; Koréli	150
Avignon	43.9°N	4.8°E	Apache	100, 200, 300, 400

2.2 Image processing

For the 14 cases considered, plant coordinates along the row were measured using high resolution RGB images. A Sigma SD14 RGB camera with a resolution of 4608 by 3072 pixels was installed on a light moving platform (**Fig. 1**). The camera was oriented at 45° inclination perpendicular to the row direction and was focused on the central row from a distance of about 1.5 m (**Fig. 1**). The 50 mm focal length allowed to sample about 0.9 m of the row with a resolution at the ground level close to 0.2 mm. Images were recorded along the row with at least 30% overlap to allow stitching of the images together to describe three to five rows over 5 m length (Brown and Lowe 2006). For each treatment approximately 20 pictures were collected. For each setup, one picture was taken over a reference chessboard put on the soil surface to calibrate the image: the transformation matrix derived from the chessboard image was applied to all the images acquired within the same setup. It enables to remove perspective effects and to scale the pixels projected on the soil surface. Then the central row was extracted. The actual plant density expressed in plants per square meter horizontal ground (plants·m^{-2}) was computed simply as the number of plants counted on the segments, divided by the product of the length of the segments and the row spacing.

Fig. 1 The moving platform used to take the images in the field in 2014.

The coordinates of the plants were interactively retrieved from the photos displayed on the screen. They correspond to the position of the intersection between the bottom of the plant and the soil surface. To ensure that the horizontal direction of the image was actually parallel to the row direction, the center row line was computed by fitting a simple linear regression on the plant coordinates. The deviation from the horizontal direction was corrected by applying the corresponding rotation. For each of the 14 cases, the coordinates of at least 150 successive plants from the same row were measured along (X axis) and across (Y axis) of the row. The coordinates x_n of plant *n* (noted $Plant_n$) along the row axis allow to compute the spacing $\Delta x_n = (x_n - x_{n-1})$ between $Plant_n$ and $Plant_{n-1}$. The precision associated to the interactive measurement of the plant spacing was around 1 to 2 mm.

2.3 Development and calibration of the plant distribution model

2.3.1 Distribution of plant spacing

The autocorrelation technique was used to explore the spatial dependency of spacing between successive plants. It allows testing linear correlation between Δx_{n-m} and Δx_n where m is the lag. Results illustrated in **Fig. 2** over the Toulouse experiment show that the autocorrelation coefficient of inter-plant distance is not significant at 95% confidence interval. The same is observed over the other 13 cases acquired. It is therefore concluded that the positions among plants along the row direction are independent: each observation Δx could be considered as one independent realization of the random variable ΔX.

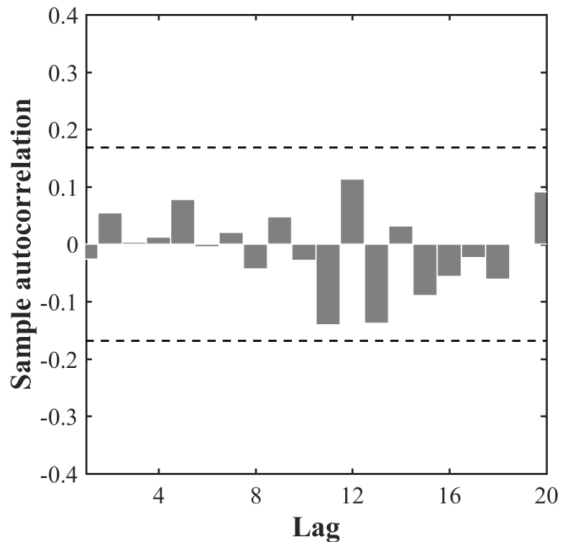

Fig. 2 The autocorrelation of the spacing among plants along the row direction illustrated with sowing density of 300 seeds·m-2 observed over the Toulouse experiment. The lag is expressed as the number of plant spacing between 2 plants along the row direction (X axis). Lags 1 to 20 are presented. The upper and lower horizontal line represent the 95% confidence interval around 0.

The distribution of the plant spacing is positively right-skewed (**Fig. 3**). A simple exponential distribution with only one scale parameter was first tentatively fitted to the data using a maximum likelihood method. However, the Chi-square test at the 5% significance level showed that the majority of the 14 cases do not follow this simple exponential distribution law. Weibull and gamma distributions are both a generalization of the exponential distribution requiring an extra shape parameter. Results show that Weibull and gamma distributions describe well (Chi-square test at the 5% significance level successful) the empirical distributions over the 14 cases (**Fig. 3**). However, the tail of the Weibull distribution tends toward zero less rapidly than that of an exponential distribution. Conversely, gamma distribution behaves the opposite with light-tail. Heavy-tailed distributions such as Weibull often show few samples with very large values (Papalexiou et al. 2013), increasing the risk of overestimation for the larger plant spacing. Hence, the gamma distribution was preferred to describe the actual distribution of plant spacing. The probability distribution function can therefore be written as:

$$f(\Delta x|a,b) = \frac{1}{b^a \Gamma(a)} \Delta x^{a-1} e^{\frac{-\Delta x}{b}} \quad \Delta x, a, b \in R^+ \qquad (1)$$

where a and b represent the shape and scale parameters respectively. The expectancy $E(\Delta X)$ and variance $Var(\Delta X)$ are simple expressions of the two parameters:

$$E(\Delta X) = a \cdot b \quad (2)$$

$$Var(\Delta X) = a \cdot b^2 \quad (3)$$

As a consequence, the coefficient of variation $CV(\Delta X) = \frac{\sqrt{var(\Delta X)}}{E(\Delta X)}$ is a simple function of the shape parameter:

$$CV(\Delta X) = 1/\sqrt{a} \quad (4)$$

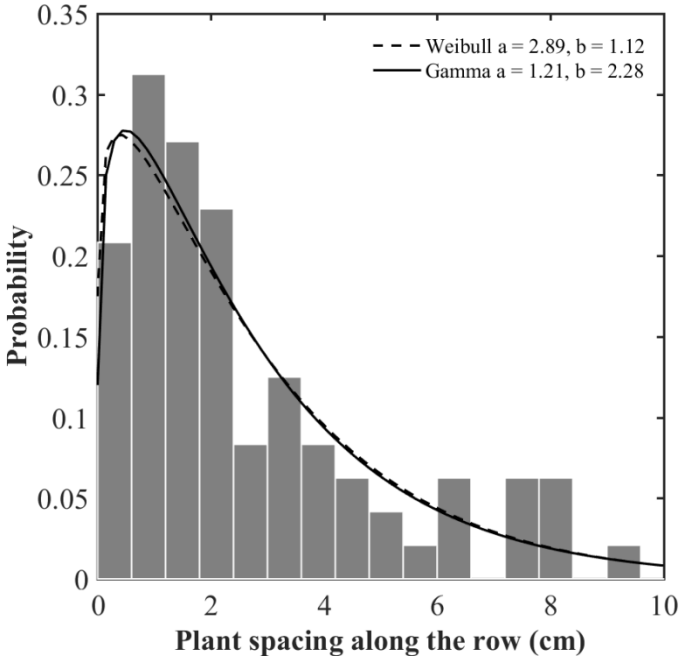

Fig. 3 Empirical histogram of the spacing along the row (gray bars). The solid (respectively dashed) line represents the fitted gamma (resp. Weibull) distribution. Case of the sowing density 300 seeds·m^{-2} observed over the Toulouse experiment. a and b represent the shape and scale parameters respectively.

2.3.2 Modeling the distribution of the number of plants per row segment

For the sake of simplicity, the row spacing, r, will be considered perfectly known. This assumption is reasonable since the row spacing is defined after tuning the sowing machine. This leads to uncertainties of few millimeters over the 175 mm row spacing distance. Therefore, for a given row spacing, the plant density depends only on the number of plants per unit linear row length. Estimating the number of plants within a given row length is a count data problem analogous to the estimation of the number of events during a specific time interval (Winkelmann 1995, 1996). Counts are common random variables that are assumed to be non-negative integer or continuous values representing the number of times an event occurs within a given spatial or temporal domain (Zeviani et al. 2014). The gamma-count model suits well our problem with intervals independently following a gamma distribution as in our case. The probability, $P\{N_l = n\}$, to get n plants over a segment of length l, writes:

$$P\{N_l = n\} = \begin{cases} 1 - IG\left(a, \frac{l}{b}\right) & \text{for } n = 0 \\ IG\left(a \cdot n, \frac{l}{b}\right) - IG\left(a \cdot n + a, \frac{l}{b}\right) & \text{for } n = 1,2, \ldots \end{cases} \quad (5)$$

where N_l is the number of plants over the segment of length l, and $IG\left(a \cdot n, \frac{l}{b}\right)$ is the incomplete gamma function:

$$IG\left(a \cdot n, \frac{l}{b}\right) = \frac{1}{\Gamma(a \cdot n)} \int_0^{l/b} t^{a \cdot n - 1}\, e^{-t} dt \quad (6)$$

where Γ is the gamma Euler function. The expectation and variance of the number of plants over a segment of length l is given by:

$$E(N_l) = \sum_{n=1}^{\infty} IG\left(a \cdot n, \frac{l}{b}\right) \quad (7)$$

$$Var(N_l) = \sum_{n=1}^{\infty}(2n-1)\, IG\left(a \cdot n, \frac{l}{b}\right) - \left[\sum_{n=1}^{\infty} IG\left(a \cdot n, \frac{l}{b}\right)\right]^2 \quad (8)$$

Finally, the expectation and variance of the plant density, D_l, estimated over a segment of length l can be expressed by introducing the row spacing distance, r, assumed to be known:

$$E(D_l) = \frac{E(N_l)}{l \cdot r} \quad (9)$$

$$Var(D_l) = \frac{Var(N_l)}{(l \cdot r)^2} \quad (10)$$

The expectation, $E(D_l)$, converges toward the actual density of the population when $l \to \infty$.

The transformed gamma-count model allows evaluating the uncertainty of plant density estimation as a function of the sampling size. The uncertainty can be characterized by the coefficient of variation (CV) as follows:

$$CV(D_l) = \frac{\sqrt{Var(D_l)}}{E(D_l)} = \frac{\sqrt{Var(N_l)}}{E(N_l)} \quad (11)$$

Several combinations of values of a and b may lead to the same plant density, but with variations in their distribution along the row (**Fig. 4**). The fitting of parameters a and b over the 14 cases using the transformed gamma-count model (**Eq. 9**) shows that the shape parameter, a, varies from 0.96 to 1.39 and is quite stable. Conversely, the scale parameter, b, appears to vary widely from 0.96 to 6.38, mainly controlling the plant density (**Fig. 4**). Since the CV depends only on the shape parameter a (**Eq. 4**), it should not vary much across the 14 cases considered. This was confirmed by applying a one-way analysis of variance on the CV values of the 14 cases available ($F = 1.09, P = 0.3685$): no significant differences are observed. This result may be partly explained by the fact that the same type of seed drill was used for all the three sites.

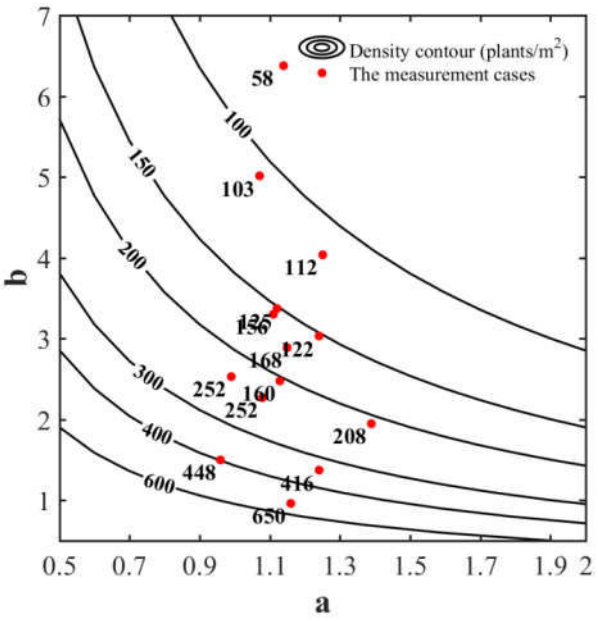

Fig. 4 Relationship between parameters a and b of the gamma-count model for a range of plant density (from Eq. 6-7 and Eq. 9). The lines correspond to, 100, 150, 200, 300, 400 and 600 plants·m-2. The red dots correspond to the actual density measured over the 14 cases considered.

The adjustment of the gamma-count model on the measured plant spacing using a maximum likelihood method provides an estimate of the plant density (**Eq. 9**). The comparison to the actual plant density (**Fig. 5**) simply computed as the number of plants per segment divided by the area of the segments (segment length by row spacing), shows a good agreement, with an RMSE around 50 plants·m^{-2} over the 14 cases available. The model performs better for the low density with a RMSE of 21 plants·m^{-2} for density lower than 400 plants·m^{-2}. These discrepancies may be mainly explained by the accuracy in the measurement of the position

of individual plants (around 1-2 mm). Uncertainties on individual plant spacing will be high in relative values as compared to that associated with the measurement of the length of the segment used in the simple method to get the 'reference' plant density. Hence it is obviously even more difficult to get a good accuracy in plant spacing measurements for high density, i.e. with a small distance among plants. In addition, small deviations from the gamma-count model are still possible, although the previous results were showing very good performance.

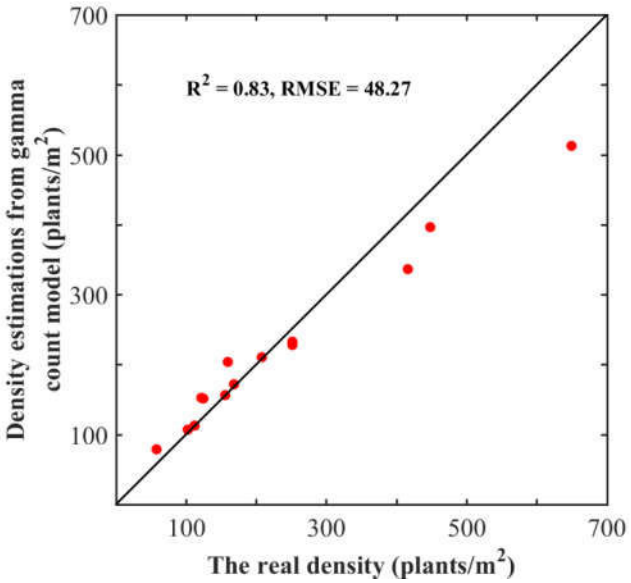

Fig. 5 Comparison between the actual density and that estimated from the gamma-count model

3 Results

3.1 Optimal sample size to reach a given precision for plant density estimation

The transformed gamma-count model provides a convenient way to investigate the effect of the sampling size on the precision of the density estimates. The precision will be quantified here using the coefficient of variation (CV). The sample size can be expressed either as a given length of the segments where the (variable) number of plants should be counted, or as a (variable) length of the segment to be measured corresponding to a given number of consecutive plants. The two alternative sampling approaches will be termed FLS (Fixed Length of Segment) for the first one, and FNP (Fixed Number of Plants) for the second one.

When considering the FLS approach, the sample size is defined by the length of segment, L, where plants need to be counted. The optimal L value for a given target precision quantified by the CV will mainly depend on the current density as demonstrated in **Fig. 6a**: longer segments are required for the low densities. Conversely, shorter segments are needed for high values of the plant density to reach the same precision. The scale parameter, b, that controls the plant density drives therefore the optimal segment length L (**Fig. 6a**). Counting plants over a $L = 5\ m$ (500 cm) provides a precision better than 10% for densities larger than 150 plants·m^{-2} for the most common conditions characterized by a shape coefficient $a > 0.9$. These figures agree well with the usual practice for plant counting as reviewed in the introduction (Gate 1995; Reynolds et al. 2001b; Whaley et al. 2000). Increasing the precision quantified by the CV will require longer segments L to be sampled (**Fig. 7a**).

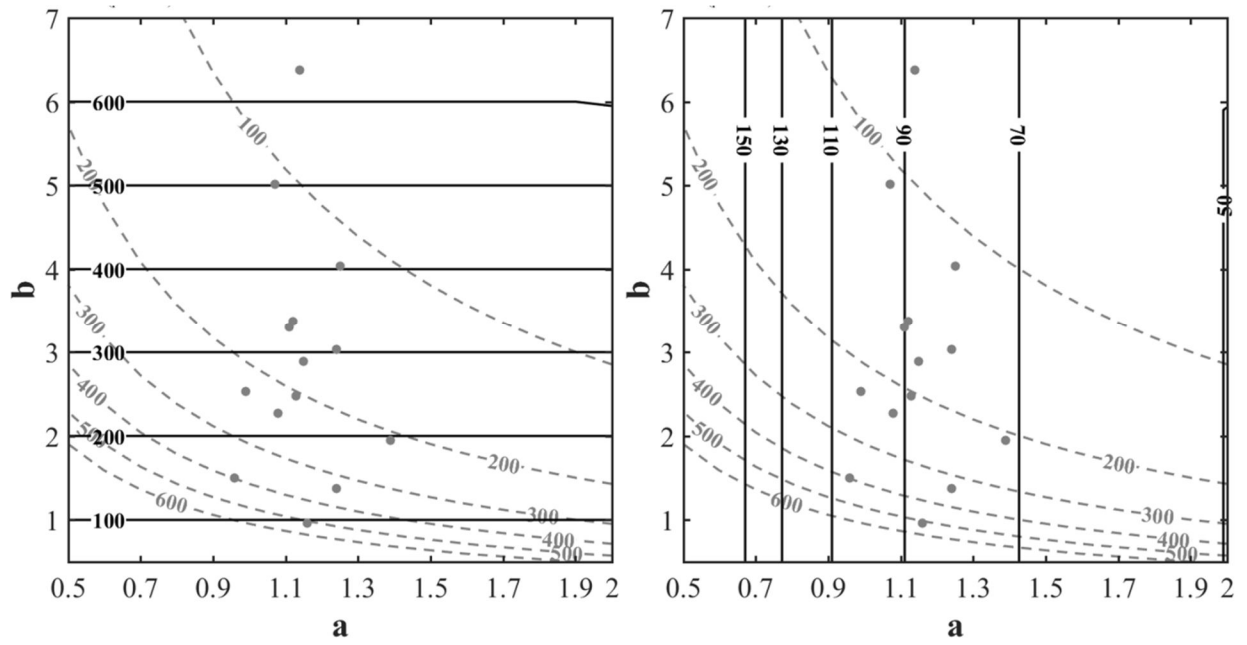

Fig. 6 (a): The optimal sampling size length (the horizontal solid lines, the length being indicated in cm) used in the FLS approach as a function of parameters a and b to get $CV = 10\%$ for the density estimation. (b): Idem as on the left but the sample is defined by the number of plants to be counted (the vertical solid lines with number of plants indicated) for the FNP approach. The gray dashed lines correspond to the actual plant density depending also on parameters a and b. The row spacing is assumed perfectly known and equal to 17.5 cm. The gray points represent the 14 cases measured.

When considering the FNP approach, the sample size is driven by the number, N, of consecutive plants that defines to a row segment whose length need to be measured. The simulations of the model (**Fig. 6b**) show that N is mainly independent from the plant density. For the 14 cases considered in this study, segments with $70 < N < 110$ plants should be measured to reach a precision of $CV = 10\%$. The shape parameter a influences dominantly the sample size: more heterogeneous distribution of plants characterized by small values of the shape parameter will require more plants to be counted (**Fig. 7b**). To increase the precision (lower CV), more plants will also need to be counted (**Fig. 7b**).

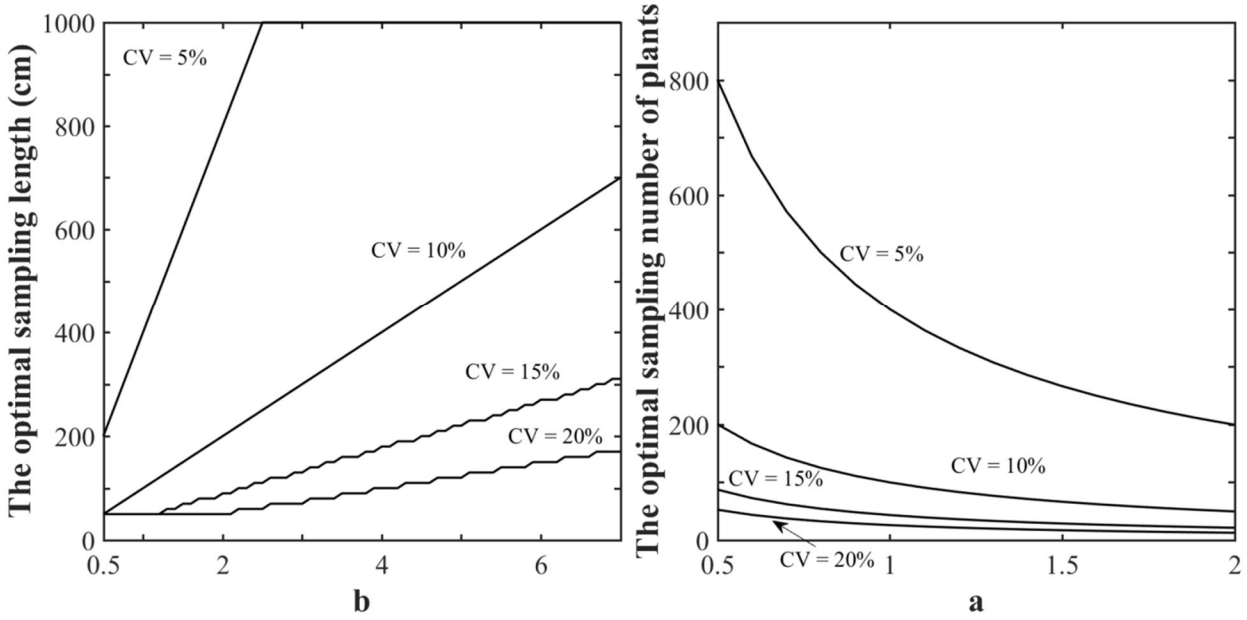

Fig. 7 The optimal sampling length for the FLS approach (a) and the number of plants for the FNP approach (b). The dominant parameter is used (the scale parameter for FLS and the shape coefficient for FNP). The precision is evaluated with the CV = 5%, 10%, 15% and 20%.

3.2 Sampling strategy to quantify plant spacing variability on the row

The previous sections demonstrated that the scale and shape parameters could be estimated from the observed distribution of the plant spacing. However, the measurement of individual plant spacing is tedious and prone to errors as outlined earlier. The estimation of these parameters from the variability observed between small row segments containing a fixed number of plants will therefore be investigated here. This FNP approach is preferred here to the FSL one because there will be no additional uncertainties introduced by the position of the first and last plants of the segment with the corresponding start and end of the segment. These uncertainties may be significant in the case of small segments in the FLS approach.

The probability distribution of a gamma distribution can be expressed as the sum of an arbitrary number of independent individual gamma distributions (Steutel and Van Harn 2003). This property allows to compute the distribution of a segment of length $\mathrm{L_n}$ corresponding to n plant spacing between $(n+1)$ consecutive plants with $\mathrm{L_n} = \sum_{i=1}^{n} \Delta x_i$, as a gamma distribution with $n \cdot a$ as shape parameter and the same scale parameter *b* as the one describing the distribution of ΔX.

$$\mathrm{L_n} \sim \mathrm{Gamma}(n \cdot a, b) \quad (12)$$

The parameters a and b will therefore be estimated by adjusting the gamma model described in **Eq. 12** for the given value of n.

The effect of the sampling size on the precision of a and b parameters estimation was further investigated. A numerical experiment based on a Monte-Carlo approach was conducted considering a standard case corresponding to the average of the 14 cases sampled in 2014 with $a = 1.10$ and $b = 2.27$. The sampling size is defined by the number of consecutive plants for the FNP approach considered here and by the number of replicates. For each sampling size 300 samples were generated by randomly drawing in the gamma distribution (**Eq. 12**) and parameters a and b were estimated. The standard deviation between the 300 estimates of a and b parameters was finally used to compute the corresponding CV. This process was applied to a number of replicates varying between 20 to 300 by steps of 10 and a number of plants per segment varying between 2 (i.e. spacing between two consecutive plants) to 250 within 12 steps. This allows describing the variation of the coefficient of estimated values of parameters a and b as a function of the number of replicates and the number of plants (**Fig. 8**).

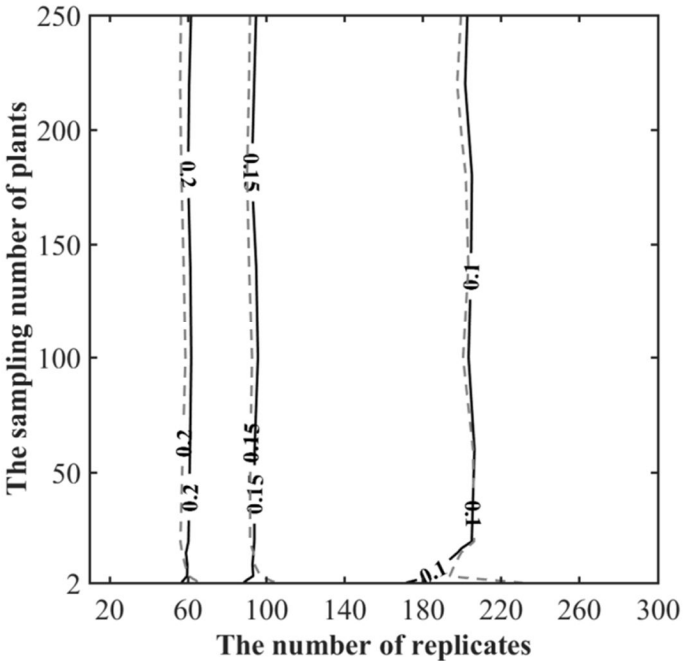

Fig. 8 Contour plot of the CV associated to the estimates of parameters a (solid line) and b (dashed line) as a function of the number of replicates of individual samples made of n plants (the y axis). The solid (respectively dashed) isolines correspond to the of CV of a parameter (respectively b) These simulations were conducted with $[a, b] = [1.10, 2.27]$.

Results show that the sensitivity of the CV of estimates of parameters a and b are very similar (**Fig. 8**). The sensitivity of parameters a and b is dominated by the number of replicates: very little variation of CV is observed when the number of plants per segment varies (**Fig. 8**). Parameters a and b require about 200 replicates independently from the number of plants per segment. It seems therefore more interesting to make very small segments to decrease the total number of plants to count.

Additional investigations not shown here for the sake of brevity, confirmed the independency of the number of replicates to the number of plants per segment when parameters a and b are varying. Further, the number of replicates need to be increased as expected when the shape parameter b decreases (i.e. when the plant spacing is more variable) to keep the same precision on estimates of a and b parameters.

4 Discussion and Conclusions

A method was proposed to estimate plant density and sowing pattern from high resolution RGB images taken from the ground. The method appears to be much more comfortable as compared with the standard outdoor methods based on plant counting in the field. Images should ideally be taken around Haun stage 1.5 for wheat crops when most plants have already emerged and tillering has not yet started. Great attention should be paid to the geometric correction in order to get accurate ortho-images where distances can be measured accurately. The processing of images here was automatic except the last step corresponding to the interactive visual extraction of the plants' coordinates in the image. However, recent work (Jin et al. 2016; Liu et al. 2016b) suggests that it will be possible to automatize this last step to get a fully high-throughput method.

The method proposed is based on the modeling of the plant distribution along the row. It was first demonstrated that the plant spacing between consecutive plants are independent which corresponds to a very useful simplifying assumption. The distribution of plant spacing was then proved to follow a gamma distribution. Although the Weibull distribution showed similar good performance, it was not selected because of the comparatively heavier tails of the distribution that may create artefacts. Further the Weibull model does not allow to simply derive the distribution law of the length of segments containing several consecutive plants (Rinne 2008). The gamma model needs a scale parameter that drives mostly the intensity of the process, i.e. the plant density, and a shape parameter that governs the heterogeneity of plant spacing. This model was transformed into a count data model to investigate the optimal sampling required to get an estimate of plant density for a given precision level.

The model proposed here concerns mainly relatively nominal sowing, i.e. when the sowing was successful on average on the row segments considered: portions of rows with no plants due to sowing problems or local damaging conditions (pests, temperature and moisture) should be taken into account separately. Similarly, portions of rows with excess of plants due to sowing problems should also be considered separately. This may eventually lead to a two scale model: a first scale describes the distribution of places with significant portions of missing plants or excess plants. The second scale will describe the distribution of plants on the row where the sowing was considered nominal and where the gamma distribution model may apply. In our experimental situations the sowing was considered as nominal on most of the fields. However, it will be possible to automatically identify from the images the nominal row segments from the places characterized by missing plants or excessive concentration of plants (Liu et al. 2016b).

This study investigated the sampling strategy to estimate the plant density. Focus was put on the uncertainty of density inference introduced by the non-uniformity of plant spacing along the row, corresponding to the sampling error. Non-sampling errors were not considered in this work. This is due to the difficulties in assessing the non-sampling errors which may come from measurement biases, non-randomness in the sample selection and so on (Lohr 2009; Särndal et al. 2003). Unlike sampling error, it could not be minimized by increasing sampling size but by combining a random sampling selection procedure and a measurement method ensuring high accuracy (Scheuren and Association 2004). Besides, the row spacing was perfectly known in this work. As a matter of fact, some uncertainties also exist in the row spacing in the range of 1 or 2 mm. This contributes to the non-sampling errors but negligible.

Optimal sampling requires a tradeoff between minimum sampling error obtained with maximum sampling size and minimum cost obtained with minimum sampling size (Biemer 2010). The optimal sampling strategy

should first be designed according to the precision targeted here quantified by the coefficient of variation (CV) characterizing the relative variability of the estimated plant density between several replicates of the sampling procedure. The term 'optimal' should therefore be understood as the minimum sampling effort to be spent to achieve the targeted precision. Two approaches were proposed: the first one considers a fixed segment length (FSL) over which the plants have to be counted; the second one considers a fixed number of successive plants (FNP) defining a row segment, the length of which needs to be measured. The first method (FLS) is the one generally applied within most field experiments. However, we demonstrated that it is generally sub-optimal: since the segment length required to achieve a given CV depends mainly on the actual plant density, the sampling is either too large for the targeted precision, or conversely too small, leading to possible degradation of the precision of plant density estimates. Nevertheless, for the plant density (>150 plants·m^{-2}) and shape parameter ($a > 0.9$) usually experienced, a segment length of 5 m will ensure a precision better than 10%. The second approach (FNP) appears generally more optimal: it aims at measuring the length of the segment corresponding to a number of consecutive plants that will depend mainly on the targeted precision. Results demonstrate that in our conditions, the density should be evaluated over segments containing 90 plants to achieve a 10% precision. The sampling size will always be close to optimal as compared to the first approach where optimality requires the knowledge of the plant density that is to be estimated. Further, the FNP approach is probably more easy to implement with higher reliability: as a matter of facts, measuring the length of a segment defined by plants at its two extremities is easier than counting the number of plants in a fixed length segment, where the extremities could be in the vicinity of a plant and its inclusion or not in the counting could be prone to interpretation biases by the technician. The total number of plants required in a segment could be split into subsamples containing smaller number of plants that will be replicated to get the total number of plants targeted. This will improve the spatial representativeness. Overall, the method proposed meets the requirements defined by (Cobb et al. 2013; Fiorani and Schurr 2013) for the next generation of phenotyping tools: increase the accuracy, the precision and the throughput while reducing the labor and budgetary costs.

The gamma-count model proved to be well suited to describe the plant spacing distribution along the row over our contrasted experimental situations. It can thus be used to describe the heterogeneity of plant spacing as suggested by (Liu et al. 2016a). This may be applied for detailed canopy architecture studies or to quantify the impact of the sowing pattern heterogeneity on inter-plant competition (Olsen et al. 2006b; Olsen and Weiner 2007). The heterogeneity of plant spacing may be described by the scale and shape parameters of the gamma model. Quantification of the heterogeneity of plant spacing requires repeated measurements over segments defined by a fixed number of plants. Our results clearly show that the precision on estimates of the gamma count parameters depends only marginally on the number of plants. Conversely, it depends mainly on the number of replicates to be measured. For the standard conditions experienced in this study, the optimal sampling strategy to get a CV lower than 10% on the two parameters of the gamma distribution would be to repeat 200 times the measurement of plant spacing between 2 consecutive plants.

Reference

Biemer, P.P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly, 74*, 817-848

Brown, M., & Lowe, D.G. (2006). Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision, 74*, 59-73

Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics, 126*, 867-887

Fiorani, F., & Schurr, U. (2013). Future Scenarios for Plant Phenotyping. *Annual Review of Plant Biology, 64*, 267-291

Gate, P. (1995). *Ecophysiologie du blé*. Tec & Doc Lavoisier

Haun, J. (1973). Visual quantification of wheat development. *Agronomy Journal, 65*, 116-119

Jin, J., & Tang, L. (2009). Corn plant sensing using real-time stereo vision. *Journal of Field Robotics, 26*, 591-608

Jin, X., Liu, S., Baret, F., Hemerlé, M., & Comar, A. (2016). Estimates of plant density from images acquired from UAV over wheat crops at emergence. *Remote sensing of the environment, submitted*

Liu, S., Baret, F., Andrieu, B., Abichou, M., Allard, D., Solan, B.d., & Burger, P. (2016a). Modeling the distribution of plants on the row for wheat crops: Consequences on the green fraction at the canopy level. *Computer and Electronic in Agriculture, Under review*

Liu, S., Baret, F., Andrieu, B., Burger, P., & Hemerlé, M. (2016b). Estimation of plant density from high resolution RGB imagery over wheat crops. *Frontiers in Plant Science, under review*

Lohr, S. (2009). *Sampling: design and analysis*. Cengage Learning

Ma, Y., Wen, M., Guo, Y., Li, B., Cournede, P.-H., & de Reffye, P. (2007). Parameter Optimization and Field Validation of the Functional-Structural Model GREENLAB for Maize at Different Population Densities. *Annals of Botany, 101*, 1185-1194

Marshall, M.N. (1996). Sampling for qualitative research. *Family practice, 13*, 522–526

Nakarmi, A.D., & Tang, L. (2012). Automatic inter-plant spacing sensing at early growth stages using a 3D vision sensor. *Computers and Electronics in Agriculture, 82*, 23-31

Nakarmi, A.D., & Tang, L. (2014). Within-row spacing sensing of maize plants using 3D computer vision. *Biosystems Engineering, 125*, 54-64

Norman, D.W. (1995). *The farming systems approach to development and appropriate technology generation*. Food & Agriculture Org.

Olsen, J., Kristensen, L., & Weiner, J. (2006). Influence of sowing density and spatial pattern of spring wheat (Triticum aestivum) on the suppression of different weed species. *Weed Biology and Management, 6*, 165-173

Olsen, J., & Weiner, J. (2007). The influence of Triticum aestivum density, sowing pattern and nitrogen fertilization on leaf area index and its spatial variation. *Basic and Applied Ecology, 8*, 252-257

Papalexiou, S.M., Koutsoyiannis, D., & Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrol. Earth Syst. Sci., 17*, 851-862

Reynolds, M.P., Ortiz-Monasterio, J.I., McNab, A., & CIMMYT (2001). *Application of physiology in wheat breeding*. Mexico: CIMMYT

Rinne, H. (2008). *The Weibull distribution: a handbook*. CRC Press

Särndal, C.-E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media

Scheuren, F., & Association, A.S. (2004). What is a Survey? In: American Statistical Association

Steutel, F.W., & Van Harn, K. (2003). *Infinite divisibility of probability distributions on the real line*. CRC Press

Tang, L., & Tian, L. (2008a). Plant Identification in Mosaicked Crop Row Images for Automatic Emerged Corn Plant Spacing Measurement. *Transactions of the ASABE*, 2181-2191

Tang, L., & Tian, L.F. (2008b). Real-time crop row image reconstruction for automatic emerged corn plant spacing measurement. *Transactions of the ASABE, 51*, 1079

Van der Heijden, G., De Visser, P.H.B., & Heuvelink, E. (2007). Measurements for functional-structural plant models. *Frontis, 22*, 13–25

Whaley, J.M., Sparkes, D.L., Foulkes, M.J., Spink, J.H., Semere, T., & Scott, R.K. (2000). The physiological response of winter wheat to reductions in plant density. *Annals of Applied Biology, 137*, 165-177

Winkelmann, R. (1995). Duration Dependence and Dispersion in Count-Data Models. *Journal of Business & Economic Statistics, 13*, 467-474

Winkelmann, R. (1996). A count data model for gamma waiting times. *Statistical Papers, 37*, 177–187

Zeviani, W.M., Ribeiro Jr, P.J., Bonat, W.H., Shimakura, S.E., & Muniz, J.A. (2014). The Gamma-count distribution in the analysis of experimental underdispersed data. *Journal of Applied Statistics, 41*, 2616-262

3. Estimating wheat Green area index from ground-based LiDAR measurement through 3D ADEL-Wheat model

GAI dynamics along the growth cycle is closely linked with the canopy functioning, thereby serving as an important trait of interest for breeders, crop modelling and decision making for field management. It is relatively easy to achieve accurate GAI estimate using passive observations at early stages. However, the performances degrade for high GAI conditions due to the saturation problem. The use of LiDAR with its capacity to bring information on the third dimension was investigated as a possible way to alleviate the saturation effect based on the regularities between top and deeper canopy layers as described by the ADEL-Wheat model. The LiDAR used is equipping the phenomobile phenotyping platform. Focus was put on the stage of maximum GAI development when saturation effects are the largest. The Sowing pattern described by a statistical model developed in the previous chapter was exploited here. This chapter corresponds to a draft journal article to be submitted to Agricultural and Forest Meteorology[5].

[5] Liu, S., Baret, F., Boudon, F., Thomas, S., Zhao, K., Fournier, C., Andrieu, B., Kamran, I., & de Solan, B. (2017). Estimating wheat Green area index from ground-based LiDAR measurement through 3D ADEL-Wheat model. To be submitted to Agricultural and Forest Meteorology

3.1. **Estimating wheat Green area index from ground-based LiDAR measurement through 3D ADEL-Wheat model**

Shouyang Liu[1*], Fred Baret[1], Fred Boudon[2], Samuel Thomas[7], Kaiguang Zhao[3], Christian Fournier[4], Bruno Andrieu[5], Irfan Kamran[1], Matthieu Hemmerlé[6] and Benoit de Solan[7]

1 INRA EMMAH, UMR 1114 Domaine Saint-Paul, Site Agroparc, 84914 Avignon Cedex 9, France

2 CIRAD, EPI INRIA Virtual Plants & UMR AGAP, Campus St Priest, 34095 MONTPELLIER Cedex 5, France

3 School of Environment and Natural Resources, Ohio Agricultural and Research Development Center, The Ohio State University, Wooster, OH 44691, USA

4 INRA-SupAgro, UMR 759, Place Viala, 34060 Montpellier Cedex 1, France

5 INRA, AgroParisTech, UMR ECOSYS, 78850 Thiverval-Grignon, France

6 Hi-Phen, 84914, Avignon, France

7 ARVALIS-Institut du végétal, Station expérimentale, 91720 Boigneville, France

Abstract

Green area index (GAI) has been difficult to estimate accurately at large scales due to the cost prohibitive nature of classical in-situ methods. Using the passive remote sensing technique of light detection and ranging (LiDAR), the difficulty of mapping GAI on ecosystems could be overcome. However, it has not yet been addressed on agricultural systems, especially for conditions with large GAI values. Through this work, we proposed a self-learning method to estimate GAI using LiDAR-derived metrics over a wheat field. Specifically, we developed a LiDAR simulator to carry out scanning on digital 3D objects, mimicking the measuring principle and setups of actual LiDAR sensors. Further, coupling the LiDAR simulator with the 3D ADEL-Wheat model made it possible to produce *in silico* scan experiments over varied canopies. We then propose to use a machine learning algorithm to correlate LiDAR-derived metrics to GAI over synthetic datasets. Eventually, the model performance was evaluated with both independent synthetic datasets and in-situ destructive measurements. We achieved a mean RMSE of 0.41 and mean rRMSE of 0.11. The model is further compared with the classically used Beer-Lambert approach. Our method is preferable for its better accuracy and capability of generalization. This study illustrated how LiDAR data could provide interesting interpretation for agricultural system.

Keywords

GAI, ADEL-Wheat, wheat, LiDAR, Neural Networks

1 Introduction

Green area index (GAI) is defined as the total one-sided area of green vegetation elements per unit ground horizontal surface. GAI is here preferred to leaf area index since it includes other green organs such as stems and ears (Baret et al. 2010a; Verger et al. 2014a) that significantly contribute to the canopy photosynthesis, respiration and transpiration (Bonan 1993; Weiss et al. 2004). Further, GAI excludes the senescent leaf parts that are no more functioning. GAI is commonly estimated in-situ using green fraction measurements (the fraction of green elements seen from a given direction) derived from downward looking hemispherical digital photography (Weiss et al. 2004) or imagery acquired with longer focal length (Baret et al. 2010a). The canopy is generally assumed to be a turbid medium to ease the derivation of the GAI based on a well-established theory (Nilson 1971). However, these GAI estimation methods are limited due to their prohibitive cost (Zheng and Moskal 2009) when repeated a large number of times , and the associated accuracy may be compromised by the possible violation of the turbid medium assumption. Further, in the case of field phenotyping applications where the GAI dynamics along the growth cycle is highly desired (Araus and Cairns 2014a), structural differences between genotypes that are not well represented by the turbid medium assumption may degrade the accuracy of the GAI estimates.

Remote sensing techniques including optical passive sensors and active light detection and ranging (LiDAR) provide efficient ways to estimate GAI over larger spatial domains. Passive optical methods are generally based on the calculation of vegetation indices such as the Normalized Vegetation Difference Index (NDVI) (Baret and Guyot 1991) or the green fraction when high enough spatial resolution is available (Baret et al. 2010a). These methods are generally showing a degradation of their performances for the larger GAI values because of the saturation of the signal; Further, these techniques may be sensitive to the variation in the illumination conditions that affect vegetation indices (Stark et al. 2016) as well as the separation between green elements and the background (Baret et al. 2010a). Conversely to these passive remote sensing techniques, LiDAR is an active method independent from the illumination conditions. Further, LiDAR offers a more detailed description of the canopy architecture by providing the third dimension from which more canopy traits of interest could be potentially retrieved. Recent work has demonstrated its capability for GAI estimates over a wide range of values on both homogenous and heterogeneous forest stands (Béland et al. 2014; Griebel et al. 2015; Ilangakoon et al. 2015; Korhonen et al. 2011; Richardson et al. 2009; Zhao and Popescu 2009) and orchards (Arnó et al. 2013). However, little work has been dedicated to the application of airborne or ground-based LiDAR systems for GAI estimates for staple crops.

The relatively little use of LiDAR over crops was mainly due the significant footprint of the available systems that were limiting the description of the fine architecture. This limitation is now partly overcome due to the reduction of footprint size and the increase of scanning frequency of some of the current LiDARs (Eitel et al. 2014; Lin 2015; Saeys et al. 2009). Further, the interaction of the laser beam with the canopy structure is complex (Kukko and Hyyppä 2009) and specific interpretation systems should be developed to reach the expected level of accuracy and precision required (Baret and Buis 2008; Zhao et al. 2009; Zhao and Popescu 2009).

Empirical based methods have been derived from available data sets acquired under specific conditions and instruments (Richardson et al. 2009). Although these methods are tractable and may perform well for the cases similar to those prevailing during the calibration experiments, they will provide uncertain performances outside the calibration conditions. Alternatively, laser–canopy interactions could be described using physically based models. The turbid medium assumption has been exploited recently by (Zhao et al. 2015) to estimate GAI from the gap fraction measured by LiDAR. Although this approach does not require any calibration process, the LiDAR-derived gap fraction, hence the estimated GAI, may be greatly affected by LiDAR intrinsic setup and extrinsic scanning pattern (Morsdorf et al. 2006; Van der Zande et al. 2006; Zheng and Moskal 2009). Moreover, GAI estimates using the turbid medium assumption provide an effective GAI. Possible clumping effect has to be corrected to get the actual GAI (Chen et al. 1997).

The limitations of the two approaches to estimate GAI from LiDAR measurements may be overcome using machine learning techniques trained over simulations of the LiDAR signal achieved with a realistic 3D description of the canopy structure. The turbid medium assumption will therefore be no more necessary

allowing accessing the true GAI, while model simulations allow considering the whole range of possible cases conversely to what is currently achievable using actual experimental observations. The objective of this study is to propose such a synthetic-learning approach to estimate the true GAI from ground-based LiDAR measurements as sketched in **Fig. 1** . The synthetic learning data set is generated using ADEL-Wheat 4D model (Abichou et al. 2013a; Fournier et al. 2003) coupled to a LiDAR simulator mimicking the actual LiDAR measurements. Machine learning techniques were then trained over the synthetic data set to estimate the true GAI corresponding to the LiDAR measurements. For this purpose, the LiDAR observations were described by few features. Further the space of canopy realization may be reduced to speed up the computation using the available prior information on the range of possible cases. Finally, the performances of the method were evaluated over both an independent synthetic dataset as well as over actual field observations to evaluate its performances over wheat crops cultivated under field conditions within phenotyping experiments.

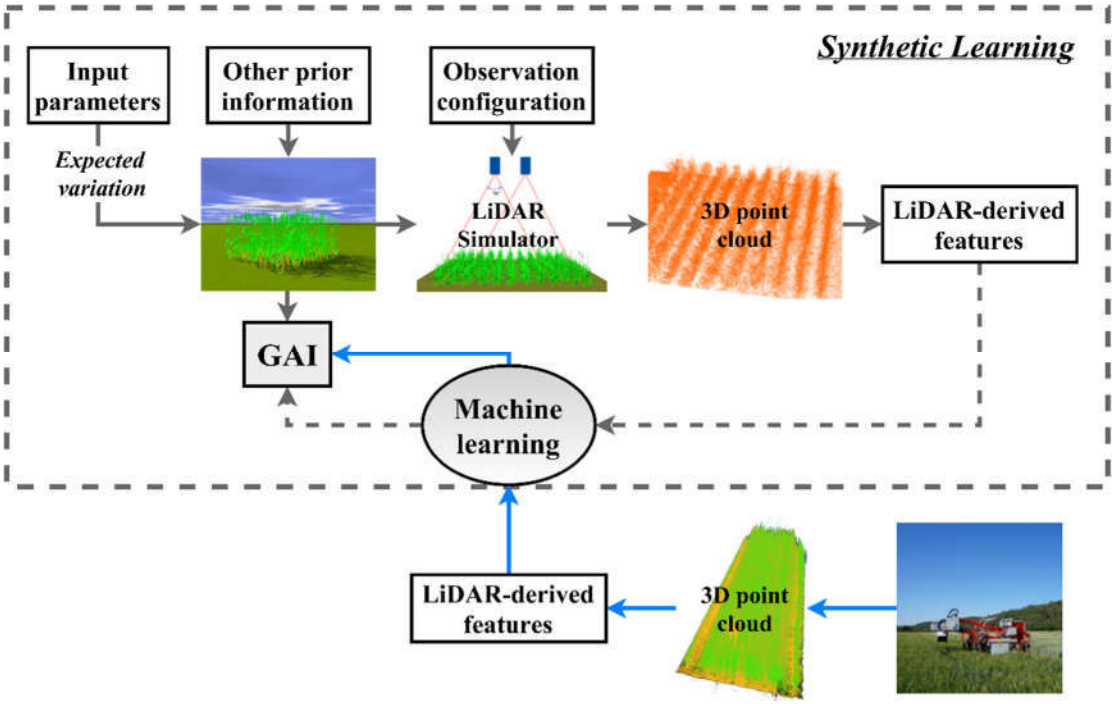

Fig. 1 Diagram showing the principles of the synthetic-learning methodology developed to estimate GAI from ground-based LiDAR measurements

2 Materials and methods

2.1 The field experiment

2.1.1 Experimental conditions and GAI measurement for validation

The experiment was conducted in the wheat field in Gréoux-les-Bains (N: 43°45′33″; E: 5°53′03″), France. It was sown October the 29th 2015 with row spacing 17.5 cm at density 300 seeds·m^{-2}. The field measurements were carried out on April the 14th 2016 around thermal time 1200 °C · d over 14 different cultivars. It corresponds the flag leaf ligulation stage when GAI reaches its maximum over the growth cycle (Abichou et al. 2013a). On that day, field measurements were first conducted using a fully automated rover robot called “phenombile” equipped with two LiDARs (**Table 1**). Then GAI was measured destructively for the validation of the GAI estimation from the methods proposed in this study. Plants were collected over one 0.5 m × 0.5 m sampling area. The green leaves were separated from the stems and the senesced leaves were discarded since they were represented a negligible contribution of the leaf area at the time of the experiment. The area of a sample of 30 leaves was estimated by multiplying the measured length and width with a shape factor of 0.65. The sample of 30 leaves was then weighed. The total leaf area was computed by multiplying the ratio of leaf area to its mass measured over the 30 leaves by the total weight of the leaves collected on the 0.5 m × 0.5 m sampling area. Similarly, the area of stems and ears was estimated over a subsample of 10 stems by multiplying half the average perimeter by the length of the stem according to (Lang 1991) and

(Stenberg 2006). The 10 stems were weighed. Note that at this stage the stems and ears are green. The total green area of stems and ears was then computed by multiplying the ratio of stem and ear area to its mass measured on the 10 stems, by the total weight of the stems (and ears) collected on the sampling area. Finally, the total area of leaves and stems and ears were summed up and divided by the 0.5 m × 0.5 m sampling area.

2.1.2 The artificial canopy

Describing the 3D structure of an actual wheat canopy is a very tedious task almost impossible to complete within a reasonable time without compromising the accuracy and realism of the description. It was therefore preferred to validate the LiDAR simulator over an artificial canopy built specifically. The artificial canopy (**Fig. 2a**) represented an area of 1.2 m x 1.2 m where 45 pseudo-plants were evenly distributed along 6 rows spaced by 15 cm with a plant spacing of 15 cm. Plants are made of a cylindrical wooden stem of 1cm radius with height ranging between 0.105 m to 0.92 m. One or two rectangular leaves made of thick white paper board with dimension 15 cm x 2.5 cm, were randomly attached in height and orientation to each stem. The orientation and position of each leaf was accurately measured. This allowed to build a faithful computer representation of the artificial canopy using the Python library, PlantGL (**Fig. 2b**) (Pradal et al. 2009). This representation of the artificial canopy is referred as the "virtual-artificial canopy".

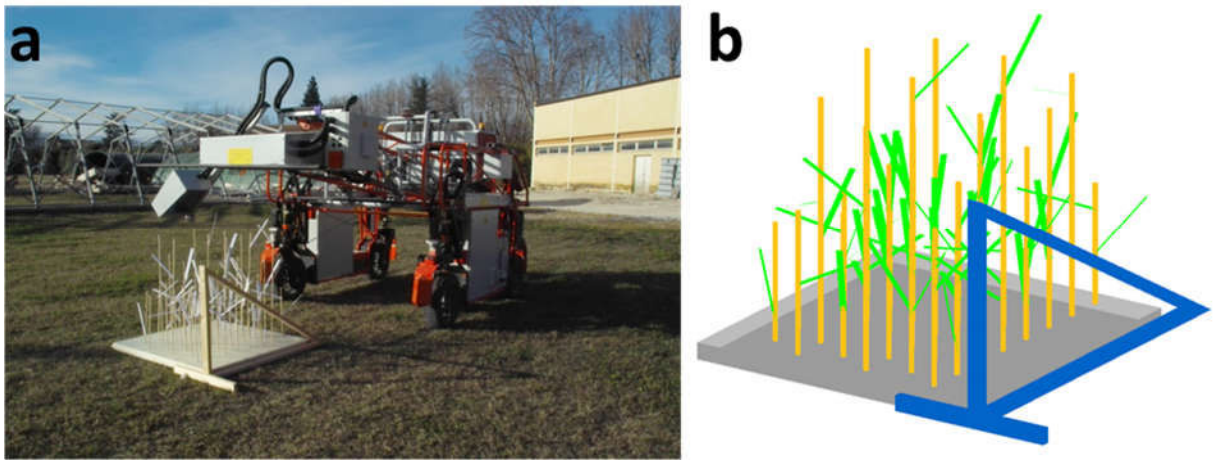

Fig. 2 a) Phenomobile measurement on the artificial canopy; b) The virtual artificial canopy simulated with PlantGL. The triangle on the side of the artificial canopy was used to check the LiDAR calibration in terms of distance to the sensor.

2.2 LiDAR measurements and processing

2.2.1 LiDAR measurements

The measurement head of the phenomobile (**Fig. 2a**) was maintained automatically at a constant distance to the top of the canopy. The system steps over the microplots with a maximum 1.35 m clearance and an adjustable width of 2.0±0.5 m. The phenomobile automatically follows a predefined trajectory in the experimental fields using a centimetric accuracy RTK-GPS and accelerometers. The measurement head is equipped with several instruments including two LMS400 LiDARs (SICK, Germany) operating at 650 nm and scanning downward with $\pm 35°$ zenith angle in a direction perpendicular to the rows at a frequency of 290 scans per second (Lefsky et al. 2002). The two LiDARs provide denser sampling of the scene. As the platform moves forward (**Fig.3**) at a speed of 0.3 $m \cdot s^{-1}$ as recorded with the GPS information, the distance between two consecutive scans of a LiDAR along the row direction is around 1 mm. Measurements are taken every 0.2° along the scanning direction. The footprint of the LiDAR is defined by the instantaneous field of view. Therefore, the size of the footprint will slightly depend on the distance to the sensor that varies from 2.4 mm x 5 mm at 0.7 m minimum measuring distance up to 7 mm x 5 mm at 2 m maximum measuring distance. The LiDAR operates at 650 nm and the distance between the sensor and the target is measured from the phase shift principle (Neckar and Adamek 2011). The intensity of the reflected signal is recorded at the same time as the distance. When the target is not flat and horizontal, the distance and the intensity computed by the LiDAR is approximately the average distance value over the LiDAR footprint. The nominal measuring error on

the distance is 4 mm in our experimental conditions. **Table 1** lists the several variables defining the measurement configuration.

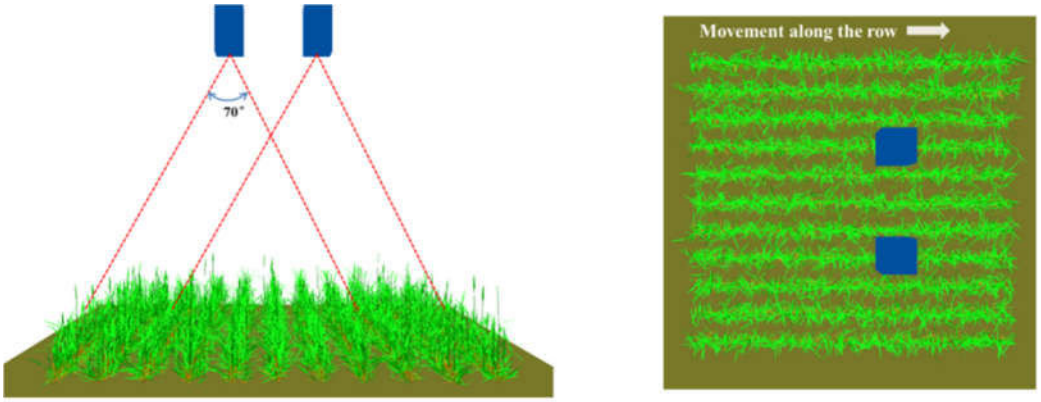

Fig.3 Scheme showing the measurement configuration. The two blue boxes correspond to two LiDARs looking downward.

Table 1 LiDAR observational configuration used in Gréoux

System	Variables	Value	Unit	Comment
Phénomobile	Speed	30	$cm \cdot s^{-1}$	Speed of the phenomobile
	Start position	Lidar1: [-22.2,-100]	cm	Defines the LiDAR trajectory from start to end
		Lidar2: [23.4,-100]	cm	
	Stop position	Lidar1: [-22.2,100]	cm	
		Lidar2: [23.4,100]	cm	
	Height	200	cm	Height above the top of the canopy
	Azimuth	0	°	perpendicular to the row direction
	Zenith	0	°	Nadir looking
LiDAR	Angular resolution	0.2	°	$\text{angle}_{\text{res}}$
	Footprint width	5	mm	$\text{Width}_{\text{foot}}$
	Scanning frequency	290	Hz	scans/second
	Scanning range	± 35	°	Range

2.2.2 Data processing

Each LiDAR scan over a microplot is associated to a position and orientation thanks to the GPS and accelerometers. This allows producing a 3D cloud of points of the whole microplot as illustrated by **Fig. 4** . Then, each point of the 3D point cloud was classified as either green, senescent or background using the intensity information as proposed by (Coren and Sterzai 2006). The green fraction may thus be computed afterwards (Meng et al. 2010). The use of the intensity was proved to be here more efficient as compared to the use of the height (results not shown for the sake of brevity). A radiometric correction was first applied to the measured intensity to account for the decrease of the incoming intensity as a function of the distance to the sensor (Baltsavias 1999; Donoghue et al. 2007; Jelalian 1992). Finally, the Otsu thresholding method with two thresholds (Otsu 1975) was employed on the corrected intensity points to distinguish green vegetation, senescent vegetation and the background surface (**Fig. 5**).

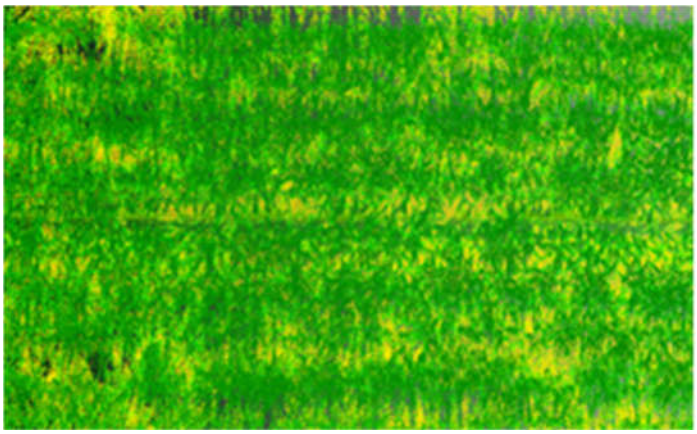

Fig. 4 Visualization of 3D point cloud, classified into vegetation (green) and soil (yellow)

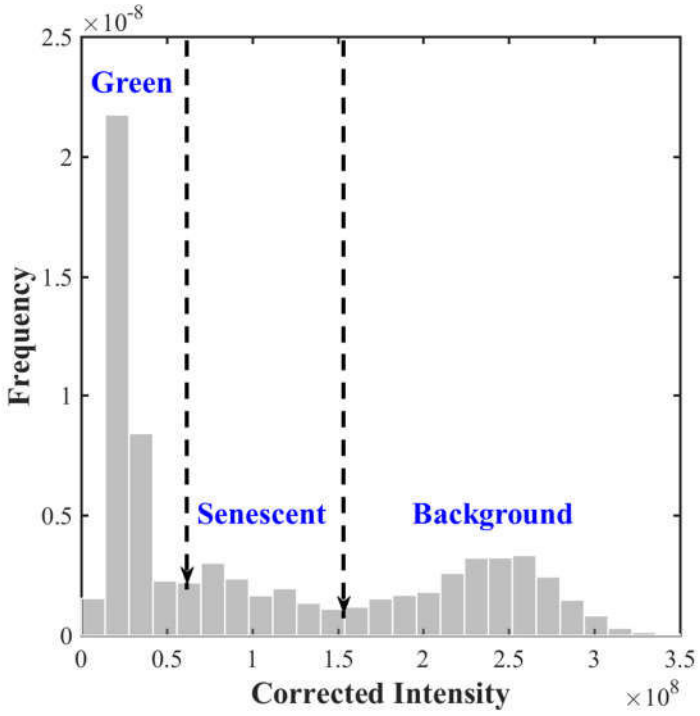

Fig. 5 Histogram of the corrected intensity derived from the LiDAR measurements. The two arrows correspond to the thresholds values used to separate the green from senescent and background points.

2.3 LiDAR simulations from ADEL-Wheat model 3D canopy description

2.3.1 ADEL-Wheat simulations

Realistic wheat canopies were generated with ADEL-Wheat and GenPlant models in OpenAlea, an open source plant architectural modelling platform (Pradal et al. 2008). The ADEL-Wheat 4D model simulates the development of wheat crops from emergence to maturity. It describes both the development of organs with time including the position, size, shape and orientation based on regularities observed over a range of contrasted experimental conditions. More details on the ADEL-Wheat 4D model can be found in (Abichou et al. 2013a; Fournier et al. 2003). The implementation of the model was taken into account the spatial distribution of plant positions according to the distribution laws proposed by (Liu et al. 2016a).

One thousand canopies were simulated using Latin Hypercube experimental sampling over the 12 parameters listed in **Table 2**. Parameters were assumed to be evenly distributed between their maximum and the minimum values. The bounds of each parameter were derived from the previous set of experiments (Abichou et al. 2013a; Fournier et al. 2003). Simulations were completed for 1200 ℃. d that roughly corresponds to the beginning of earing. The scenes correspond to 11 rows of 2 m length with 17.5 cm inter-row spacing. The number of rows and row spacing are consistent with actual experimental observations. The

length of the scene corresponds to a trade-off between the precision of the simulations and the required computing cost. The output of each simulation includes the 3D canopy scene and the corresponding GAI information (**Fig. 6**).

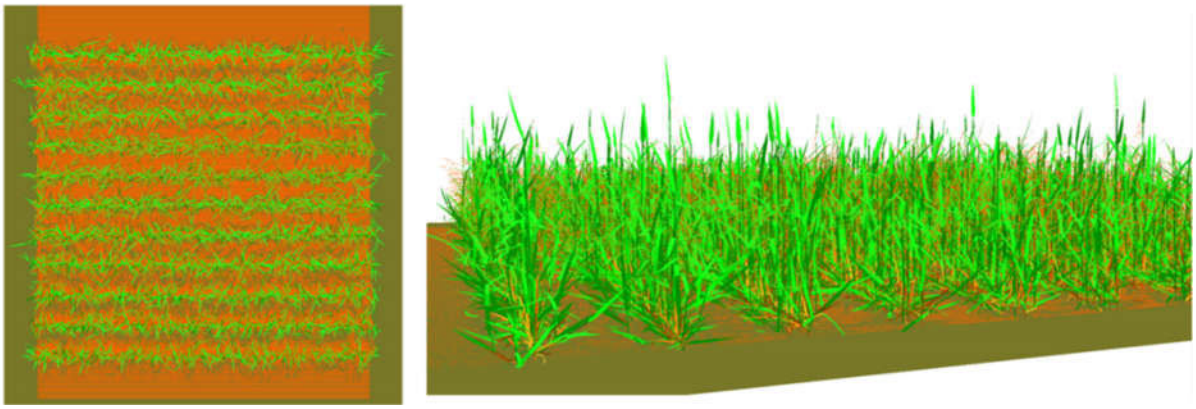

Fig. 6 3D wheat canopy generated by ADEL-Wheat model. On the left, top view. On the right, perspective view from the ground level.

Table 2 Parameters of ADEL-Wheat to generate the training dataset

Abbr.	Descriptions	Min	Max	Unit
TT_{phy}	Phyllochron: thermal time interval between the emergence of successive leaf	90	110	℃ · d
Ang_{shift}	The rotation of leaf basal inclination relative to the measurement	-10	20	degree
N_{leaf_main}	Number of leaves on the mainstem	9	12	scalar
L_{lamina_1}	The length of lamina one	3	8	cm
α_{lamina_N1}	The increase rate of lamina length during phase 1	0	3.5	scalar
α_{lamina_N2}	The increase rate of lamina length during phase 2	2.5	4.5	scalar
N2	The number of leaves corresponding to the turning point from α_{lamina_N1} to α_{lamina_N2}	4.5	6	scalar
N_{tiller}	Number of tillers per plant	0	5	scalar
Ang_{tiller}	The inclination of the base of the tiller relative to main stem (final angle at flowering)	10	85	degree
S_{start}	Number of green leaves at the start of leaf senescence on the mainstem	3.5	6.5	scalar
S_{min}	Number of green leaves at Haun stage $N_{leaf\_main} - 4$ on the main stem	1.5	3.5	scalar
S_{flag}	Number of green leaves at Haun stage N_{leaf_main} on the main stem	3.5	6.5	scalar

2.3.2 LiDAR simulator

The Z-buffer (Catmull 1974) is the most common method to get the 3D point cloud of the canopy parts that are visible from the camera. The method searches the minimum view depth for each pixel and stores the corresponding coordinates and intensity in a buffer. The result is thus analogous to what measures the LMS400 LiDAR. The Z-buffer provides 3D point cloud that may be with much finer resolution that that of the LiDAR. It should thus be resampled according to the LiDAR scanning range and footprint size. To model faithfully the LiDAR measurement, the camera used to record the z-buffer image is positioned on the virtual scene where it is actually located on the measurement head (**Fig. 3**). Its view angle corresponds to the angular scanning range of the LiDAR. The depths of each pixel of the z-buffer are converted to points in polar coordinates using the camera parameters. These points are then aggregated according to the angular

resolution of the LiDAR. The centroid of all the pixels corresponding to a LiDAR footprint is then computed. Once a scan is simulated, the virtual cameras are shifted along the row direction by a distance determined from the actual platform speed and scanning frequency to simulate the next scan. The schematic diagram of the simulation algorithm is sketched in **Fig. 7** with the input variables listed in **Table 2**. The simulation algorithm was coded with the PlantGL python library (Pradal et al. 2009).

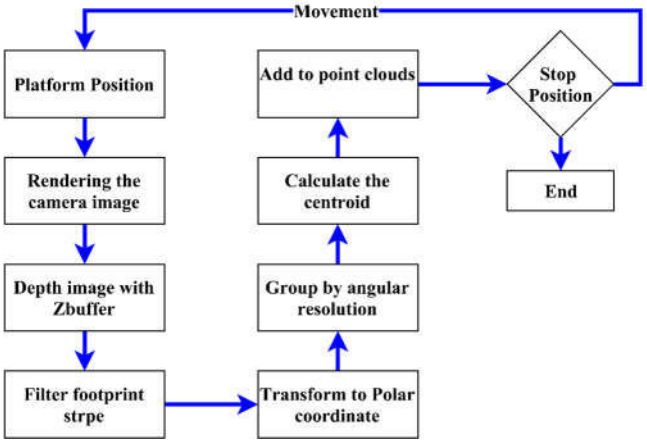

Fig. 7 Schematic diagram of the LiDAR simulator

2.4 Machine learning algorithm

The 3D cloud of points derived from the LiDAR needs first to be described with a set of features. They will then be used as the input variables for the machine learning algorithm that will be trained to estimate GAI.

2.4.1 LiDAR-derived features

Two types of features were extracted from the LiDAR derived 3D point cloud, mostly focusing on the vertical distribution of the 3D point clouds: (1) the height corresponding to six quantiles of the green points, H^{q}_{green}, with $q = [0.15,\ 0.30,\ 0.45,\ 0.60,\ 0.75, 0.90]$ for all the zenith directions available and (2) the directional green fractions, $f_g\ (\Theta) = \mathrm{P}_{\mathrm{green}}(\Theta)/(\mathrm{P}_{\mathrm{green}}(\Theta) + \mathrm{P}_{\mathrm{soil}}(\Theta) + \mathrm{P}_{sen}(\Theta))$ where $\mathrm{P}_{\mathrm{green}}(\Theta)$, $\mathrm{P}_{\mathrm{soil}}(\Theta)$ and $\mathrm{P}_{\mathrm{sen}}(\Theta)$ are respectively the total number of green, soil or senescent points of the 3D point cloud for a range of zenith angles $\Theta = [\theta < |\theta| < \theta + 5°]$ where θ is the zenith angle corresponding to each point of the 3D point cloud. Four ranges of zenith angles are considered: $\Theta = [0°, 5°, 10°, 15°]$. The range of angles was limited to 20° since for $\Theta > 20°$ the LiDAR beams are getting outside the wheat plot sampled (**Fig. 8**). As a result, each 3D point cloud will be characterized by a vector of 10 values, including 6 heights and 4 green fractions.

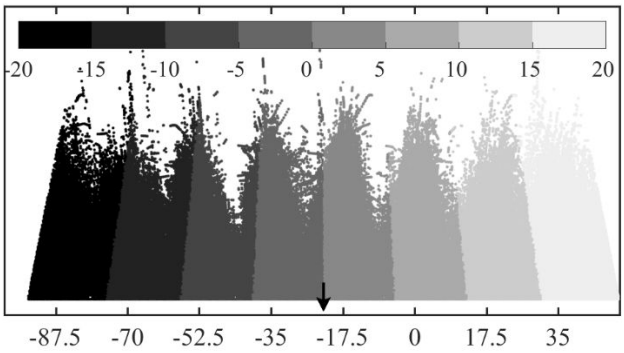

Fig. 8 Groups of incident angles illustrated by LiDAR2. The X axis represents the horizontal distance to the center of the plot.

2.4.2 Artificial neural networks

Machine learning algorithms, and among them Artificial Neural Networks (ANNs), have proven their efficiency to estimate biophysical characteristics from remote sensing observations (Baret and Buis 2008)

including LiDAR-derived features (Zhao et al. 2011). ANNs combine the input with interconnected neurons to best match the output values over a training database. In our case, inputs correspond to the features of the 3D point cloud, and the output is the GAI. The training process requires first to define the network architecture (the number of hidden layers and nodes per layer and the type of transfer function of each neuron). Then the synaptic weights and biases are tuned to get a good performance. In our case, a one-layer feed-forward network was used. The first layer has three neurons with tangent sigmoid transfer function. The output layer has one neuron with a linear transfer function. The training was achieved with 85% of the whole data set. The remaining 15% was used for the test data set to evaluate the performances.

2.5 Estimates of GAI based on the turbid medium assumption

The turbid medium assumption allows computing the green fraction (i.e. the complementarity to the unity of the gap fraction) as a function of the GAI and the leaf angle function distribution (LIDF):

$$f_g\ (\Theta) = 1 - e^{-\frac{G(\Theta,LIDF(ALIA))}{\cos(\Theta)}GAI} \qquad [1]$$

The LIDF was assumed to follow an ellipsoidal distribution (Campbell 1986) characterized by the average leaf inclination angle (ALIA). Both GAI and ALIA were estimated from the directional variation of the green fraction with $\Theta = [0°, 5°, 10°, 15°]$. GAI and ALIA were retrieved from the green fractions using the interior-point iterative algorithm (Byrd et al. 2000) used to minimize the Euclidian distance between the estimated green fraction and the measured one.

3 Results

3.1 Validation of LiDAR simulator on the artificial canopy

The LiDAR simulator was run on the virtual artificial canopy over configurations consistent with the actual LiDAR observations. The cumulated density functions (CDF) for the 3 axes [X, Y, Z] were computed from both the measured and simulated 3D point clouds. These metrics were selected since they represent the statistical properties of the scene. As a matter of facts, the two 3D point clouds could not be compared at the point level for obvious reasons on the accuracy of the positioning. Further, the CDF along the Z axis provides the basic information for the extraction of the 6 first features used in the following to estimate GAI. Results show a high degree of consistency between the measured and simulated CDFs as illustrated in **Fig. 9**. Two-sample Kolmogorov-Smirnov tests were further conducted to quantify the agreement: the null hypothesis that the samples are drawn from the same distribution is accepted with a probability of 99%. In addition, the comparison between the 4 directional green fractions shows also a very good agreement. Similar results not shown for the sake of brevity were found for other measurement configurations where the scanning frequency, phenomobile speed and measurement heights were changed. This comparison demonstrates that the simulator provides 3D point clouds very consistent with the actual measurements. Note that the simulator describes as well the 'shadow points' that are observed at the border of the leaves because of the dimension of the LiDAR beam footprint.

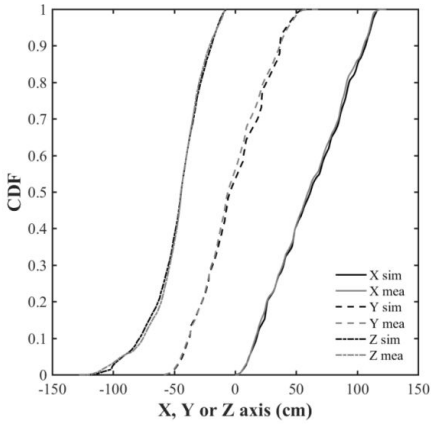

Fig. 9 The cumulative distribution function (CDF) of measured (mea) and simulated (sim) points along X, Y and Z axis.

3.2 Performances of GAI estimates using Poisson model

The performance of GAI and average leaf angle estimates using Poisson model was evaluated using both synthetic dataset and actual measurement. Results achieved over the synthetic data set shows that the retrieval of GAI from green fractions observed around the nadir direction is relatively inaccurate with RMSE=1.0 (**Fig. 10**). As a matter of facts, the restricted range of observation angles provides only little information of the canopy structure. Further, the turbid medium assumption is strongly violated for directions close to nadir for wheat crops for which plants are grouped into rows where leaves are clumped. The retrieved values of ALIA are always close to 80°, corresponding to an erectophile leaf inclination distribution, quite different from the actual distribution that is closer to a spherical distribution. Results obtained using the actual LiDAR measurements with the phenomobile confirms the difficulty to estimate GAI accurately from green fractions observed from near nadir directions (**Fig. 10**).

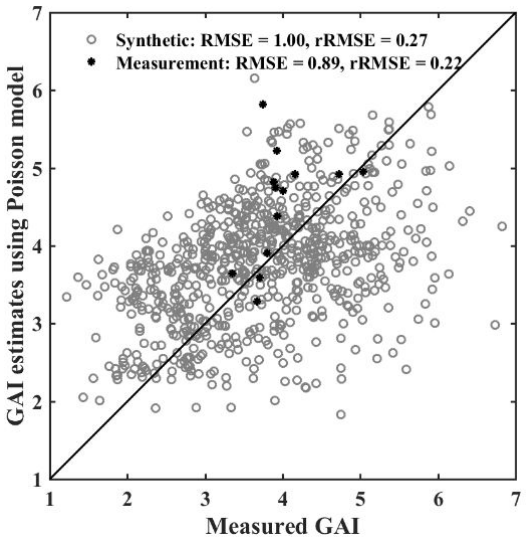

Fig. 10 Performance of GAI estimates using the turbid medium assumption. The empty symbols correspond to the synthetic test data set. The filled black symbols correspond to the actual measurements.

3.3 Performances of GAI estimates using synthetic learning approach

The performance of GAI estimates using ANN was evaluated for different sets of the 3D point cloud features over both the synthetic test dataset (the 15% of the simulated scenes) and over the destructive measurements. When using the directional green fraction (**Fig.11a**), performances evaluated over the synthetic test data set appear improved very significantly as compared to the estimates based on the turbid medium assumption (**Fig.10**). However, the performances evaluated over the measurement data set show a large overestimation of GAI values. This may be due to unrealistic representation of the leaf curvature in the ADEL-Wheat model that may impact very significantly the green fraction, particularly for directions close to nadir as in the cases considered. The relatively simple leaf curvature profiles and the random leaf azimuth need probably to be revisited to improve the realism of the canopy structure representation in ADEL-wheat. When using the 6 quantiles describing the height distribution of the green points, estimation performances evaluated over the test synthetic dataset improved significantly as compared to the previous situation when using only the green fractions (**Fig.11b**). However, more scattering is observed for the higher GAI values. Application of the neural network using the actual LiDAR height quantiles measured from the phenomobile provides GAI estimates very consistent with the destructive GAI measurements, with RMSE values of the same order of magnitude as those associated to the synthetic test data set (**b**). Adding the 4 directional green fractions to the 6 height quantiles to the inputs of the neural network improves significantly the GAI estimates as evaluated over the test synthetic data set (**Fig. 11c**). The larger discrepancies observed for the high GAI values when using only heights (**Fig.11a**) and the overestimation observed when using the green fraction as compared to the measurements (**Fig. 11b**) are significantly reduced. However, when applying the neural network to the actual LiDAR measurements increases only marginally the GAI estimation accuracy.

The lack of realism in the description of the green fraction explains why the GAI retrieval performances are not improved.

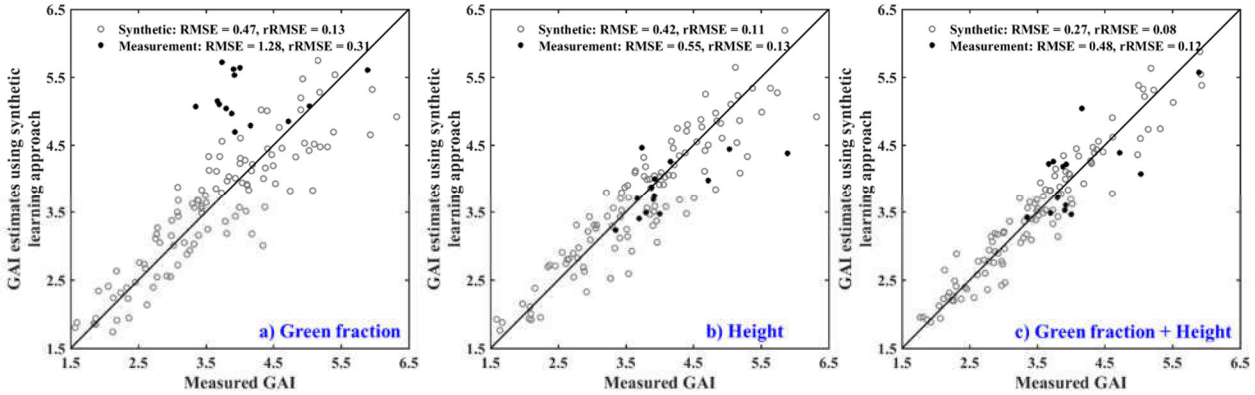

Fig. 11 Performances of LiDAR-based GAI estimates using synthetic learning approach. a) the 4 directional green fractions are used as inputs to the ANN. b) The 6 height quantiles are used as inputs. c) both the 4 directional green fractions and the 6 height quantiles are used as inputs.

4 Discussions and Conclusion

A synthetic-learning methodology was developed in this study to estimate GAI from ground-based LiDAR observations over wheat. A neural network machine learning method was employed to derive GAI from features characterizing the LiDAR-3D point cloud. The neural network was trained over a synthetic dataset generated by a 3D crop model, ADEL-Wheat, coupled with a LiDAR simulator. This approach allows circumventing the limits of training the neural network over actual experimental measurements that are prone to uncertainties and are limited by the number and range of cases that can be sampled. Another important advantage of the method proposed is that it takes explicitly into account the sensor specifications to simulate the measurements as it is actually acquired. In our case, the LiDAR footprint size is on the same order of magnitude as leaf and stem widths, creating 'shadow points' at the border of the leaves. A direct comparison with the actual distribution of height or green fraction computed from the 3D synthetic canopy scene would have therefore revealed biases as compared to the LiDAR measurements. Since these effects have been taken into account in the LiDAR simulator, the simulated 3D cloud of points is fully consistent with the LiDAR measurements: LiDAR simulator plays a similar role as that of the observation operator in data assimilation techniques.

The method presented here was calibrated over a specific development stage. However, the continuity observed in growth and development of the canopy structure indicates that the calibration would be valid within a limited temporal window around the nominal stage considered. Further, the development stage in wheat is well described by the thermal time since sowing, which will ensure to apply the method at a date very close to the nominal one. However, the success of the method will highly bear on the realism and plasticity of the 3D architecture model used. In ADEL-Wheat model, the plant to plant variability is simulated stochastically without representing the correlations that likely exist between adjacent plants. This may exaggerate leaf clumping and thereby bias the corresponding green fraction with consequences on GAI estimates as reported in this study. Progress is therefore needed to improve this aspect in canopy 3D architecture models. Finally, the training database used to relate the LiDAR 3D point cloud features to GAI should represent well the possible variability of canopy architecture, including the possible genotypic and environmental dependency. The range of values used for each parameter to generate the training database is therefore of prime importance. They were here derived from a collection of available experiments and the good results obtained in GAI estimation demonstrate that they were pertinent. Nevertheless, the results of the available experiments should be organized to facilitate their compilation for the generation of the prior distributions and co-distributions of the derived model parameters.

For the low to medium GAI values, the methods based on the green fraction are associated to an accuracy close to 12% for wheat crops (Baret et al. 2010a). The use of the LiDAR for these early stages (100, 300, 500

and 900 ℃ · d) was providing an accuracy very similar using the method developed here (results not reported for the sake of brevity). However, at the flag leaf ligulation stage(1200 ℃ · d) the GAI reaches maximum values. In such conditions, the use of techniques based on the reflected or transmitted radiation measured is known to be less accurate for the larger GAI values due to saturation problems. The LiDAR techniques combined with 3D realistic models of canopy architecture present the advantage to use the distribution in height of the non-occulted parts to infer the corresponding state of the canopy, thus its GAI value. This allows preventing from the saturation effect at the condition, apparently verified here, that the non-occulted parts are consistently related to the occulted parts.

As a matter of facts, this study concentrates on the GAI variable that is highly desired by plant breeders. In the ADEL-Wheat model, GAI is not a parameter but results from the canopy architecture driven by the actual model parameters. This synthetic learning technique has thus the potential to retrieve other model parameters as well as their combinations into integrated variables such as the radiation interception efficiency. Further, simple features extracted from the 3D point cloud and used as input into the machine learning process correspond only to a limited fraction of the information content provided by LiDAR observations. Investigations should focus on the use of efficient methods now available such as auto-encoders (Pascal et al. 2010) to derive the latent variables of the 3D point cloud to access more model parameters with a better accuracy.

This study concentrated on wheat crops at a given stage using LiDAR measurements only. The proposed approach offers the potential to extend this demonstration case to the combined use of several sensors including RGB cameras and multispectral cameras, to other development stages and to other species providing a realistic 3D canopy architecture is available.

Acknowledgments

This study was supported by “Programme d’investissement d’Avenir” PHENOME (ANR-11-INBS-012) and Breedwheat (ANR-10-BTR-03) with participation of France Agrimer and “Fonds de Soutien à l’Obtention Végétale”. The grant of the principal author was funded by the Chinese Scholarship Council. The work was completed within the UMT-CAPTE funded by the French ministry of Agriculture.

References

Abichou, M., Fournier, C., Dornbusch, T., Chambon, C., Baccar, R., Bertheloot, J., Vidal, T., Robert, C., David, G., & Andrieu, B. (2013). Re-parametrisation of Adel-wheat allows reducing the experimental effort to simulate the 3D development of winter wheat. In, *7th International Conference on Functional-Structural Plant Models* (pp. 304-306)

Araus, J.L., & Cairns, J.E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *TRENDS in Plant Science, 19*, 52-61

Arnó, J., Vallès, J.M., Llorens, J., Sanz, R., Masip, J., Palacín, J., & Rosell-Polo, J.R. (2013). Leaf area index estimation in vineyards using a ground-based LiDAR scanner. *Precision Agriculture, 14*, 290-306

Béland, M., Baldocchi, D.D., Widlowski, J.-L., Fournier, R.A., & Verstraete, M.M. (2014). On seeing the wood from the leaves and the role of voxel size in determining leaf area distribution of forests with terrestrial LiDAR. *Agricultural and Forest Meteorology, 184*, 82-97

Baltsavias, E.P. (1999). Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing, 54*, 199-214

Baret, F., & Buis, S. (2008). Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems. *Advances in land remote Sensing* (pp. 173–201): Springer

Baret, F., de Solan, B., Lopez-Lozano, R., Ma, K., & Weiss, M. (2010). GAI estimates of row crops from downward looking digital photos taken perpendicular to rows at 57.5° zenith angle: Theoretical considerations based on 3D architecture models and application to wheat crops. *Agricultural and forest meteorology, 150*, 1393-1401

Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment, 35*, 161-173

Bonan, G.B. (1993). Importance of leaf area index and forest type when estimating photosynthesis in boreal forests. *Remote Sensing of Environment, 43*, 303-314

Byrd, R.H., Gilbert, J.C., & Nocedal, J. (2000). A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming. *Mathematical Programming, 89*, 149-185

Campbell, G.S. (1986). Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution. *Agric. For. Meteorol., 36*, 317-321

Catmull, E. (1974). A subdivision algorithm for computer display of curved surfaces. In: University of Utah

Chen, J.M., Rich, P., Gower, S.T., Norman, J.M., & Plummer, S. (1997). Leaf area index of boreal forests: theory, techniques and measurements. *Journal of Geophysical Research, 102*, 29429-29443

Coren, F., & Sterzai, P. (2006). Radiometric correction in laser scanning. *International Journal of Remote Sensing, 27*, 3097-3104

Donoghue, D.N., Watt, P.J., Cox, N.J., & Wilson, J. (2007). Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. *Remote Sensing of Environment, 110*, 509-522

Eitel, J.U., Magney, T.S., Vierling, L.A., Brown, T.T., & Huggins, D.R. (2014). LiDAR based biomass and crop nitrogen estimates for rapid, non-destructive assessment of wheat nitrogen status. *Field Crops Research, 159*, 21-32

Fournier, C., Andrieu, B., Ljutovac, S., & Saint-Jean, S. (2003). ADEL-wheat: a 3D architectural model of wheat development. *Proceedings of the 2003 Plant Growth Modeling, Simulation, Visualization, and Applications*, 54–63

Griebel, A., Bennett, L.T., Culvenor, D.S., Newnham, G.J., & Arndt, S.K. (2015). Reliability and limitations of a novel terrestrial laser scanner for daily monitoring of forest canopy dynamics. *Remote Sensing of Environment, 166*, 205-213

Ilangakoon, N.T., Gorsevski, P.V., & Milas, A.S. (2015). Estimating Leaf Area Index by Bayesian Linear Regression Using Terrestrial LiDAR, LAI-2200 Plant Canopy Analyzer, and Landsat TM Spectral Indices. *Canadian Journal of Remote Sensing, 41*, 315-333

Jelalian, A.V. (1992). *Laser Radar Systems*. Artech House

Korhonen, L., Korpela, I., Heiskanen, J., & Maltamo, M. (2011). Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sensing of Environment, 115*, 1065-1080

Kukko, A., & Hyyppä, J. (2009). Small-footprint laser scanning simulator for system validation, error assessment, and algorithm development. *Photogrammetric Engineering & Remote Sensing, 75*, 1177–1189

Lang, A. (1991). Application of some of Cauchy's theorems to estimation of surface areas of leaves, needles and branches of plants, and light transmittance. *Agricultural and forest meteorology, 55*, 191-212

Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. (2002). Lidar Remote Sensing for Ecosystem Studies Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience, 52*, 19–30

Lin, Y. (2015). LiDAR: An important tool for next-generation phenotyping technology of high potential for plant phenomics? *Computers and Electronics in Agriculture, 119*, 61-73

Liu, S., Baret, F., Andrieu, B., Abichou, M., Allard, D., Solan, B.d., & Burger, P. (2016). Modeling the distribution of plants on the row for wheat crops: Consequences on the green fraction at the canopy level. *Computer and Electronic in Agriculture, Under review*

Meng, X., Currit, N., & Zhao, K. (2010). Ground filtering algorithms for airborne LiDAR data: A review of critical issues. *Remote Sensing, 2*, 833-860

Morsdorf, F., Kötz, B., Meier, E., Itten, K.I., & Allgöwer, B. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment, 104*, 50-61

Neckar, P., & Adamek, M. (2011). Software and hardware specification for area segmentation with laser scanner SICK LMS 400. *JOURNAL OF SYSTEMS APPLICATIONS, ENGINEERING & DEVELOPMENT, 5*, 674-681

Nilson, T. (1971). A theoretical analysis of the frequency of gaps in plant stands. *Agricultural meteorology, 8*, 25-38

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica, 11*, 23-27

Pascal, V., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *The Journal of Machine Learning Research, 11*, 3371-3408

Pradal, C., Boudon, F., Nouguier, C., Chopard, J., & Godin, C. (2009). PlantGL: A Python-based geometric library for 3D plant modelling at different scales. *Graphical Models, 71*, 1-21

Pradal, C., Dufour-Kowalski, S., Boudon, F., Fournier, C., & Godin, C. (2008). OpenAlea: a visual programming and component-based software platform for plant modelling. *Functional Plant Biology, 35*, 751-760

Richardson, J.J., Moskal, L.M., & Kim, S.H. (2009). Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR. *Agricultural and forest meteorology, 149*, 1152-1160

Saeys, W., Lenaerts, B., Craessaerts, G., & De Baerdemaeker, J. (2009). Estimation of the crop density of small grains using LiDAR sensors. *Biosystems Engineering, 102*, 22-30

Stark, B., Zhao, T., & Chen, Y. (2016). An Analysis of the Effect of the Bidirectional Reflectance Distribution Function on Remote Sensing Imagery Accuracy from Small Unmanned Aircraft Systems. In, *2016 International Conference on*

Unmanned Aircraft Systems (ICUAS) (pp. 1342-1350). Arlington, VA USA

Stenberg, P. (2006). A note on the G-function for needle leaf canopies. *Agricultural and Forest Meteorology, 136*, 76

Van der Zande, D., Hoet, W., Jonckheere, I., van Aardt, J., & Coppin, P. (2006). Influence of measurement set-up of ground-based LiDAR for derivation of tree structure. *Agricultural and forest meteorology, 141*, 147-160

Verger, A., Vigneau, N., Chéron, C., Gilliot, J.-M., Comar, A., & Baret, F. (2014). Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sensing of Environment*

Weiss, M., Baret, F., Smith, G.J., Jonckheere, I., & Coppin, P. (2004). Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. *Agricultural and forest meteorology, 121*, 37-53

Zhao, K., García, M., Liu, S., Guo, Q., Chen, G., Zhang, X., Zhou, Y., & Meng, X. (2015). Terrestrial lidar remote sensing of forests: Maximum likelihood estimates of canopy profile, leaf area index, and leaf angle distribution. *Agricultural and forest meteorology, 209–210*, 100-113

Zhao, K., Popescu, S., Meng, X., Pang, Y., & Agca, M. (2011). Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sensing of Environment, 115*, 1978-1996

Zhao, K., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment, 113*, 182-196

Zhao, K.G., & Popescu, S. (2009). Lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA. *Remote Sensing of Environment, 113*, 1628-1645

Zheng, G., & Moskal, L.M. (2009). Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors (Basel), 9*, 2719-2745

4. Extraction of traits from the assimilation of multi-temporal GF measurements into ADEL-Wheat during the early stages

Monitoring the dynamics of canopy architecture during the early stages is highly desirable for breeders to select cultivars of high drought-tolerance, resistance to weed infestation, improved water-use efficiency and potentially high grain yield. These traits may be grouped into an early vigor trait. The retrieval of these traits requires assimilating multi-temporal observations within a dynamic model that simulates development of the canopy structure. The estimates of the parameters from the proposed assimilation approach corresponds to traits highly desired by the breeders because of their expected higher heritability since the effects of the environmental factors have already been taken into account through the processes described by the FSPM ADEL-Wheat model. In addition, knowledge of the ADEL-Wheat parameters will allow deriving emerging properties of the crop.

The objective of this chapter is to evaluate the feasibility and interest of a phenotyping data assimilation approach. For this purpose, an in silico experiment will be completed using the ADEL-Wheat model where the sowing pattern was described by the model developed in Chapter 2. The green fraction observed from several view directions and dates is the variable that is assimilated. It was preferred to the LiDAR measurements since for these early stages when the stems are not yet elongated, the Lidar provides poorer information as compared to RGB high resolution imagery. A sensitivity analysis was conducted to evaluate the effect of the number and spacing of the observation dates as well as the number of view directions used. It will allow to compare with the simpler methods proposed in chapter 3 for which only instantaneous observations are exploited. This chapter corresponds to a draft journal article to be published in Environmental modelling and software: Extraction of traits from the assimilation of multi-temporal GF measurements into ADEL-Wheat during the early stages[6].

[6] Liu, S., Baret, F., Buis, S., Abichou, M., Andrieu, B., Kamran, I., & de Solan, B. (2017). Extraction of traits from the assimilation of multi-temporal GF measurements into ADEL-Wheat during the early stages. To be submitted to Environmental Modelling & Software.

4.1. **Extraction of traits from the assimilation of multi-temporal GF measurements into ADEL-Wheat during the early stages**

Shouyang Liua,*, Fred Baret[a], Samuel Buis[a], Mariem Abichou[b], Bruno Andrieu[b], Irfan Kamran[a] and Benoit de Solan[a, c]

a INRA EMMAH, UMR 1114 Domaine Saint-Paul, Site Agroparc, 84914 Avignon Cedex 9, France

b INRA-AgroParisTech, UMR 1091 EGC, 78850 Thiverval-Grignon, France

c ARVALIS-Institut du végétal, Station expérimentale, 91720 Boigneville, France

Abstract

Monitoring the dynamics of canopy architecture during the early stages is highly desirable for breeders to select cultivars of high drought-tolerance, resistance to weed infestation, improved water-use efficiency and potentially high grain yield. These desired traits may be grouped into early vigor trait. A dedicated in silico study was completed to evaluate the feasibility and interest of a phenotyping data assimilation approach. A sequential assimilation scheme was developed to limit the space of canopy realization since the simulations of the 3D canopy architecture as well as the rendering of the phenotyping measurements is computer demanding. The early growth was decomposed into 2 consecutive stages. For each stage, 5 main influential parameters were identified. The phenotyping measurements were consisting into RGB images taken at several dates along the growth cycle and from several view inclination angles. For the first stage, the 5 influential parameters were estimated using a neural network machine learning technique. For the second phase, the 5 additional influential parameters were adjusted using the same neural network based technique while exploiting as prior information the posterior values of the 5 parameters defined during the first phase assimilation process. Results show that the parameters defining the size and rate of change of the green canopy crop elements were estimated with a relatively good accuracy. Conversely, the parameters driving the orientation and inclination of the elements were retrieved with a poor accuracy. However, some important emerging properties were very well retrieved, including the GAI and the number of tillers with more than 3 leaves at the beginning of stem elongation. Further detailed sensitivity analysis highlights the importance of getting enough dates of measurements covering the growth cycle, while dates separated by less than a phyllochron do not provide significant additional information. The sensitivity analysis highlighted also the importance of the inclined observations. The assimilation approach proposed appears very general and could be applied to a number of phenotyping measurements completed along the growth cycle. Nevertheless, because our results were achieved using in silico experiments, further validation is necessary using actual field measurements.

Keywords

Assimilation, ADEL-Wheat, Plant vigor, Canopy architecture, Neural network

1 Introduction

The rapid development of canopy architecture at early stages allows the crop to efficiently colonize resources, and limit soil evaporation and weed competition by a rapid canopy closure (Luquet et al. 2012). Hence monitoring the evolution of canopy architecture since emergence is desirable for breeders to select cultivars of high drought-tolerance, resistance to weed infestation, improved water-use efficiency and potentially high grain yield (Botwright et al. 2002; Ludwig and Asseng 2010). However, this task is often accomplished by manual field investigations which are generally too expensive to be applied over large spatial scales or large of micro-plots. Fortunately, recent advances in phenotyping techniques offer opportunities to facilitate the characterization of early vigor.

Image-based methods are the mainstream of modern phenotyping techniques (Paproki et al. 2012). A large diversity of image processing tools have been developed to assist the extraction of morphological traits from imagery, such as wheat lamina shape (Dornbusch and Andrieu 2010), plant height (Hartmann et al. 2011; Reuzeau et al. 2010), leaf angle (Cabrera-Bosquet et al. 2016). These tools are easy to use but not always fully automatic, limiting their wide acceptation and use by the community. More importantly these methods do not exploit the whole information content in the images. Nevertheless, the characteristics already accessible could be instrumental to identify early vigor which depends on complex and dynamic trade-offs among several elementary traits. Alternatively, three dimensional (3D) reconstruction of the canopy may provide enough details on the plant structure to provide ways to estimate some structural variables (Gibbs et al. 2016). Following this approach, Duan et al. (2016) applied multiview images to reconstruct 3D wheat architecture at early stage to extract morphological traits of interest. However, this method has been only validated on individual plant under well controlled illumination condition (Duan et al. 2016), while 3D reconstruction techniques appear to be vulnerable to the variations in the illumination conditions, suffer from errors and omissions introduced by occlusions and have difficulties on reflective surfaces because of saturation of the measured signal (Gibbs et al. 2016). Additionally, this approach is relatively demanding in terms of computation cost (Cobb et al. 2013). As a consequence of all these limitations, existing phenotyping techniques still cannot fully satisfy the need of breeders to characterize plant vigor at canopy level under field conditions.

Despite the advances in sensor technology and phenotyping platforms, the data interpretation appears to be the bottleneck to break for offering more traits of interests to breeders. Radiative transfer (RT) models describe the interaction between canopy characteristics and light. The satellite remote sensing community has been demonstrating since the last decades that RT bridges remote sensing observations with canopy traits (Baret and Buis 2008; Combal et al. 2003; Prévot et al. 2003; Weiss et al. 2000). Further the assimilation of RT models coupled with crop models provides access to more eco-physiological traits (Moulin et al. 1998; Zhang et al. 2016). This approach could be directly used in the interpretation of phenotyping observations (Verger et al. 2014a) due to the fact that phenotyping observations (proximal sensing) use the similar sensors as those used for aerial or satellite remote sensing. However, the strong assumptions on canopy architecture and approximations made in most RT models limit the use of this approach to relatively simple traits. This may satisfy the needs for ecosystem observation at large scale, while over close range observations more details are expected with more accuracy to answer the breeders requirements. Alternatively functional-structural plant models (FSPMs) explicitly describe both the 3D canopy structure and functioning processes (Vos et al. 2010). FSPMs have been inverted to estimate input parameters in the context of model calibration (Guo et al. 2006; Kang et al. 2008a; Kang et al. 2008b; Martre et al. 2015). However, their use has not been extended to retrieve canopy traits from phenotyping observations. Potentially more detailed

parameterization of structural and functioning processes in FSPMs will assist the interpretation of phenotyping observations. This will provide access to more traits of interests and with a better accuracy than the classical RT models could achieve from instantaneous observations and without the exploitation of our knowledge on the physiological processes involve in canopy structure development and plant functioning.
Given RT model, the retrieve scheme is generally driven by multi-spectral observations, while through FSPMs, it can theoretically be driven by different single or multi phenotyping observations. Simulators mimicking properties of typical imaging sensors could couple with FSPMs to simulate corresponding observations. Digital camera has been popular in phenotyping due to its high resolution and affordable cost (Baret et al. 2010a; Li et al. 2014b). Hence in this work, our objective is to assimilate sequential RGB imagery with 4D wheat model during the early stage to facilitate the identification of early vigor.

2 Materials and methods

2.1 Adaptation of ADEL-Wheat model

ADEL-Wheat 4D model (Abichou et al. 2013a; Fournier et al. 2003) was used to simulate the evolution of the 3D canopy structure. Wheat development is primarily driven by temperature but also affected by vernalization and photoperiod which impacts the final number of leaves on the main stem (Brooking et al. 1995; Jamieson et al. 2008). The thermal time elapsing between the visual appearance of two successive leaf tips, i.e. phyllochron is constant from seedling to flag leaf expansion (Hokmalipour 2011; Jamieson et al. 2008). In the phenological framework of ADEL-Wheat model, the plant development is mainly characterized by the Haun stage of the mainstem as a function of the thermal time or the phyllochron (Abichou et al. 2013a; Fournier and Andrieu 1999). ADEL-Wheat enables to simulate a large range of final dimensions of organs and as consequence a wide variety of canopy structure under different climate, soil and management practices conditions. However, the large number of parameters required induces a severe dimensionality problem when parameters need to be estimated from ground experiments. Parameterization efforts were made to reduce the number of input parameters and thereby better suits ADEL-Wheat model to model calibration.
Temperature significantly impacts the final dimension of organs on the mainstem. Long photoperiod and high temperature normally leads to very large increase in lamina length and sheath length (Hay 1990; Pararajasingham and Hunt 1996). However, the mechanism is still not clear enough to model the dimension of organs as a function of local environmental conditions (Nagelmüller et al. 2016). Therefore, statistical models are here proposed to parameterize the main stem structure based on a representative dataset covering 28 different winter wheat experiments with varied sowing date, cultivars and nitrogen levels from several years in Grignon, France. Details on the experimental conditions could be found in **Table 1**. Additionally, this statistical analysis would add prior knowledge to the proper range of some input parameters.

Table 1 Treatments in the dataset

Experiment date	Genotype	Sowing Date Level	Density	Nitrogen Level
1998-1999	Soisson	standard	250	standard
	Soisson	standard	70	standard
2005-2006	Soisson	standard	250	standard
	Caphorn	standard	250	standard
	Caphorn	standard	250	low
	Soisson	standard	250	low
2007-2008	Soisson	late	228	standard
	Soisson	late	514	standard
	Soisson	late	70	standard
	Soisson	early	228	standard

	Soisson	early	514	standard
	Soisson	early	77	standard
2008-2009	Soisson	late	228	standard
	Soisson	late	514	standard
	Soisson	late	77	standard
	Soisson	early	514	standard
	Soisson	early	77	standard
2012-2013	Apache	standard	220	standard
	Caphorn	standard	220	standard
	Maxwell	standard	220	standard
	Renan	standard	220	standard
	Soisson	standard	220	standard
2013-2014	Apache	standard	70	standard
	Caphorn	standard	70	standard
	Caphorn	standard	200	standard
	Renan	standard	70	standard

2.1.1 Lamina length

The final lamina length with respect to phytomer order (**Fig. 1**) from the bottom to the top leaf can be grouped into four phases. For the first phase, the lamina length varies gradually up to the second leaf (phytomer order N1=3). Then, the length plateaued up to approximately leaf 6 (N2=5.6±0.6). This is followed by a significant increase of leaf length with the leaf order (phytomer order) until reaching a maximum for N3=10 or 11). Finally, the length of the last one or two leaves decreases, corresponding to the flag and penultimate leaves. This change pattern is quite general for winter wheat the cases available and is consistent with finding of Evers et al. (2005b).

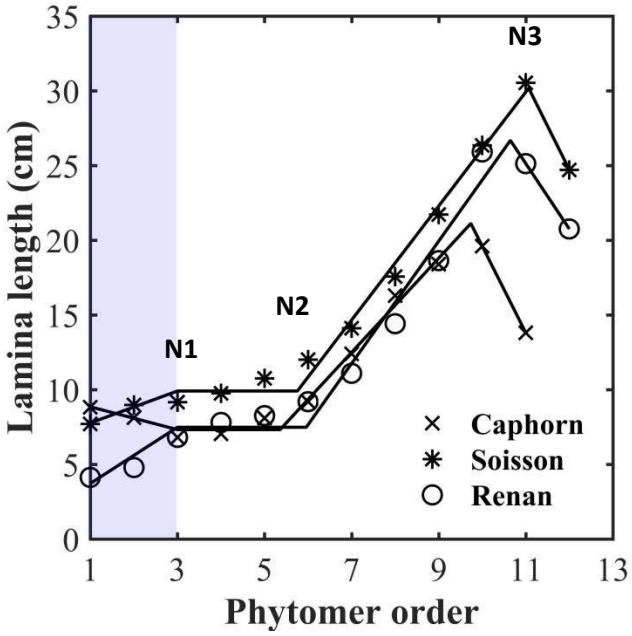

Fig. 1 Final lamina length with respect to phytomer order from bottom to top. The transitions between phases are called N1, N2 and N3.

A four-phases linear model was consequently proposed to describe leaf final lamina length as a function of the phytomer order (**Eq. 1-3**). All the treatments can be well fitted with $R^2 > 0.95$. N_1 is quite consistent with three leaves corresponding to the start time of tillering (Kirby et al. 1985). N_2 represents the division between juvenile and adult phase (Kerstetter and Poethig 1998). For our cases, N_2 is observed for a number of phytomers equals to 5.6 with a standard deviation 0.8.

$$L_{lamina\_n} = \begin{cases} L_{lamina\_1} + \alpha_{lamina\_N1}(n-1) & 1 \le n < 3 \\ L_2 & 3 \le n < N_2 \\ L_2 + \alpha_{lamina\_N3}(n-N_2) & N_2 \le n < N_3 \\ L_3 + \alpha_{lamina\_N4}(n-N_3) & N_3 \le n \le N_{phy} \end{cases} \quad (1)$$

$$L_2 = L_{lamina\_1} + 2\alpha_{lamina\_N1} \quad (2)$$

$$L_3 = L_2 + \alpha_{lamina\_N2}(N_3 - N_2) \quad (3)$$

Where L_{lamina_n} if the final length of the lamina for the phytomer number n, N_2 and N_3 refer to the phytomer order at the end of the 2nd and 3rd phases. N_{phy} is the maximum phytomer number of mainstem. α_{lamina_N1}, α_{lamina_N3} and α_{lamina_N4} correspond to the change slope of lamina length during 1st, 3rd and 4th, respectively.

2.1.2 Lamina width

Lamina width is linearly related to the lamina length from the first phytomer up to $n = N_3$. (**Fig. 2**). For $n > N_3$, which corresponds to the flag leaf and penultimate leaf lamina, the lamina width continues to increase although the lamina length decreases (Dornbusch et al. 2011c).

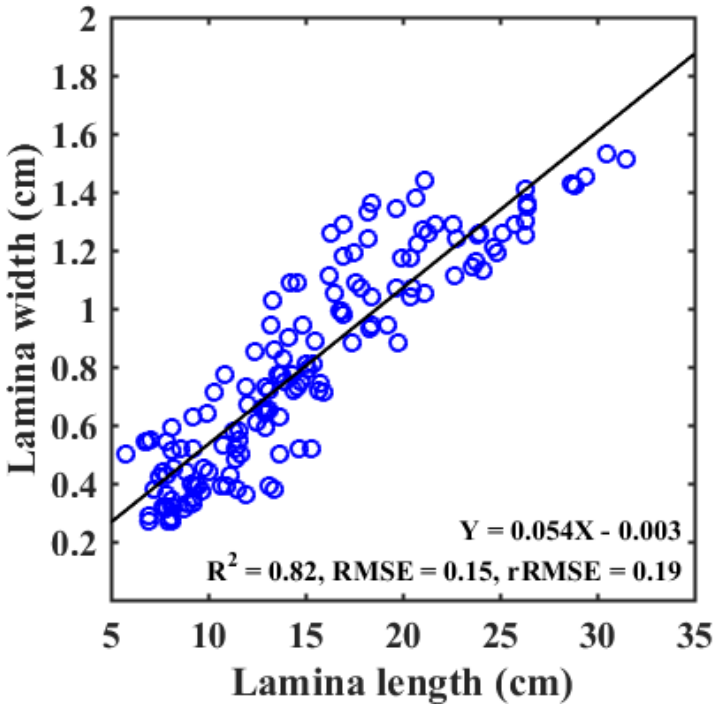

Fig. 2 Lamina length as a function of lamina width for n<N3.

2.1.3 Sheath length and width

Sheath length linearly correlates with lamina width (**Fig. 3**). Similar correlation was reported by previous work (Evers et al. 2005b; Lock 2003). Regarding sheath width, it normally increases continuously as $n < N_2$ and then it varies. No significant correlation was found between sheath width and sheath length or other variables for $n > N_2$. However, as it impacts marginally the canopy structure, a linear model was calibrated to estimate the sheath width from the sheath length.

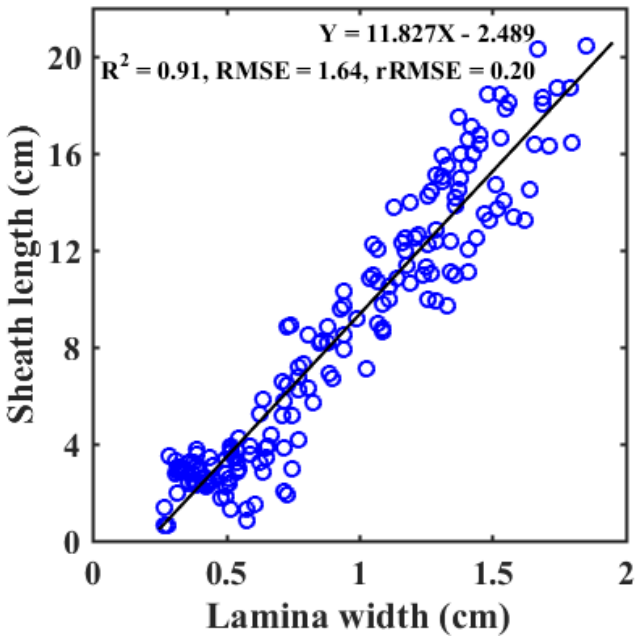

Fig. 3 Lamina width VS. sheath length

2.2 Assimilation of ADEL-Wheat

The green fraction (GF) refers to the fraction of green vegetation in the image. The GF in the sun direction corresponds to the light fraction intercepted by the green elements which drives photosynthesis and transpiration processes (Baret et al. 1993a; Baret et al. 1993b). Hence sequential observations of multi-directional GF were employed as the features to be assimilated in the ADEL-Wheat model (**Fig. 4**). Focus was put on the early stages spanning from emergence to the start of stem elongation that would allow characterizing the early vigor that plays a key role in the development of the rooting system and for the competition with weeds. This period was further divided into two distinct phases: before (P_{bf_tiller}) and after (P_{af_tiller}) tillering. The start of tillering was assumed occurring at Haun stage 3 according to several authors including (Masle 1985b). The assimilation process (that may be regarded as some kind of inversion of the ADEL-Wheat model) was firstly conducted for parameters involved in the P_{bf_tiller} phase. Similar assimilation process was run for the P_{af_tiller} phase using the parameters previously derived for the P_{bf_tiller} phase.

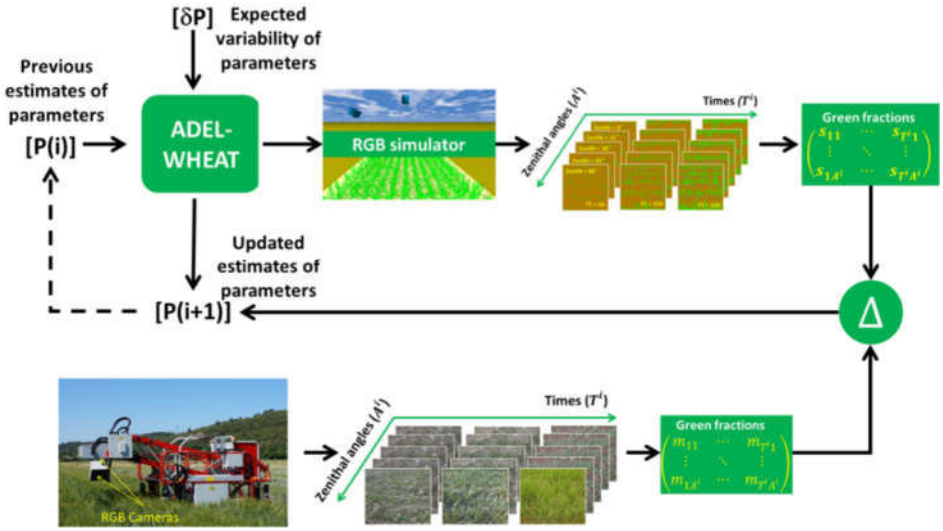

Fig. 4 Diagram of the assimilation of RGB imagery into ADEL-Wheat model

2.2.1 Parameters to be estimated

The selection of assimilation parameters is always a difficult task that most of the time results from a trade-off between completeness and realism on one side, and accuracy and operationalization on the other side. With the development of the crop, more parameters gradually should be involved to characterize the evolution of the canopy architecture within the ADEL-Wheat model. During P_{bf_tiller}, thanks to our efforts in the parameterization of the mainstem structure (at this stage only the main stem is present), we demonstrated that five parameters would be sufficient (**Table 2**):

- ***The phyllochron***, TT_{phy} controls the time of leaf appearance and the rate of leaf extension.
- ***The lamina length*** of the first 3 leaves is assumed to change linearly, which is parameterized by the lamina length of the first leaf L_{lamina_1} and the slope, α_{lamina_N1} (**Eq. 1**).
- ***Leaf orientation and inclination***. The leaf orientation is initialized from the seeding stage depending on seed orientation. However, seeds are assumed to be sown with a random azimuth which is further corroborated by (Ledent and Moss 1977). Hence for the sake of simplicity, the azimuth of the first leaves was orientated randomly. Evers et al. (2005a) found that the azimuth of successive leaves was mainly opposite for the from the 1st to the 3rd phytomers. However different genotypes show a variability in the leaf orientation distribution pattern depending on the phenotypic plasticity (Maddonni et al. 2001). The azimuth angle of leaves relative to the previous one was assumed drawn from a Gaussian distribution with mean 180° and standard deviation $Ang_{\Delta leaf}$ accounting for the plasticity found among cultivars. the leaf inclination was described based on experimental observations. Variations of leaf inclination was manipulated by shifting the basal inclination controlled with parameter $Ang_{\Delta leaf}$.

During P_{af_tiller}, five additional parameters were introduced to depict the structural and quantitative dynamic of tillers (**Table 2**). Therefore, a total of 10 parameters are necessary to characterize canopy structure during this stage: 5 parameters for the stage P_{bf_tiller}, and the five additional ones specific to the stage P_{af_tiller}.

Table 2 Parameters of ADEL-Wheat that will be tuned during the assimilation process for the 2 stages considered. The corresponding range of variation used in the generation of training dataset is also indicated.

Stages	Abbr.	Descriptions	Min	Max	Unit
$\boldsymbol{P_{bf\_tiller}}$	TT_{phy}	Phyllochron: thermal time interval between the emergence of successive leaf	80	120	°C · d
	$Ang_{\Delta leaf}$	The standard deviation of the leaf azimuth compared to the previous one with mean 180°	0	90	degree
	Ang_{shift}	The shift of leaf basal inclination	-15	15	degree
	L_{lamina_1}	The length of lamina one	4	8	cm
	α_{lamina_N1}	The increase rate of lamina length during phase N1	-3	3	scalar
$\boldsymbol{P_{af\_tiller}}$	N_{tiller}	Number of tillers per plant	0	5	scalar
	Ang_{tiller}	The inclination of the base of the tiller relative to main stem	10	85	degree
	α_{tiller}	The change rate of tiller inclination angle (it starts at Haun stage =1 of the corresponding axes)	10	50	degree/Haun stage
	S_{start}	Number of green leaves at the start of leaf senescence on the main stem	3.5	6.5	scalar
	S_{min}	Number of green leaves at $T_{HS=N_{leaf\_main}-4}$ on the main stem	1.5	3.5	scalar

2.2.2 Synthetic dataset

In silico experiments were conducted by simulating RGB images of wheat canopies using the ADEL-Wheat model coupled with the POV-Ray (Persistence of Vision Raytracer, Version 3.7) ray tracing code. To cover the range of variation of canopy architecture, 2500 sets of parameters in stage P_{bf_tiller} were generated using a Latin Hypercube sampling scheme. The 5 parameters listed in **Table 2** for P_{bf_tiller} were assumed following uniform distributions within the sampling range (**Table 2**). The range was defined according to the expert knowledge developed by the Grignon INRA team. For the P_{af_tiller} stage, similar sampling strategy was used for the 5 additional parameters involved (**Table 2**). However, parameters derived from the previous stage P_{bf_tiller} were considered associated to uncertainties assumed Gaussian distributed with the standard deviation equal to the RMSE associated to their estimates over the P_{af_tiller} period.

For each subset of parameters, ADEL-Wheat model was run at 5 different thermal time, corresponding to the time of the GF measurements (**Table 3**). The output of each run consists in 3D scenes of the canopy and the corresponding emergent properties, including green area index (GAI), density of axes, Haun stage of each axis. The scenes simulated are 2 m by 2 m size, including 11 rows with density 250 plants·m^{-2} with an inter-row spacing of 17.5 cm. The 3D scenes were then rendered by the GF simulator under different configurations (**Table 3**). The camera with a 20° field of view was set looking across the row direction from 5 zenith angles, 0°, 15°, 30°, 45° and 60°, respectively. To reduce computation cost, images were rendered at a relatively low resolution with 500 by 500 pixels: the spatial resolution finally achieved is around 1 mm which appears sufficient for the computation of the GF over the scene. As a matter of facts, the green color in the rendered images is perfectly defined. Consequently, no segmentation error will occur in the calculation of GFs. Conversely, the variation of illumination under field conditions may induce errors into GFs because of possible confusions between green vegetation and non-green elements or the soil surface. Hence, different levels of noise were added to the 'measured' green fraction to mimic the actual observations. The noise was assumed following Gaussian distribution with zero mean and standard deviation as 0%, 5% or 10% which are typical accuracies reported in green segmentations. The total number of simulations for each stage was 2500.

Table 3 Experimental design used to generate the synthetic dataset

Stages	Measurements' dates (℃ · d)	Zenithal angles (°)	Number of combinations
P_{bf_tiller}	50, 100, 150, 200 and 250	0, 15, 30, 45, 60	2500
P_{af_tiller}	300, 400, 500, 600 and 700		

2.2.3 Assimilation method

The assimilation will here consist in adjusting some of the parameters of the ADEL-Wheat model (**Table 2**) to get a good agreement with the GF observed from phenotyping sensors at several dates along the growing cycle. The adjustment of the parameters can be achieved using either using a Look-Up-Table (LUT) or a Neural Network (NN) methods that are both proved to be computationally efficient as compared with other methods (Kimes et al. 2000). The LUT technique focuses on searching for the solution that corresponds to the minimum distance between the GF observed and those simulated. Conversely, NN minimizes the distance in the space of the model parameters. NN shows often better performances for problems of high level complexity given a small dimensionality of input metric (Atzberger 2004; Kimes et al. 2002), which is the case of our retrieval problem. This is the reason why we selected this technique.

NN can be considered as a black box that fits a relationship between observations and predictions. The training process requires first to define the network architecture (the number of hidden layers and nodes per layer and the type of transfer function of each neuron). Then the synaptic weights and biases are tuned to best match the output values over a training database. A one-layer feed-forward network was employed here with tangent sigmoid transfer functions in the first layer and a linear transfer function in the output

layer. In the training process, the weight and bias values were updated according to Levenberg-Marquardt optimization. Unfortunately, there are no general rules to select the optimal number of neurons. A rough approximation can be obtained by the geometric pyramid rule proposed by Masters (1993). For a network with n inputs and m outputs neurons, the optimal configuration may be $\sqrt{nm}$ neurons in the hidden layer. To compare the performance of NNs with different number of inputs, this method was employed to define the number of neurons in the hidden layer. Calibration and validation processes were conducted with respectively 85% and 15% of the synthetic dataset.

3 Results and discussion

3.1 Assimilation results of P_{bf_tiller} stage

Green area index (GAI) is one of the key variables to characterize the development of the canopy structure. As shown in **Fig. 5**, very good accuracy was achieved with rRMSE 6%. Besides, the performance seems quite robust as the accuracy degrades slightly with the increasing of noise. Baret et al. (2010a) demonstrated that based on the Poisson model, GAI could be estimated with rRMSE 12% using single green fraction measurements from 57.5° zenith angle. The improved performances shown in our work should come from the increasing information quantity used in the assimilation scheme, with more directions and more dates of observation. In addition, the assimilation method ensures to get a consistent time course of GAI during P_{bf_tiller} stage based on the temporal constraints provided by the use of the dynamic ADEL-Wheat model.

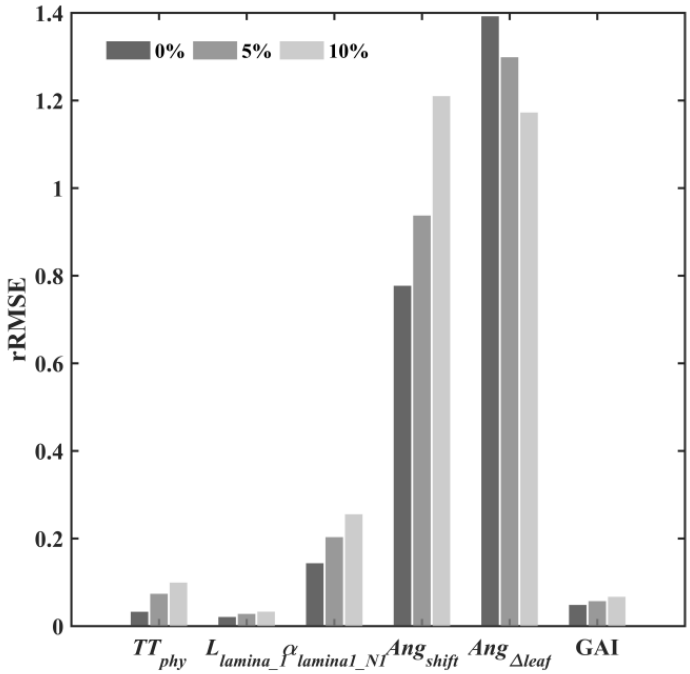

Fig. 5 Accuracy of 5 the estimation of the 5 parameters and GAI during P_{bf_tiller} stage using the GF contaminated with different levels of noise and using the 5 directions concurrently.

GAI linearly is equivalent to the (total) leaf area index during the P_{bf_tiller} stage when senescence has not yet started. Hence accurate estimates of GAI imply good estimation of the parameters controlled, namely by TT_{phy}, L_{lamina_1} and α_{lamina_N1}. **Fig. 5** confirms that very good estimates of TT_{phy} (rRMSE=10%) and L_{lamina_1} (rRMSE = 3%) were achieved even with 10% noise on GF measurements. Regarding α_{lamina_N1}, the accuracy is acceptable until medium noise, 5%, with rRMSE=20% (**Fig. 5**). The phyllochron varies among

cultivars (Hay and Kirby 1991) and may be impacted by the combination of temperature and day length (Cao and Moss 1989). Likewise, environmental factors have significant effects on the final length of leaves. Hence, the heritability of TT_{phy} and L_{lamina_1} may be doubtful, thereby questioning their qualification as secondary traits in characterizing plant vigor or the adaptiveness to abiotic stresses (Edmeades 1996; Monneveux et al. 2012). Conversely, the correlation between the lamina length of different orders, α_{lamina_N1}, may be more heritable determined and may serve as a robust trait.

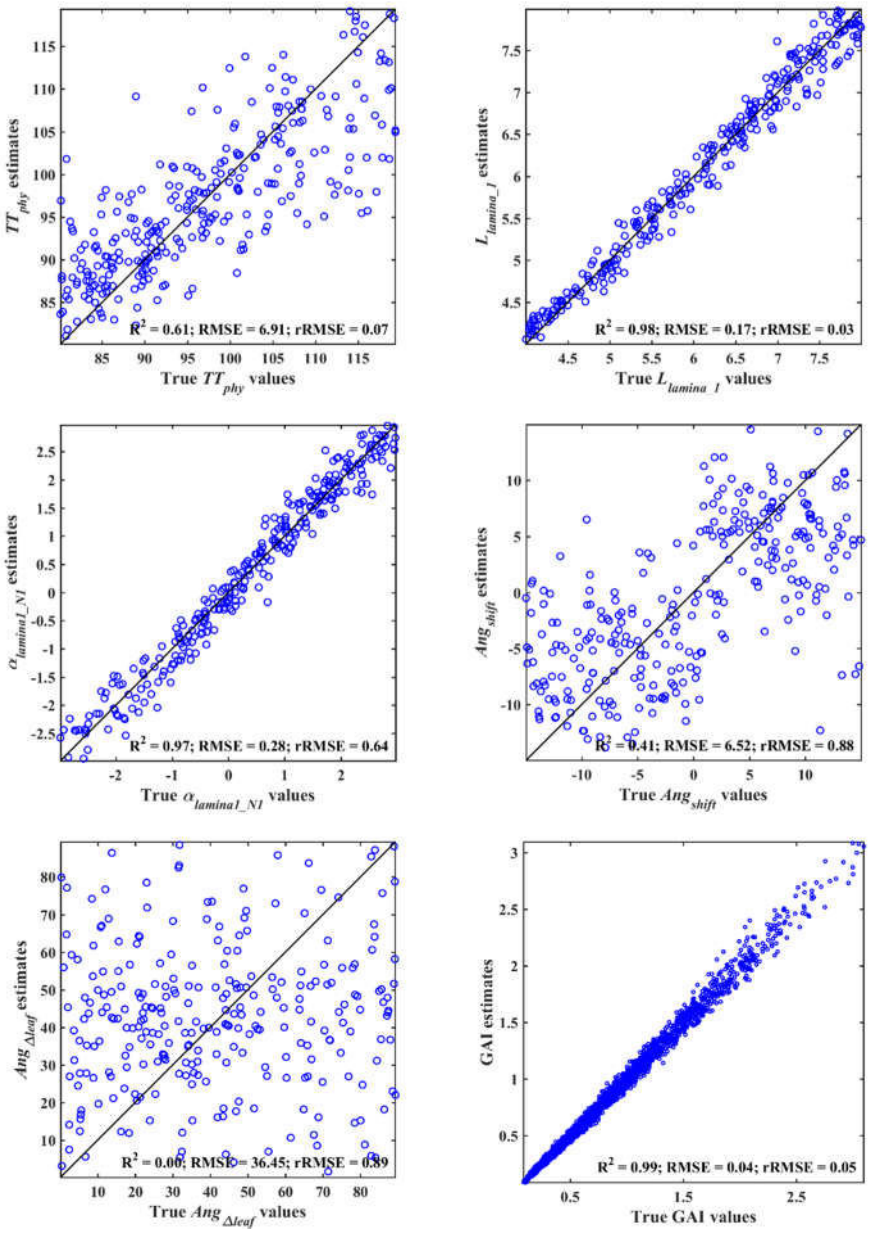

Fig. 6 Comparison of the estimated values of the 5 parameters and GAI (computed over the 5 dates of GF measurements) with the true values. This corresponds to the assimilation of GF measurements with 5% noise in 5 directions for the stages before 3 leaves.

Orientation and inclination of leaves are important features that characterize the canopy structure, manipulated by parameters Ang_{shift} and $Ang_{\Delta leaf}$ in the ADEL-Wheat model (**Table 2**). Leaf inclination, Ang_{shift}, controls the inclination of leaves around the basal and determines the average inclination angle of the canopy. The relatively narrow range of variation of the Ang_{shift} values (**Table 2**) explains the difficulty to retrieve accurately these values (**Fig. 5** and **Fig. 6**). This situation is even worse for $Ang_{\Delta leaf}$ that is

impossible to retrieve accurately from GF observations (**Fig. 5** and **Fig. 6**). Parameter $Ang_{\Delta leaf}$ characterizes the variation of relative azimuth angle at the plant (and even leaf) scale. As a matter of facts, the canopy is not well developed during the P_{bf_tiller} stage: there are almost no interactions between neighboring plants. This explains why the azimuthal orientation pattern of leaves is not accessible from observations at the canopy scale during this early development phase. The initial orientation pattern of the first leaves follows an almost random distribution. This situation will change with the growth and development of plants. Different leaf orientation patterns will later result in significant effect on light interception (Maddonni et al. 2001).

3.2 The optimal configuration

We demonstrated earlier that the additional information coming from the concurrent exploitation of the temporal and directional dimensions was improving significantly the estimates of integrated canopy characteristics such as the GAI. However, the impact of this additional information on canopy characteristics retrieval performances need to be better quantified. In the following, we focused on the performances quantified as the average rRMSE measured on the 3 parameters that were demonstrated to be well retrieved (TT_{phy}, L_{lamina_1} and α_{lamina_N1}) as well as the GAI. For GAI, rRMSE was computed over all the fates of GF observations available. For the sake of brevity and simplicity, the GF used in this exercise were corresponding to the medium 5% Gaussian noise. The several combinations of view directions (31 combinations) and dates of observations (31 combinations) resulted in 961 total cases.

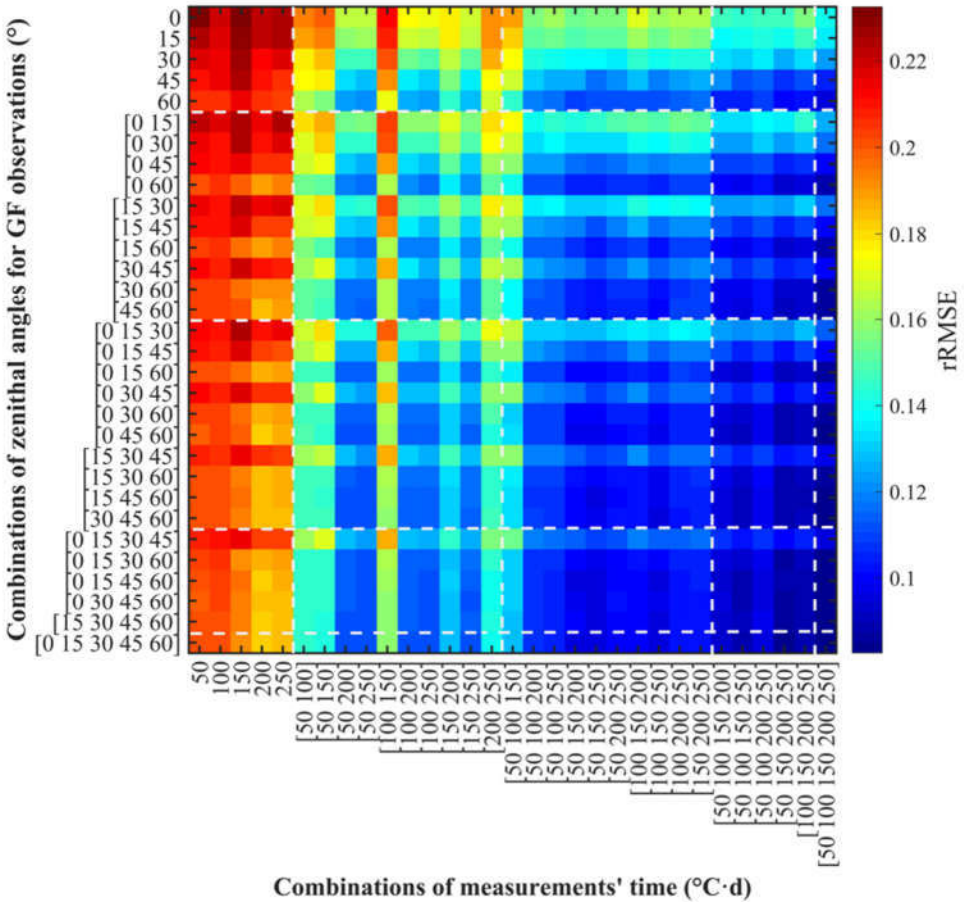

Fig. 7 The average rRMSE of the estimation of the three ADEL-Wheat parameters (TT_{phy}, L_{lamina_1} and α_{lamina_N1}) and GAI obtained from the several combinations of dates and directions.

Results (**Fig. 7**) shows that single dates observations lead to very poor estimates, which is well explained by the fact that it is obviously impossible to get a reliable estimates of parameters driving the canopy structure dynamics such as TT_{phy} just from a single date GF measurement. At least 2 dates are required and if possible as different as possible: when just 50 °C · d separates two consecutive measurements, performances are much lower (higher rRMSE) as compared to more distant observations. Again, this appears quite understandable since the information on canopy dynamics when observations are too close do not bring much additional information on the evolution of the canopy. Most of the effects will be to reduce the influence of the GF noise on parameter and GAI estimations. At least 100 °C · d should separate 2 consecutive observations. This corresponds roughly to a Phyllochron. Best performances are achieved when the span of dates is maximal. Best performances are anyway achieved with all the dates are used since it allows the best estimates of dynamic parameters such as TT_{phy} in addition to reduce the effect of the GF noise par multiplying independent observations.

The sampling of directions appears also critical. More inclined directions are generally performing better than more vertical ones. This is verified when using single directions as well as wen using concurrently several directions. The more inclined directions provide always improved performances for parameter and GAI retrievals. At least 2 directions are required to reach optimal performances (**Fig. 7**).

In the following, the assimilation during the P_{af_tiller} stage will be completed with the full set of dates and directions. The sensitivity analysis presented in this section was not repeated because most of the conclusions are expected to be very similar for this phase.

3.3 Assimilation results during P_{af_tiller} stage

The assimilation of 5 additional GF observations made under the 5 directions after the start of tillering phase (**Table 2**) was achieved by tuning 5 additional parameters (**Table 2**) in addition to the 5 parameters previously estimated during the P_{bf_tiller} phase. However, the range of variation of the previously estimated parameters was restricted to the confidence interval associated to their estimation from the first assimilation phase. In addition to the GAI emerging property already computed during the first development phase, another emerging property was computed: $\mathrm{D_{axes}}$. It characterizes the density of tillers having at least three leaves at the very beginning of stem elongation. As a matter of facts, tillers having 3 leaves can continue to grow while the others will regress due to the competition between neighboring plants (Masle 1985a). The variable $\mathrm{D_{axes}}$ is therefore strongly related the ear density and consequently to the potential yield.

Results (**Fig. 8**) show that the parameters [TT_{phy}, L_{lamina_1}, α_{lamina_N1}, S_{start}] and the integrated variables [GAI, D_{axes}] have an average rRMSE<20%. The parameters corresponding to the first period P_{bf_tiller} were generally better estimated during this second assimilation phase (compare **Fig. 6** and **Fig. 8**): the additional constraints imposed by the observations during the second phase allowed to improve their estimation. However the 2 parameters driving the leaf inclination [Ang_{shift}, $Ang_{\Delta leaf}$] are still poorly estimated. The same is observed for geometry parameters [Ang_{tiller}, α_{tiller}] that active for the second phase P_{af_tiller}. More surprisingly, the parameter driving the number of tillers produced per plant, N_{iller}, was not well estimated. The parameter S_{min} driving the number of green leaves of the tillers is not well retrieved since its influence may be limited to the very last GF observations. Conversely, the parameter defining the start of senescence, S_{start}, is well retrieved since it drives the GAI dynamics during this second phase, hence the GF observed. Parameter S_{start} has been proposed as a secondary trait for drought tolerance (Araus et al. 1997; Hafsi et al. 2007).

As a consequence of the parameter estimation during this second development stage, the corresponding emerging properties such as GAI and $\mathrm{D}_{\mathrm{axes}}$ are well estimated (**Fig. 8**).

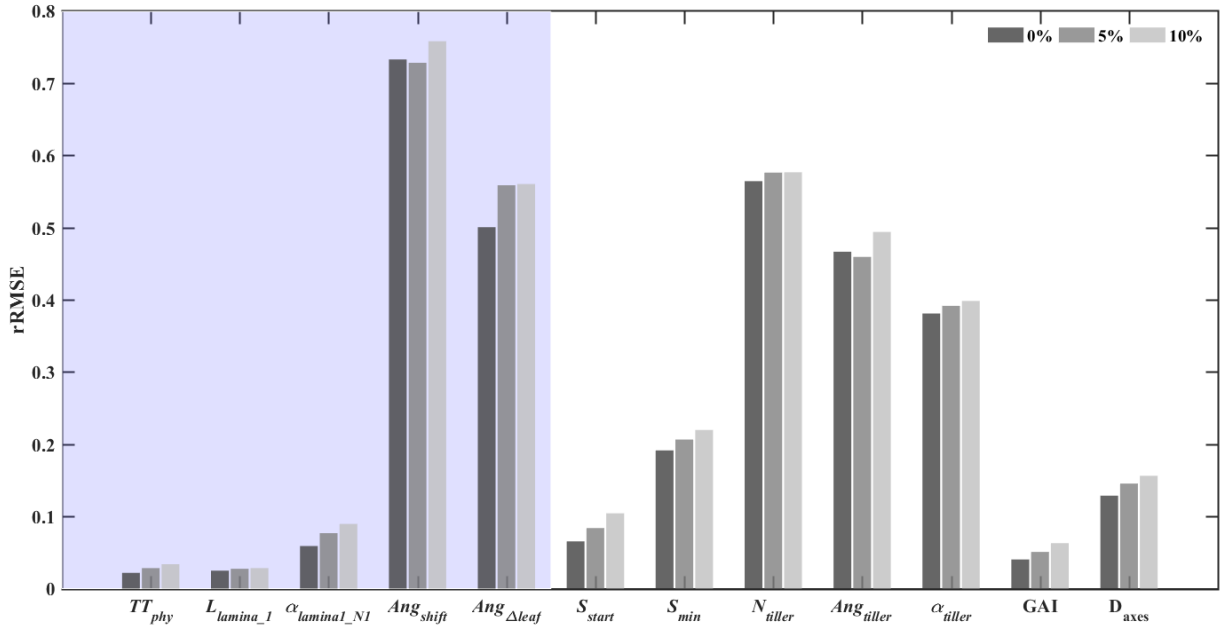

Fig. 8 The retrieval accuracy during the second development stage P_{af_tiller}**.**

4 Conclusions

The main objective of this study was to develop a method to estimate some of the ADEL-Wheat parameters and corresponding integrated variables such as GAI from the observations of green fractions (GF) made over time. Focus was put on the early stages corresponding to the installation of the plants. These early stages of wheat crops are critical to explain the later performances of the crop, particularly in terms of rate of soil coverage, root development (Baret et al. 1992) and the relationship with the potential biomass and yield. The proposed assimilation approach integrates into a consistent system the information coming from the GF observations performed along the growth cycle. It allows accessing some parameters of the ADEL-Wheat model including those driving the rate of leaf emission (the phyllochron), early vigor traits such the length of the first lamina and its rate of elongation. In addition, the parameter governing the delayed senescence of leaves is also accessible and that has already been proposed as a secondary trait for drought resistance (Araus et al. 1997; Hafsi et al. 2007). Emerging properties could be computed from the ensemble of parameters estimated. They include at least GAI and the number of tillers with more than 3 leaves at the start of stem elongation that may be used by plant breeders and agronomist into their applications (Masle 1985a).

The performances of retrieval of these variables are generally improved as compared to the use of single date or single direction observations because of the constraints imposed by the knowledge on the development processes embedded in the ADEL-Wheat model. In addition, the multiplication of measurements exploited allows mitigating the noise associated to each individual measurement. The sensitivity analysis on the sample of dates and directions required to get good estimation performances show that dates should be separated by at least 100 $℃ \cdot \mathrm{d}$ while the inclined directions appear important at these early stages. However, this study shows also that the GF information provides only little information of the orientation and inclination of the leaves. Other information should probably be input to the assimilation

system, particularly regarding the features contained in the images. As a matter of facts, green fractions represent a very small fraction of the information content in the RGB images used to compute GF.

The sequential assimilation scheme proposed here that decomposes the growth cycle into a series of successive development phases each of them driven by specific parameters allows to simplify the problem by splitting it into sub-problems. This non-global optimization approach could be considered as sub-optimal since the possible interactions between the several phases are not explicitly accounted for. However, the restricted degree of liberty given here in the second phase to the parameters estimated during the first phase provides a way to refine these estimates while not increasing too much the size of the problem. The NN approach used to relate the parameters to the input parameters or the emerging properties could have been more simply replaced by a look up table approach that would skip the training phase that is relatively demanding particularly in the case of the sensitivity analysis on the date and directions sampling. However, the main limiting factor in such study is still the building of the training database that requires a large number of simulations to be completed. The sequential approach proposed allows restricting the number of simulations for each phase by the restricted range of variation of the parameters that were estimated already during the previous phases. Further, the use of now currently available high computing capacity clusters will help solving this problem.

This in silico experiment has demonstrated the feasibility and interest of such an assimilation approach. However, its transfer to the actual world needs obviously to be tested in future works. Several limitations may already be pointed out. The first one corresponds indeed to the capacity of the ADEL-Wheat model to simulate with enough realism the variables that are measured with phenotyping techniques. In a previous work, we pointed out the limits of ADEL-Wheat to get realistic GF values because the possible interactions between leaves was not always accurately described. Further, the interval between the emergence of a leaf and the tiller in its cohort is 3 phyllochrons but it tended to be longer for the coleoptile and first secondary tiller. Additionally, tillering typically ceased at the elongation of the first internode when it is the terminal spikelet stage in wheat (Kirby et al. 1985). But this time is known to be strongly affected by external factors. Tillering ceases at specific light conditions within the wheat canopy, independent of population density (EVERS et al. 2006). Additionally shorter photoperiod will cease tillering before stem elongation (Miralles and Richards 2000). These features are not yet implemented in ADEL-Wheat. Anyway, the approach needs to be tested with the model in its current state to provide a first order evaluation in actual experimental conditions. The fixed dates used in this study should not pose a problem to be adapted to any date during these early stages. As a matter of facts, the dynamics of growth is smooth since it results from incremental growth and senescence processes. It will be therefore possible to interpolate the GF values between the fixed dates simulated in this exercise to match the actual dates of observations. Similar interpolation process could be also applied to match the actual observation directions used since the directional variation of GF is expected to be also very smooth. These GF observations could be computed from several possible vectors currently used in phenotyping applications: either using fixed sensors at the ground level, semi-automatic light systems such as the “phenotypette”, rover robot running on the ground such as the phenomobile, or UAV observations (Schirrmann et al. 2016)achieved either from very high resolution imagery or even from multispectral camera measurements that may be transformed into GF using vegetation indices as proposed by (Comar et al. 2012b) or radiative transfer model inversion (Li et al. 2015). However, in this later case, a more consistent approach based on the simulation of the multispectral signal recorded by the camera should be applied to exploit the detailed canopy architecture description provided by ADEL-Wheat.

Reference

Abichou, M., Fournier, C., Dornbusch, T., Chambon, C., Baccar, R., Bertheloot, J., Vidal, T., Robert, C., David, G., & Andrieu, B. (2013). Re-parametrisation of Adel-wheat allows reducing the experimental effort to simulate the 3D development of winter wheat. In, *7th International Conference on Functional-Structural Plant Models* (pp. 304-306)

Araus, J., Amaro, T., Zuhair, Y., & Nachit, M. (1997). Effect of leaf structure and water status on carbon isotope discrimination in field‐grown durum wheat. *Plant, Cell & Environment, 20*, 1484-1494

Atzberger, C. (2004). Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sensing of Environment, 93*, 53-67

Baret, F., Andrieu, B., Folmer, J., Hanocq, J., & Sarrouy, C. (1993a). Gap fraction measurement from hemispherical infrared photography and its use to evaluate PAR interception efficiency. *Crop structure and light microclimate*, 359-371

Baret, F., Andrieu, B., & Steven, M. (1993b). Gap frequency and canopy architecture of sugar beet and wheat crops. *Agricultural and forest meteorology, 65*, 261-279

Baret, F., & Buis, S. (2008). Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems. *Advances in land remote Sensing* (pp. 173–201): Springer

Baret, F., de Solan, B., Lopez-Lozano, R., Ma, K., & Weiss, M. (2010). GAI estimates of row crops from downward looking digital photos taken perpendicular to rows at 57.5° zenith angle: Theoretical considerations based on 3D architecture models and application to wheat crops. *Agricultural and forest meteorology, 150*, 1393-1401

Baret, F., Olioso, A., & Luciani, J.L. (1992). Root biomass fraction as a function of growth degree days in Wheat. *Plant and Soil, 140*, 137-144

Botwright, T., Condon, A., Rebetzke, G., & Richards, R. (2002). Field evaluation of early vigour for genetic improvement of grain yield in wheat. *Crop and Pasture Science, 53*, 1137-1145

Brooking, I.R., Jamieson, P.D., & Porter, J.R. (1995). The influence of daylength on final leaf number in spring wheat. *Field Crops Research, 41*, 155-165

Cabrera‐Bosquet, L., Fournier, C., Brichet, N., Welcker, C., Suard, B., & Tardieu, F. (2016). High‐throughput estimation of incident light, light interception and radiation‐use efficiency of thousands of plants in a phenotyping platform. *New Phytologist*

Cao, W., & Moss, D.N. (1989). Temperature and daylength interaction on phyllochron in wheat and barley. *Crop Science, 29*, 1046–1048

Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics, 126*, 867-887

Comar, A., Burger, P., de Solan, B., Baret, F., Daumard, F., & Hanocq, J.F. (2012). A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: description and first results. *Journal of Functional Biology*

Combal, B., Baret, F., Weiss, M., Trubuil, A., Mace, D., Pragnere, A., Myneni, R., Knyazikhin, Y., & Wang, L. (2003). Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment, 84*, 1-15

Dornbusch, T., & Andrieu, B. (2010). Lamina2Shape—An image processing tool for an explicit description of lamina shape tested on winter wheat (Triticum aestivum L.). *Computers and Electronics in Agriculture, 70*, 217-224

Dornbusch, T., Watt, J., Baccar, R., Fournier, C., & Andrieu, B. (2011). A comparative analysis of leaf shape of wheat, barley and maize using an empirical shape model. *Annals of botany, 107*, 865-873

Duan, T., Chapman, S., Holland, E., Rebetzke, G., Guo, Y., & Zheng, B. (2016). Dynamic quantification of canopy structure to characterize early plant vigour in wheat genotypes. *Journal of experimental botany, 67*, 4523-4534

Edmeades, G.O. (1996). *Developing Drought and Low N-tolerant Maize: Proceedings of a Symposium, March 25-29, 1996, CIMMYT, El Batán, Mexico*. CIMMYT

EVERS, J.B., VOS, J., ANDRIEU, B., & STRUIK, P.C. (2006). Cessation of Tillering in Spring Wheat in Relation to Light Interception and Red : Far-red Ratio. *Annals of Botany, 97*, 649-658

Evers, J.B., Vos, J., Fournier, C., Andrieu, B., Chelle, M., & Struik, P.C. (2005a). Towards a generic architectural model of tillering in Gramineae, as exemplified by spring wheat (Triticum aestivum). *New Phytologist, 166*, 801-812

Evers, J.B., Vos, J., Fournier, C., Andrieu, B., Chelle, M., & Struik, P.C. (2005b). Towards a generic architectural model of tillering in Gramineae, as exemplified by spring wheat (Triticum aestivum). *New Phytologist, 166*, 801–812

Fournier, C., & Andrieu, B. (1999). ADEL-maize: an L-system based model for the integration of growth processes from the organ to the canopy. Application to regulation of morphogenesis by light availability. *Agronomie, 19*, 313–327

Fournier, C., Andrieu, B., Ljutovac, S., & Saint-Jean, S. (2003). ADEL-wheat: a 3D architectural model of wheat development. *Proceedings of the 2003 Plant Growth Modeling, Simulation, Visualization, and Applications*, 54–63

Gibbs, J.A., Pound, M., French, A.P., Wells, D.M., Murchie, E., & Pridmore, T. (2016). Approaches to three-dimensional reconstruction of plant shoot topology and geometry. *Functional Plant Biology*

Guo, Y., Ma, Y., Zhan, Z., Li, B., Dingkuhn, M., Luquet, D., & De Reffye, P. (2006). Parameter optimization and field validation of the functional–structural model GREENLAB for maize. *Annals of Botany, 97*, 217-230

Hafsi, M., Akhter, J., & Monneveux, P. (2007). Leaf senescence and carbon isotope discrimination in durum wheat (Triticum durum Desf.) under severe drought conditions. *Cereal Research Communications, 35*, 71-80

Hartmann, A., Czauderna, T., Hoffmann, R., Stein, N., & Schreiber, F. (2011). HTPheno: an image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics, 12*, 148

Hay, R. (1990). The influence of photoperiod on the dry matter production of grasses and cereals. *New Phytologist, 116*, 233-254

Hay, R., & Kirby, E. (1991). Convergence and synchrony-a review of the coordination of development in wheat. *Crop and Pasture Science, 42*, 661-700

Hokmalipour, S. (2011). The Study of Phyllochron and Leaf Appearance Rate in Three Cultivar of Maize (Zea mays L.) At Nitrogen Fertilizer Levels. *World Applied Sciences Journal, 12*, 850–856

Jamieson, P., Brooking, I., Zyskowski, R., & Munro, C. (2008). The vexatious problem of the variation of the phyllochron in wheat. *Field Crops Research, 108*, 163-168

Kang, M., Evers, J.B., Vos, J., & Reffye, P.d. (2008a). The Derivation of Sink Functions of Wheat Organs using the GREENLAB Model. *Annals of Botany, 101*, 1099-1108

Kang, M.Z., Cournede, P.-H., Mathieu, A., Letort, V., Qi, R., & Zhan, Z.G. (2008b). A Functional-Structural Plant Model-Theory and Applications in Agronomy. In, *International Symposium on Crop Modeling and Decision Support: ISCMDS 2008*

Kerstetter, R., & Poethig, R. (1998). The specification of leaf identity during shoot development. *Annual review of cell and developmental biology, 14*, 373-398

Kimes, D., Gastellu-Etchegorry, J., & Esteve, P. (2002). Recovery of forest canopy characteristics through inversion of a complex 3D model. *Remote Sensing of Environment, 79*, 320-328

Kimes, D., Knyazikhin, Y., Privette, J., Abuelgasim, A., & Gao, F. (2000). Inversion methods for physically‐based models. *Remote Sensing Reviews, 18*, 381-439

Kirby, E.J.M., Appleyard, M., & Fellowes, G. (1985). Leaf emergence and tillering in barley and wheat. *Agronomie, 5*, 193–200

Ledent, J.F., & Moss, D.N. (1977). Spatial orientation of wheat leaves. *Crop Science, 17*, 873–879

Li, L., Zhang, Q., & Huang, D. (2014). A Review of Imaging Techniques for Plant Phenotyping. *Sensors, 14*, 20078-20111

Li, W., Weiss, M., Waldner, F., Defourny, P., Demarez, V., Morin, D., Hagolle, O., & Baret, F. (2015). A Generic Algorithm to Estimate LAI, FAPAR and FCOVER Variables from SPOT4_HRVIR and Landsat Sensors: Evaluation of the Consistency and Comparison with Ground Measurements. *Remote Sensing, 7*, 15494

Lock, A. (2003). The origin and significance of an indent on wheat leaves. *The Journal of Agricultural Science, 141*, 179-190

Ludwig, F., & Asseng, S. (2010). Potential benefits of early vigor and changes in phenology in wheat to adapt to warmer and drier climates. *Agricultural Systems, 103*, 127-136

Luquet, D., Rebolledo, M.C., & Soulié, J.-C. (2012). Functional-structural plant modeling to support complex trait phenotyping: case of rice early vigour and drought tolerance using ecomeristem model. In, *Plant Growth Modeling, Simulation, Visualization and Applications (PMA), 2012 IEEE Fourth International Symposium on* (pp. 270-277): IEEE

Maddonni, G.A., Chelle, M., Drouet, J.-L., & Andrieu, B. (2001). Light interception of contrasting azimuth canopies under square and rectangular plant spatial distributions: simulations and crop measurements. *Field Crops Research, 70*, 1-13

Martre, P., Quilot-Turion, B., Luquet, D., Memmah, M.-M.O.-S., Chenu, K., & Debaeke, P. (2015). Model-assisted phenotyping and ideotype design. *Crop Physiology* (pp. 349-373): Elsevier

Masle, J. (1985a). Competition Among Tillers in Winter Wheat: Consequences for Growth and Development of the Crop. In W. Day, & R.K. Atkin (Eds.), *Wheat Growth and Modelling* (pp. 33-54). Boston, MA: Springer US

Masle, J. (1985b). Competition among tillers in winter wheat: consequences for growth and development of the crop. *Life Sciences (series A). 86*, 33-55

Masters, T. (1993). *Practical neural network recipes in C++*. Morgan Kaufmann

Miralles, D.J., & Richards, R.A. (2000). Responses of Leaf and Tiller Emergence and Primordium Initiation in Wheat and Barley to Interchanged Photoperiod. *Annals of Botany, 85*, 655-663

Monneveux, P., Jing, R., & Misra, S.C. (2012). Phenotyping for drought adaptation in wheat using physiological traits

Moulin, S., Bondeau, A., & Delecolle, R. (1998). Combining agricultural crop models and satellite observations: from field to regional scales. *International Journal of Remote Sensing, 19*, 1021-1036

Nagelmüller, S., Kirchgessner, N., Yates, S., Hiltpold, M., & Walter, A. (2016). Leaf Length Tracker: a novel approach to analyse leaf elongation close to the thermal limit of growth in the field. *Journal of experimental botany*, erw003

Paproki, A., Sirault, X., Berry, S., Furbank, R., & Fripp, J. (2012). A novel mesh processing based technique for 3D plant analysis. *BMC Plant Biology, 12*, 63

Pararajasingham, S., & Hunt, L. (1996). Effects of photoperiod on leaf appearance rate and leaf dimensions in winter and spring wheats. *Canadian journal of plant science, 76*, 43-50

Prévot, L., Chauki, H., Troufleau, D., Weiss, M., Baret, F., & Brisson, N. (2003). Assimilating optical and radar data into the STICS crop model for wheat. *Agronomie, 23*, 297-303

Reuzeau, C., Pen, J., Frankard, V., de Wolf, J., Peerbolte, R., Broekaert, W., & van Camp, W. (2010). TraitMill: a discovery engine for identifying yield-enhancement genes in cereals. *Plant Gene and Trait, 1*

Schirrmann, M., Giebel, A., Gleiniger, F., Pflanz, M., Lentschke, J., & Dammer, K.-H. (2016). Monitoring Agronomic Parameters of Winter Wheat Crops with Low-Cost UAV Imagery. *Remote Sensing, 8*, 706

Verger, A., Vigneau, N., Chéron, C., Gilliot, J.-M., Comar, A., & Baret, F. (2014). Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sensing of Environment*

Vos, J., Evers, J.B., Buck-Sorlin, G.H., Andrieu, B., Chelle, M., & Visser, P.H.B.d. (2010). Functional–structural plant modelling: a new versatile tool in crop science. *Journal of Experimental Botany, 61*, 2101-2115

Weiss, M., Baret, F., Myneni, R., Pragnère, A., & Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie, 20*, 3-22

Zhang, L., Guo, C.L., Zhao, L.Y., Zhu, Y., Cao, W.X., Tian, Y.C., Cheng, T., & Wang, X. (2016). Estimating wheat yield by integrating the WheatGrow and PROSAIL models. *Field Crops Research, 192*, 55-66

Conclusions and perspectives

After explosion of investment in the last decade, phenotyping infrastructures are more easily accessible from the several possible platforms available including aerial, field or greenhouse ones. Further thanks to the academic effort often supported by a strong interest of private plant breeding companies or structures marketing the phenotyping tools, local and international phenotyping networks have been growing facilitating the sharing of information as well as the uptake and adoption of the new developments. As my experience in Europe and China, the concept of phenotyping has been accepted by wide range of domains, including breeding, agronomy and plant modeling, driven by their common requirement to better describe canopy structure and functions. The technical progress of sensors, computers and robotic systems still obeying the Moore's law will reduce the cost of the platform while increase the quality and throughput of the data acquisition.

The appealing pictures and videos collected by the fancy integrated complex phenotyping systems greatly contribute to the current prosperity of phenotyping. However, what is really achieved from these platforms is more questionable. The sustainable development or even the survival of phenotyping relies on whether it can satisfy the actual needs of his users before they will be too disappointed by the actual limited performances for the associated high cost required. Insufficient attention has been invested in the interpretation of phenotyping acquisitions. The majority of phenotyping techniques were mostly inheriting from the relatively mature remote sensing domain where efforts have been spent on the interpretation of satellite and airborne imagery. The corresponding methodologies have been only partly transferred to phenotyping applications to access canopy traits from imaging observations. However, the improved spatial resolution from kilometer or meter to centimeter or even fraction of millimeter needs specific techniques to extract more traits with increased accuracy. Even though the leap between scales is significant, the principles of the remote sensing methods based on physical principles can be further adapted to interpret phenotyping observations. The interpretation of phenotyping observations has not progressed as fast as the development of the sensing techniques. The phenotyping bottleneck of genetic progress recognized few years ago by the plant breeding community (Furbank and Tester 2011)[7] is itself limited now by the data interpretation bottleneck. The objective of this study was therefore to contribute to advance the interpretation techniques of the data provided by the new phenotyping sensors and platforms. I investigated three complementary aspects that illustrate the potentials of advanced image processing, model inversion for the interpretation of instantaneous measurements and data assimilation for multi-temporal measurements to provide new traits or improve the accuracy with which already accessible traits have been retrieved. Several platforms (phenotypette, phenomobile, UAV) and sensors (RGB high resolution cameras, LiDAR) were used along this study. A synthesis of the main findings is provided below:

- ***Plant density and sowing pattern estimation from high resolution RGB imagery.*** Plant density along with the distribution pattern of individual plants on the row is required to describe the competition between neighboring plants as well as allow initializing the simulations of the 3D ADEL-Wheat model. Using high spatial resolution ground-based RGB imagery, computer vision methods were applied to automate the density survey for wheat in the early stage. Accuracy and reliability of our method was

[7] Furbank, R.T., & Tester, M. (2011). Phenomics - technologies to relieve the phenotyping bottleneck. TRENDS in Plant Science, 16, 635-644

validated over a wide range of conditions for the density estimation. The method initially developed from ground level phenotypette images was adapted to the UAV platform to increase the throughput. It can be also easily transposed to the phenomobile platform. The plant distribution pattern on the row was concurrently described by a statistical model that requires six parameters. It takes into account both non-uniformity of inter-plant distance along the row direction and the deviation across the row direction. Simulations completed with the ADEL-Wheat model showed that the effects of the plant distribution pattern on the canopy GF were generally small. The statistical model developed was also exploited to investigate the optimal sampling strategy required to each a given accuracy when evaluating the plant density. Results demonstrate that measuring the length of segments containing 90 plants will achieve a precision better than 10%, independently from the plant density and the variability of plant spacing.

- ***Estimating GAI from ground-based LiDAR measurements using ADEL-Wheat model.*** GAI dynamics along the growth cycle is closely linked with the canopy functioning, thereby serving as an important trait of interest for breeders, crop modelling and decision making for field management. It is relatively easy to achieve accurate GAI estimate using passive observations at early stages. However, the performances degrade for high GAI conditions due to the saturation problem. LiDAR with its capacity to bring information on the third dimension provides a possible way to alleviate the saturation effect based on the regularities between top and deeper canopy layers. Focus was put on the stage of maximum GAI development when saturation effects are the largest. A synthetic learning approach was proposed to achieve better GAI estimate using LiDAR observations and the ADEL-Wheat model. A LiDAR simulator was first developed and validated to mimic the properties of field LiDAR observations from the outputs of the ADEL-Wheat model. Our synthetic learning approach was evaluated using both an independent synthetic dataset as well as actual field measurements completed with the phenomobile platform. Results showed that our method can achieve accuracy two times better than the classical turbid medium approach (Poisson model) with average rRMSE 10%.

- ***Extraction of traits from the assimilation of multi-temporal GF measurements into ADEL-Wheat during the early stages.*** Monitoring the dynamics of canopy architecture during the early stages is highly desirable for breeders to select cultivars of high drought-tolerance, resistance to weed infestation, improved water-use efficiency and potentially high grain yield. These desired traits may be grouped into early vigor trait. A dedicated in silico study was completed to evaluate the feasibility and interest of a phenotyping data assimilation approach. A sequential assimilation scheme was developed to limit the space of canopy realization since the simulations of the 3D canopy architecture as well as the rendering of the phenotyping measurements is computer demanding. The early growth was decomposed into 2 consecutive stages. For each stage, 5 main influential parameters were identified. The phenotyping measurements were consisting into RGB images taken at several dates along the growth cycle and from several view inclination angles. For the first stage, the 5 influential parameters were estimated using a neural network machine learning technique. For the second phase, the 5 additional influential parameters were adjusted using the same neural network based technique while exploiting as prior information the posterior values of the 5 parameters defined during the first phase assimilation process. Results show that the parameters defining the size and rate of change of the green canopy crop elements were estimated with a relatively good accuracy.

Conversely, the parameters driving the orientation and inclination of the elements were retrieved with a poor accuracy. However, some important emerging properties were very well retrieved, including the GAI and the number of tillers with more than 3 leaves at the beginning of stem elongation. Further detailed sensitivity analysis highlights the importance of getting enough dates of measurements covering the growth cycle, while dates separated by less than a phyllochron do not provide significant additional information. The sensitivity analysis highlighted also the importance of the inclined observations. The assimilation approach proposed appears very general and could be applied to a number of phenotyping measurements completed along the growth cycle. Nevertheless, because our results were achieved using in silico experiments, further validation is necessary using actual field measurements.

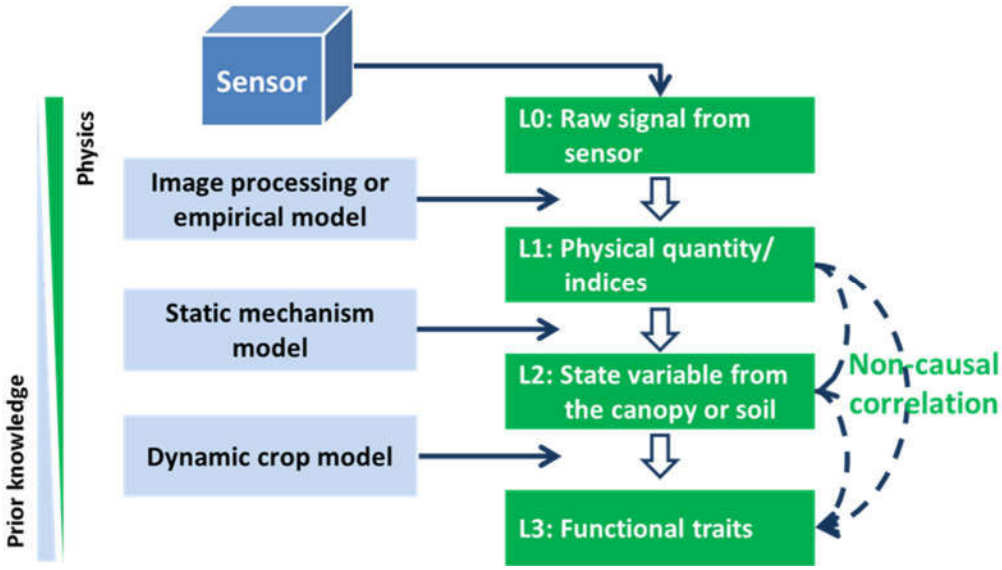

Fig. 1 Interpretation methods of phenotyping observations

The work completed in this study proposes a framework for interpreting phenotyping observations. Four levels of data interpretation could be identified (**Fig. 1**).

0. ***The level 0 corresponds to the raw signal*** recorded by the sensors. This is generally a digital count in the image or for the sensor output.
1. ***The level 1 corresponds to the calibrated data and the associated meta-information*** (time, coordinates, quality assessment of raw data, sensor characteristics, configuration measurement …). Level 1 results from the application of preprocessing algorithms to calibrate the signal into a physical quantity, or into a normalized quantity. It may be a white balance corrected RGB image corrected from distortion and vignetting effects, radiance or reflectance eventually combined into a vegetation index value such as NDVI. It could also correspond to specific features extracted from the image which encapsulates most of the spatial information contained in the image.
2. ***The level 2 corresponds to a canopy state variable*** such as plant density, GAI, green area density …. They are derived from the Level 1 using specific image processing algorithms, machine learning methods based on experimental data sets or using physical approaches such as radiative transfer modeling. The chapters 2 and 3 of this work correspond to the development of Level 2 algorithms and corresponding traits.
3. ***The level 3 corresponds to functional traits***. These are the traits that are desired by the breeders because of an expected higher heritability since the effects of the environmental factors are already taken into account through the processes described by the FSPM model. The retrieval of these traits

requires combining multi-temporal observations within a dynamic model that simulates the Level 2 variables. The retrieved model parameters are thus considered as Level 3 data, i.e. functional traits. In our work, the ADEL-Wheat parameters estimated using the assimilation approach corresponds to such functional traits. In addition to the model parameters retrieved, emerging properties could be also computed. Depending on their dependency on the environmental factors, they could either be considered as Level 2 or Level 3 data.

Note that empirical relationships are very frequently proposed in the literature between the several data levels. Such empirical relationships, if non-causal (**Fig. 1**) may lack robustness and may lead to misinterpret the phenotyping measurements. Conversely, the progression to higher levels proposed in our approach is associated with the integration of ancillary information and knowledge that corresponds also to an increase in the complexity of the data processing. Although it should provide causal relationships, hence robustness, the approach relies on a number of assumptions made through the models. As an illustration of the current limits of the model used, the dynamics and plasticity of leaves is not always well described, this may impact the realism of the canopy. More importantly, most current FSPMs focus more on the description of canopy structure than on the physiological processes. This limits the capability to access physiological traits of interest. However, the phenotyping experiments will thus provide great opportunities to calibrate, validate and eventually question the formalism of the models used which may ultimately beneficiate the interpretation methods and the associated applications.