

NPPN

The challenge of robust trait estimates in plant phenotyping with Machine Learning

DATE

AUTHOR(S)

February 2022 Etienne

Etienne David



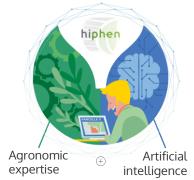


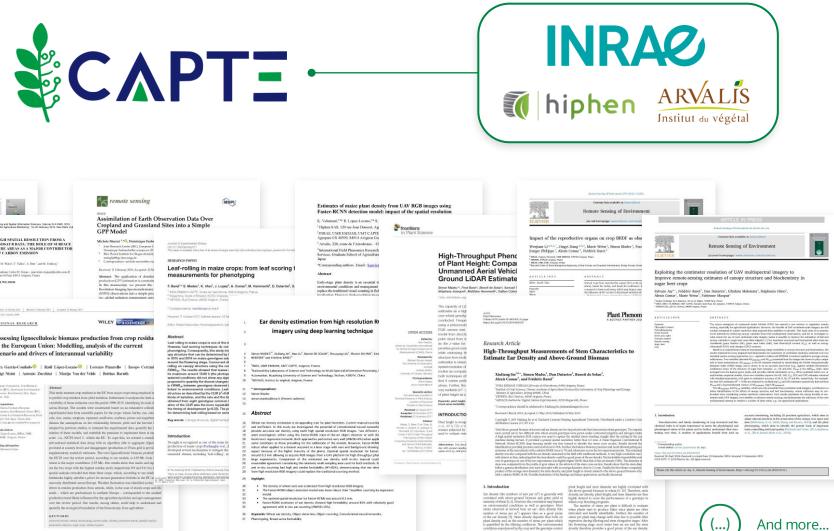
Our 20+ agronomists and data scientists are expert at turning crop images into valuable agronomic traits



hiphen

Our remote sensing expertise arises from agronomic knowledge combined with artificial intelligence

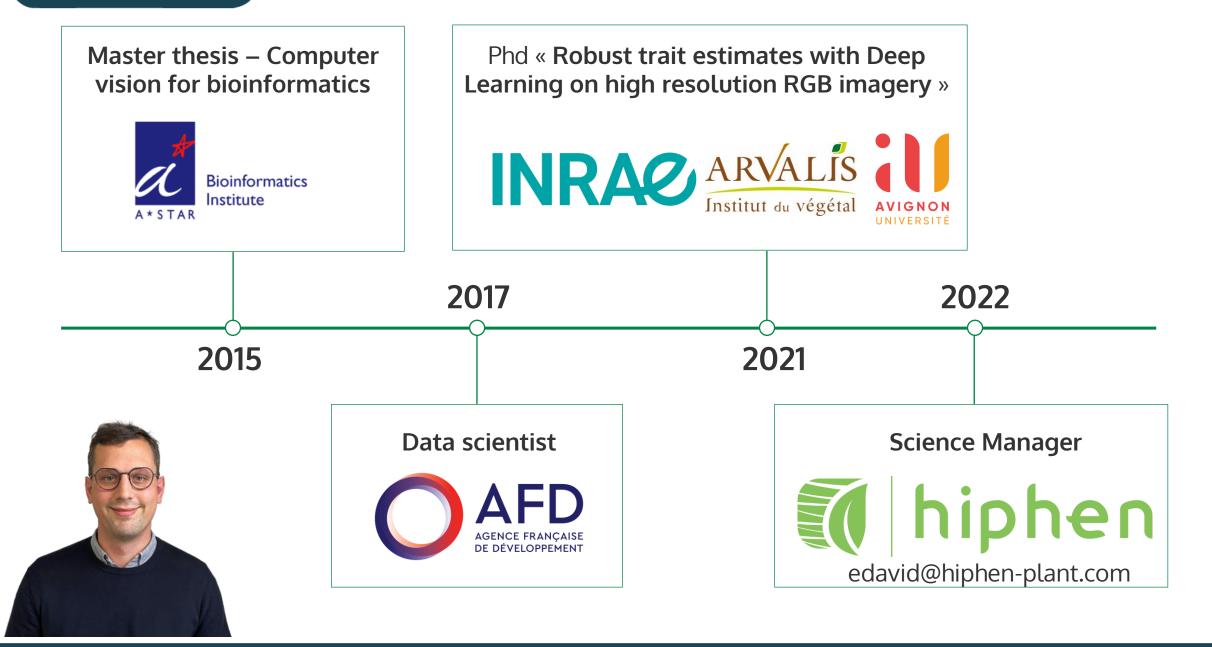




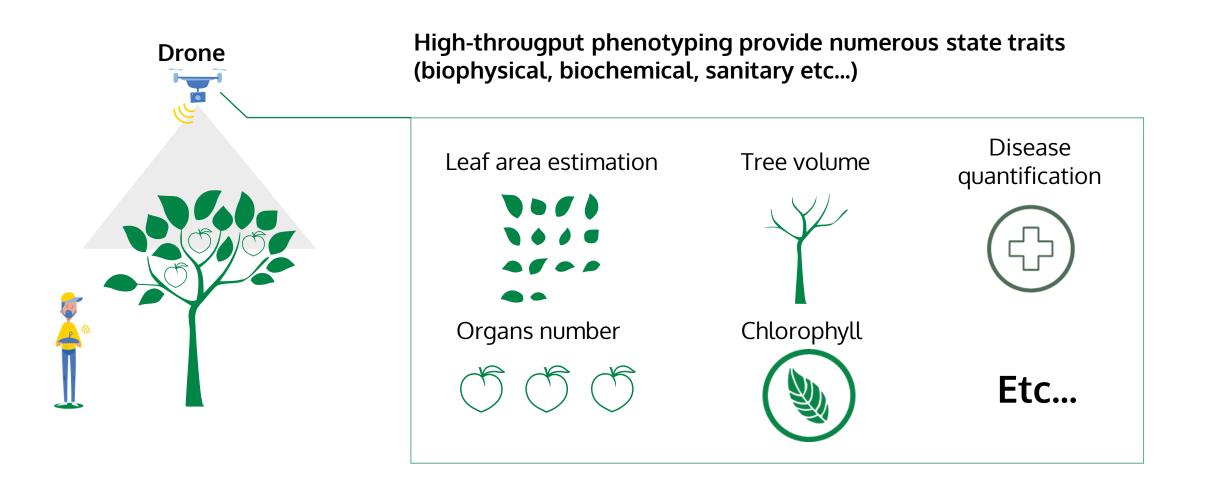
				Services.	Field Crops Research	and the second se		Antick		
Congress of Encoders & Aprilate 170 (2011) 10-138			- E-INEE				Assimilation of Earth Observation Data Over			
			The International Archives of the Photogenerativy ESPRE-GEOCLAMPISCS Jam Int. Workshop on "Sett		Remote Sensing and Spatial Information Sciences, Volume XLI-5065, 2019 Observations for Agricultural Monitoring", 19–20 February 2019, New Debi, Ind		Cropland and Grassland Sites into a Simple GPP Model			
eutorionica	Cont	ents lists available at ScienceDived		An automatic method base		T AT HIGH SPATIAL RESOLUTI		Michele Meroni ^{1, 4} 0, Dominique Fasbe		
1000				wheat heading stage	COMBINATION OF SENTINEL-2 A			1 Joint Research Centre (IRC), European C	Journal of Experimental Bolany (a) 1020/adulary071	
Computers and Electronics in Agriculture Kaaviyo		Kaaviya Velumani ^{s, har} , Simon Ma		RADIATIVE FORCING FROM AGRICULTURE AREAS AS A MAJOR CON TO AN ABATEMENT OF CARBON EMISSION		Dominique Fashenderfles europa eu (D. ¹ Mas Parisk Institute for Bogescheniste	This paper is available defew has of all access drampes have http://jib.com/			
ELSEVIER	irmal homap	inge: www.elsevier.com/tec	(ate/compag	Jeremy Labrosse ² , Stephane Jezer				nnigliøftig-jero opg de * Consepondence michele metoniftecau	BESEARCH PAPER	
				¹ MANE, Angeon Université, UNIV 2000/04, UNIV CAPTE "Avails, 220, roam il: Patricipcine - CE 40500, P4019 A	JL. Roginan ¹⁷ , A. Okoso ² , E. Coschar, I	 Hagolie¹, M. Weise¹, T. Talloc¹, A. Brat¹, and 	M. Ferliceq			
Original papers					"CESERO UMR 5126, 18 avenue E. Belia, bpi 290	t, 31055 Toulouse Cedex 09, Prame - jaon Jours.	vajoanižentio cnis ž	Received: 11 February 2019; Accepted: 22 M	Leaf-rolling in maize cro	
Modeling the spatial dis				ARTICLE IN FO		BA Domaine St Paul, 84914 Avignon, France		Abstract: The application of detailed production (GPP) estimation is constrained	measurements for phen	otypin
Consequences on the gr		15		Kopwordi: Decelopy	Ca	andssion III, WG III/10		In this manuscript, we present the	F. Baret1+ S. Madec1, K. Irfan1, J. Lopez1, A.	Comar ² , M.
Shouyang Liu ^{4,*} , Frédéric Bare Philippe Burger ^o	t", Bruno A	ndrieu ^b , Mariem Abichou ^b	¹ , Denis Allard ^e , Benoit de S	Invenue of Things for Agriculture Generalizing a neural supervise	KEY WORDS: Sentral-2, albeix agreeature, carbon			Resolution Imaging Spectroradiometer (NDWI) observations into a simple pro-	¹ NRA-EMMAN-CAPTE, Route de l'aerochome, 649	A14 Avignon, Fm
1988 DRIME ONE LIVE Downey Same Park Se	-Apapent Mille	Arignes Gales S. France		Paid search Physical endoting	ABSTRACT:			(i.e., global radiation, temperature, prec	² Bagerrine, Floute d'Emech, 83730 Chapten, ^{****} ⁴ HEHEN, Rue Chartes, 84000 Autgron, France	
¹ Mills AgroParticlus, UAB 2001 202, 78859 Plane 1986 (http://doi.org/10.1016/j.compress Same Peak, Sine A 402027, Intellige Aground Aground Agroup and Agroup	nol-Origans ****				Satellite Sentend-2 offers a global coverage of the					
*ABORES-Institut duringhist, Sontim orphytochial *1580, 0588 7248 AGB, Chewin & Birde Rouge, B	E 98720 848 P 12627, 381				meters. Such spatio-temporal resolution fisters an mean to reduce the surface sudiative foning in list	Docted 10 October 2010 Restand 10 hours	2001 Anoped Wheney 2017		* Consepondance: tensf@avignon.tvia.tr	
Anna ann an Anna Anna			Founds Souling of Devicement	26×Gelli 11289	new generated continely from the observations of methodology inherited from the Global Land Sore	TOP 301117/junctions			Received 11 October 2017, Editoral decision 10 Feb	
ARTICLE INFO	A D S	STATES SHA	Comesn lies evaluate a	(Selected and Contemport		ORIGINAL RESEARCH		WILEY BIOENERGY	Editor: Poland Panactics, Ponchangiaethan Jak	1
Received 16 May 2016 Received in revised form 16 February 2017	cultina cultina	1000	Remote Sensing of	Environment	L INTRODUCTION					2
Accepted 30 Petruary 2017 Available on line 22 March 2017	Besult	ELSEVIER	ipartal tomopage www.elia	the second s	The advent of fleet such like the satellites Sentire permits the Earth observation research to enter a	Assessing lignocell	ilosic biomass pi	roduction from crop residu-	Abstract	
Arywools Place distribution patients	distrift hie ph	ELSEVIEW.	partie readings.		owing to a frequent revisit of the globe at approp- spatial resolution. Besides, the level of quality measured from lessons learned properties at one	in the European U	tion: Modelling.	analysis of the current	Leaf rolling in maize crops is one of the m Homeser, leaf, scoring techniques do not	3
Plant distribution patient Creek Station 1955	icepact carb a				presented Exclusions instead promition, modely in the BR (High Resolution) information, modely in the European Commission's Committee memory II	scenario and drive			phenotyping. Consequently, this study ine opy structure that can be determined by I	4.4
Wasi	the off openes			counting for the differences	anyrest favorably to an increasing demand for	Sara García-Condado 0 Raúl López-Lozano		variability	in 2015 and 2016 on maize genotypes sub	5 8
	place p	between upper and	lower faces		mitable environmental data in particular, il airre permanent monitoriane ef far hard territories.			Lorenzo Penerello Iscono Cerroni	around the flowering stage. Concurrent di rolling on canopy structure using the cor	6 7 1
	Novell optimi	Jingyi Jiang ¹⁴ ¹⁴ , Alexis Comar ¹⁴ ¹⁴ , Marie Weiss ¹⁸ , Prédéric Baret ¹⁰ ¹ Joans / Guer Jene Nicasar (Julyen) of Into Kenry on Unabal Ashidemin, Sulperineey Indexis, 1990 Sulp. Class ¹ ADMONS MJ, Dr. Ver Units, 1990 Adjust, Fazer ¹ ADMONS MJ, Dr. ve Units, 1990 Adjust, Fazer			characteristics of Saniard-2 sensor allow a biophysical variables describing the vegetation o energy bodget and the water syste over the entre generic product one serve matteries applicat	Luigi Nisini Antonio Zucchini Marijn Van o			FIFMR ₆₀ . The results showed that leaves i its maximum around 15:00 h (the photops	8 1
								conte Detring Datasi	watered conditions did not show any sign	9 A 10 T
	-				agriculture and fixed scenary, weather fencast, if impact studies, water, ferest and naronal mesures	Daypon Consisten, Just Resert	1.00000000		proposed to quantify the durnal changes i in FIFAR _{in} between genotypes observed i	10
1. Introduction		ARTICLEINFO	ABSTRACT		To be collined that the Sentinol-2 mission of Franch SPOT and US Landot missions Par-	Ornier (BC), Darwitsens for Batalandar Resoaner, Pool Jonatty Unit, Japan	Abstract This study assesses crop residues in the EU from major crops using empiri-		linked to environmental conditions, Leaf structure as described by the CL/R (r ² =0.8)	11 -
The plant spacial distribution is divisionly lists ing. The seed diffi is putel by the transmit of the pla sombifunding of goars synchronized with the the diff. The variation of goars synchronized with the seed metricing, produce, Ende throm the mechanis or generative metric of this space. The throm the mechanis		Indirections is and in the second sec		that properties between apper and lower heat lates. Here	deseptint of 30 metars from Landau-8 is complia- and 30 metars pool resolution of Santiad-2, areagefue boh, spatial and sampain overage enfluence product in a spatial domain overain unfaced and minimized offers new childring culture, and the statistical offers new childring culture, and the statistical offers new childring of culture, and the statistical offers are statistical offers	Verse, Buly Correspondence Base Oersto Constacts, European Commission, Jaine European (BEC).		n yield statistics; furthermore it analyses the inter-a	litude of variation, and the rate and the tin	12 5
				ef to the non-lanar-generate distribution of absorbing and scatt articular surface forkarss of both opidemisis. We proposed the I OPINCT model to describe the differences in artifectance and to			variability of those estimates over the period 1998-2015, identifying its main d across Europe. The models were constructed based on an exhaustive collecti experimental data fraza scientific papers for the crept: wheat, backy, eye, one		obtained from eight genotypes common I ation of the CLIR was the more repeatabl	14
			31.7 model to docume the differences in articulates and to see epidemic layers are characterized by distant waveleng de of a pulsade and a spongy participants layers using t	the timing of development (p=0.33). The pr for determining leaf-rolling based on seria					15	
face and soil covering and pressing (will of the seed metering system can cause	isan, 19880	the double the deriver		general and load arraction between these two layers. As or	outcomes of the dissentination of quality checks will certainly benefit to programs blor GEOGLA	Desettanie for Danialackie Ecourose, Pool Society Urit, Jugas, Viewin, Jidy	cale, rice, maine, sorghum, mpesced, sunflower, soybean, pointo and segarber		Key words: Canopy structure, digital herriept	15 .
place to place (Karayel et al., 2006). I may contribute to the pon-uniformity	m activities			al properties between the two faces. The specific absorption on smallbashed for the NANPOCT model over a definated da	Earth Observation for Global Agricultural Mi- which main scenarios are the obset and decay-	Bual man perto-condido #+c empo es		a the relationship between yield and the harvest i	Key words. Carriey southast, repair to rep	16 ¥ 17 a
the row and deviations from the row di- of the state of the soil surface, change	ection inc		model this also accounts for the d	of compared with DECEPTICE-5 and DECEPTICE-0 as well as ifferences between the two faces through down that the P and transmittingers of the two faces has participating the	only warring. The presentation will highlight fit methodology to be implemented in order is	Present address East Liper-Leners, DRA, UME		is, to interpret the experimental data, quantify the u of establish the premises to implement them at rep		18 #
vibrations or deviations of the drill from	the pre-d				management of the HE surface allocks and a training mentioning of the worldwisk corps.	Exhibit Avgros. Fasts		within the EU. To cope this, we created a consoli	Introduction	19 v 20 h
tion. Further, the graminating process	s driven i		interior load backward is chosen if	simulations as compared to OLH. The capacity of the FAS on from orthoclastic and transmittance microarcometer tests in the estimation of elibertybyF, carotimotids, and water-conte		Fundag information		a along with an algorithm able to aggregate (figure	Drought is recognized as one of the main fac production of maize crops (Fathanglie et al., 2	21 5 22 #
* Enropending aution FreeToddeet: Television Laternal (10, Lat.)		significant improvement intervent and impre-of d				Roled Bostnerill Centure		and disaggregate (production at 25 km grid is provid timates. The total lignscellalosic biomoss product	developed several mechanisms to mitigate the	23 3
The Obstantio West owner WITAT ST								period, according to our models, is 419 Mt. from	onmental stenses, including 'leaf-etilling', in	24 a 25 b
0165-3056/c 2017 Published by Elsever &V.		1. Introduction	(pr	ties show differences in acatoring properties between				atur (155 Mt). Our results show that make and rap		26 #
		The monitoring of regulation by remote sorating inclusions at 1971) larger encode series		ddini at al., 1907; Grant, 1907; Shedoen at al., 300 (1). Leaves cannot several functions cashe a wide care	Ntps://bis.org/10.5154/sprs-on			highest residue yield, respectively 8.9 and 8.6 t ha-1	© The Autor() 2018. Published by Colord University Pres	27 a 28 1
		sharen Ettebaren, 2003: West	err al., 2010. Vecetation authorize a c	mental conditions and have therefore developed adaptate a differentiation between lines. Washers et al. (200%)			openancy descines stron Europe, weather intrastion was intrined in the		This is Open Access while dutits and under the terms any visit ungoth unustablished annual fail ballanus optimized	29 A
		tions and tomaform into structure	ind or biochemical regetation chirac- her	a hilacial na dominantal lawawa with constanting optical pr www.huth-faces indexs a significant effect on smergy reflec	basio.				profiliation forgenet in the following Aground pair and	10
		whith the signal collicited by millimeters in the savellar, drone, or assessi levels. Modellar the reductive symplex at the campe level so- mittace from a limited set of stars variables describtant the o			then of		driver in residue production from cereats, while, in the case of starch crops and scals - which are predominant in sorthern Europe - corresponded to the more			31
		quites a description of find system properties (Cont and Chan, 2017). adsorbing materials such as charagingil, water, or day matter, locates are often represented as Lamberian surfaces with the same staticting constraints at the interfaces between materials with d			Different		sends - which are predominant in northern hampe - corresponded to the mer production trend likely influenced by the agricultural policies and agro-managem			32 33
		reflectance and transmittance pro-	peties for both faces. However, many inf	raction index values. Several modeling approaches has	e lees		over the review period. Our results, among others, could help to unders			3.4
		* Compared as write or insert	h Gray of Forst Management Build of the	Ferentry and Granitani Administration, Beijing Ferentry University	00000		quantify the scological boundaries of the bioeconomy from spicalture.			35
		Riging China American (1997)			a na seconda de la constante de		KEYWORDS			36 A 37 P
		In the constant of the local sector of the local	COMPANY OF THE OWNER				advaced/aphela, reenils, Arrest accesses, addition, maps carps	oding, harvest index, siloveli, production tyrols, spitially explicit vessilies incontin		
		Revenued 13 March 2020; Received 5 Acadable online 6 December 2020 0034 4282/02 2020 Elsevier Sec. Allo	is noised form 4 theoler 2020; Accepted 27 New miles around	anw arun			a construction of the state			
		and here to any mention of the				The loss care according to be well by the terms of the	Destive Determine //Albeites Linear, wh	it preside and distribution and opposite that is any outday, provided the original		
						web a popoly met. 6 200 Ten Artico 002 Bowerg Pricides' in Net Wire & Soc Dd				
						OCH Browey, 200111-000-001.				



Who am I?

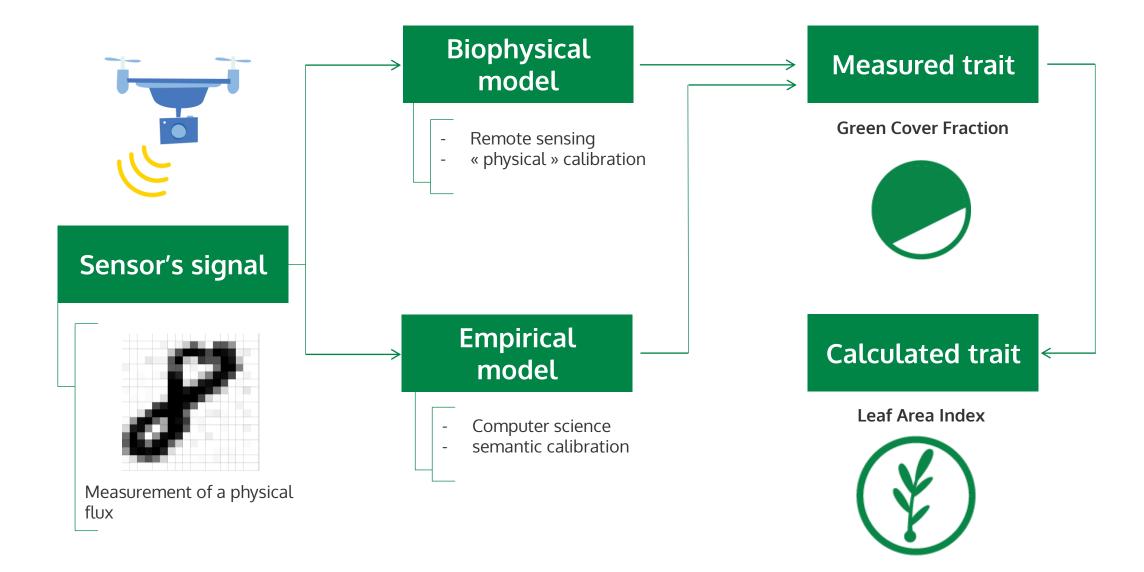


State trait estimates is the basis phenotyping



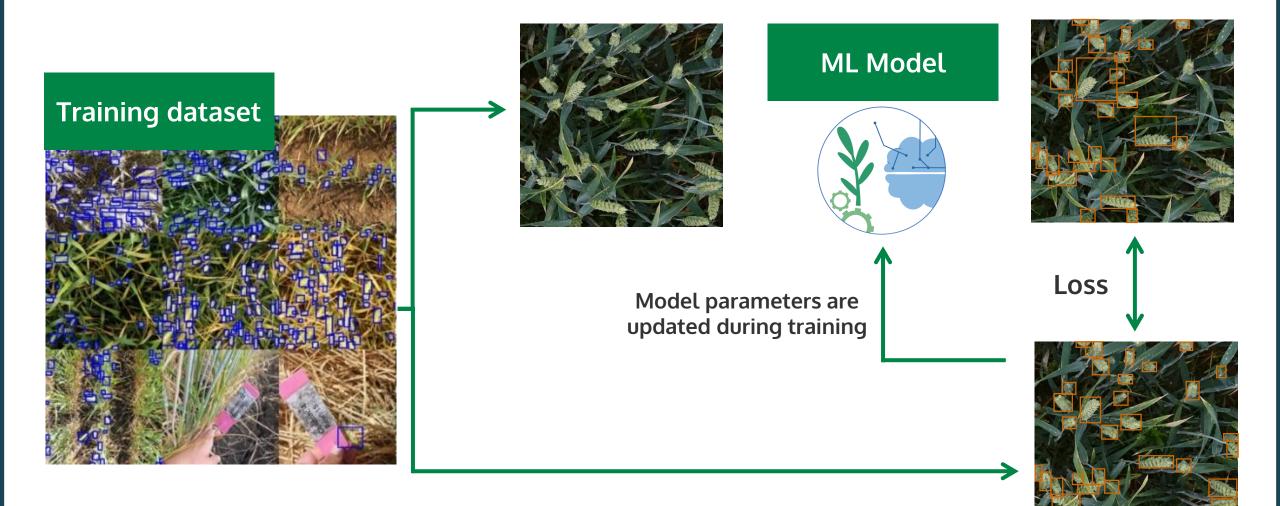


Two main algorithm's families can be used



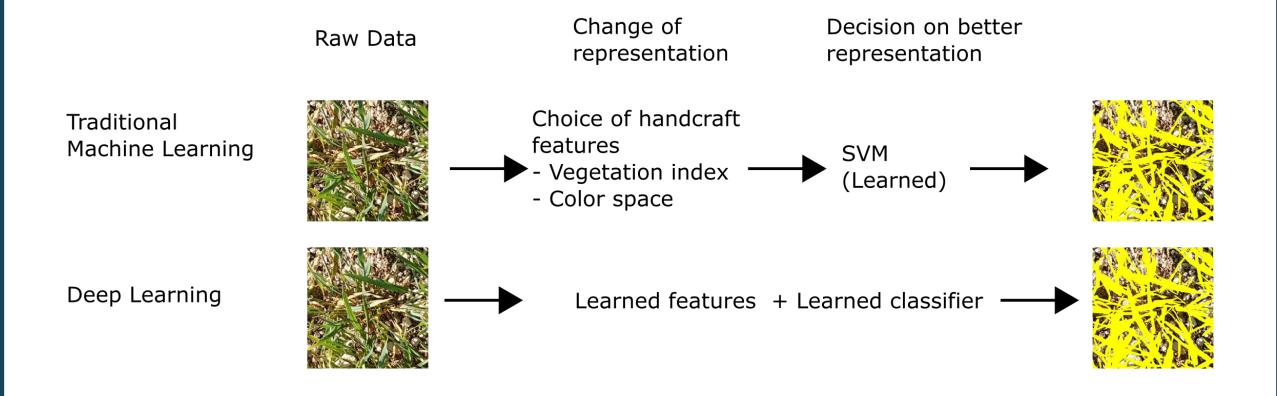
i hiphen

ML is a powerful family of empirical model





Deep Learning models learn representation





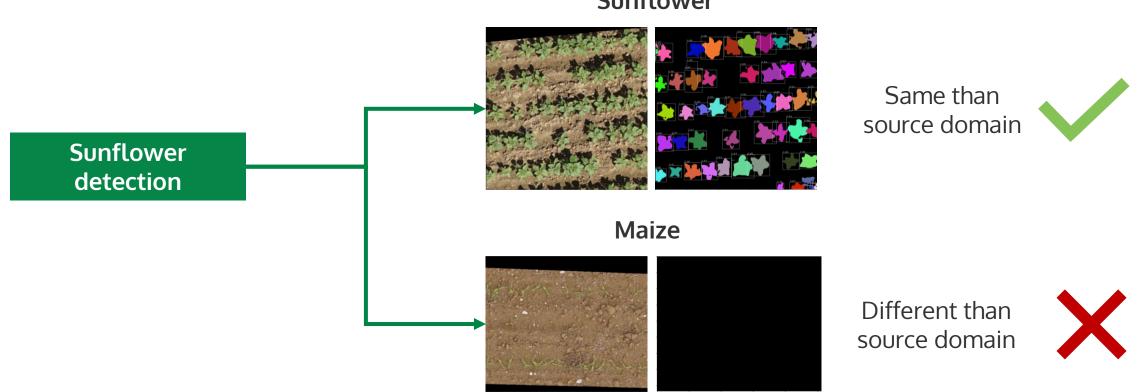
Deep Learning allows to access new traits

Trait	Piophysical	Empirical				
ITall	Biophysical	Machine Learning	Deep Learning			
Green Cover Fraction	\checkmark	\checkmark	\checkmark			
Leaf Area Index	\checkmark	\checkmark	\checkmark			
Height	\checkmark		\checkmark			
Lodging Score	\checkmark		\checkmark			
Leaf Chlorophyl Content	\checkmark	\checkmark	\checkmark			
Canopy Chlorophyll Content	\checkmark	\checkmark	\checkmark			
Plants Density		\checkmark	\checkmark			
Crop Cover Fraction			\checkmark			
Senescent Fraction			\checkmark			
Head Density			\checkmark			
Disease Fraction			\checkmark			



The domain shift problem

A **domain shift** is a change in the **data distribution** between an algorithm's training dataset and the encountered images when deployed.

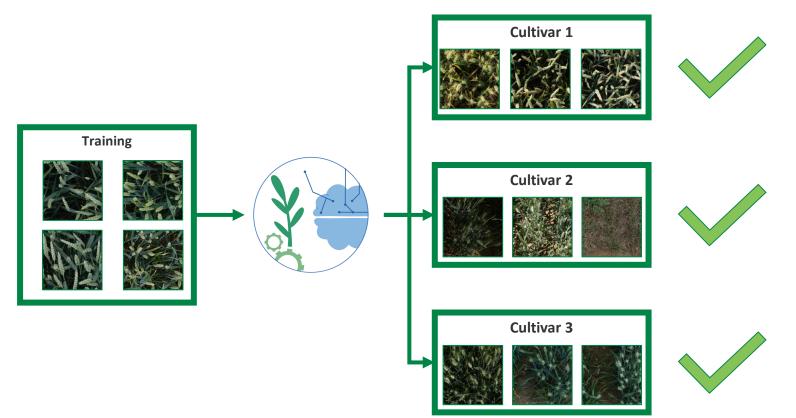


Sunflower



Does it cause a robustness problem ?

Robustness in plant phenotyping is to the capacity of an algorithm to produce an **unbiased** trait estimate for **all images** acquired with the **same** protocol of acquisition.





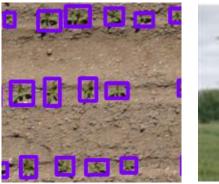
How much does domain shift affects ML algorithms performance in Plant Phenotyping ?

Session of acquisition as a proxy for diversity

Session of acquisition

Vector X Time unit X Experimental unit (ex: 1 Flight over one trial)

UAV plant counting





3-4 leaves	Sony alpha 6000			
Loam soil	Cloudy condition			
Row spacing of 45 cm	GSD of 4.5 mm			



Wheat head localization



Post-Flowering Row spacing of 17,5 cm

Sony alpha 6000

Cloudy condition

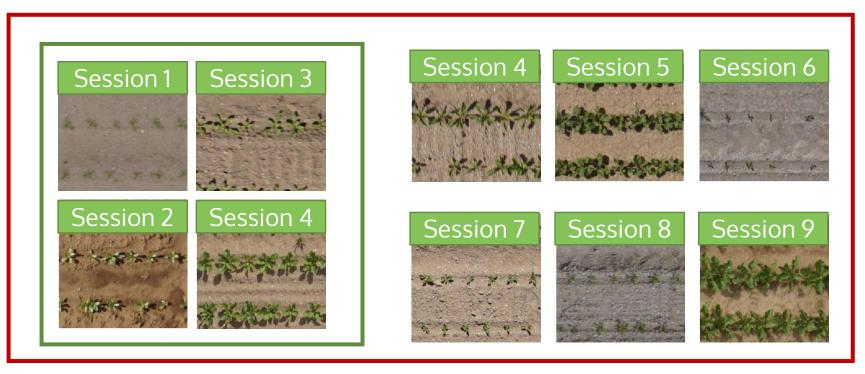
GSD of 4.5 mm

Example on plant counting

Study	UAV	Crop	Sessions	Localization	Test independency
(Quan et al., 2019)		Maize	10	\checkmark	
(Ribera et al., 2017)	\checkmark	Sorghum	2		
(Valente et al., 2020)	\checkmark	Spinach	1		
(Liu et al., 2020)	\checkmark	Maize	2	\checkmark	
(Lin and Guo, 2020)	\checkmark	Sorghum	2	\checkmark	
(Madec et al., 2019)		Wheat	2	\checkmark	\checkmark
(Xiong et al., 2019)	\checkmark	Wheat	10+		\checkmark

The ideal dataset covers all posibilities

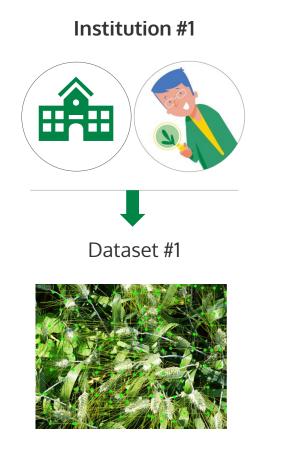
Training diversity = Application diversity



The ideal dataset should include as many sessions of acquisition as possible to cover **expected diversity** !



Sharing the labelling burden

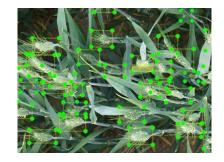


Institution #2





Dataset #2



Institution #3





Dataset #3



An example of large, collaborative dataset





Database/Software Article

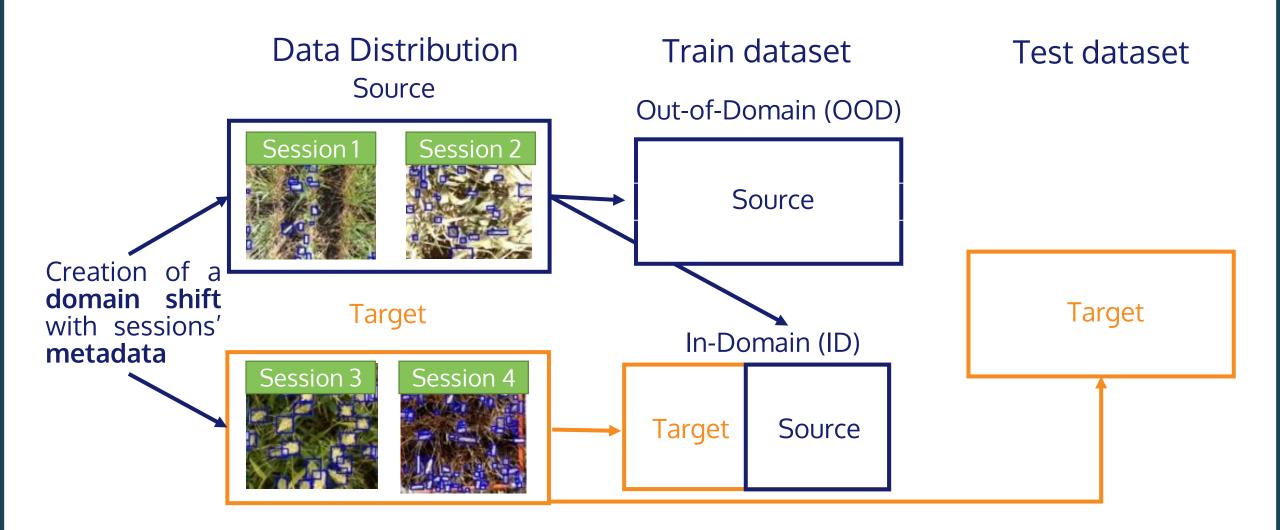
Global Wheat Head Detection 2021: An Improved Dataset for Benchmarking Wheat Head Detection Methods

Etienne David (°,^{1,2} Mario Serouart (°,^{1,2} Daniel Smith (°,³ Simon Madec (°,^{1,3} Kaaviya Velumani (°,^{2,4} Shouyang Liu (°,⁵ Xu Wang (°,⁶ Francisco Pinto (°,⁷ Shahameh Shafiee,⁸ Izzat S. A. Tahir (°,⁹ Hisashi Tsujimoto (°,¹⁰ Shuhei Nasuda (°,¹¹ Bangyou Zheng (°,¹² Norbert Kirchgessner (°,¹³ Helge Aasen (°,¹³ Andreas Hund (°,¹³ Pouria Sadhegi-Tehran (°,¹⁴ Koichi Nagasawa (°,¹⁵ Goro Ishikawa (°,¹⁶ Sébastien Dandrifosse (°,¹⁷ Alexis Carlier (°,¹⁷ Benjamin Dumont (°,¹⁸ Benoit Mercatoris (°,¹⁷ Byron Evers (°,⁶ Ken Kuroki (°,¹⁹ Haozhou Wang (°,¹⁹ Masanori Ishii,¹⁹ Minhajul A. Badhon (°,²⁰ Curtis Pozniak (°,²¹ David Shaner LeBauer (°,²² Morten Lillemo (°,⁸ Jesse Poland (°,⁶ Scott Chapman (°,^{3,12} Benoit de Solan (°,¹ Frédéric Baret (°,² Ian Stavness (°,²⁰ and Wei Guo (°)¹⁹





How to measure the domain shift ?



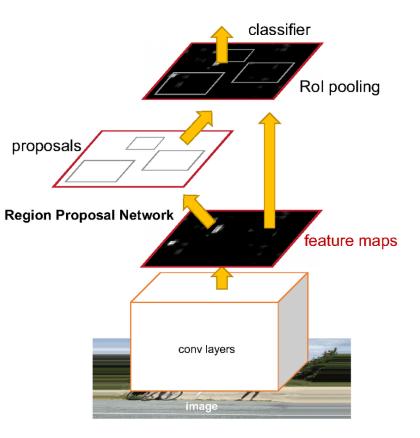


Faster-RCNN to localize and count

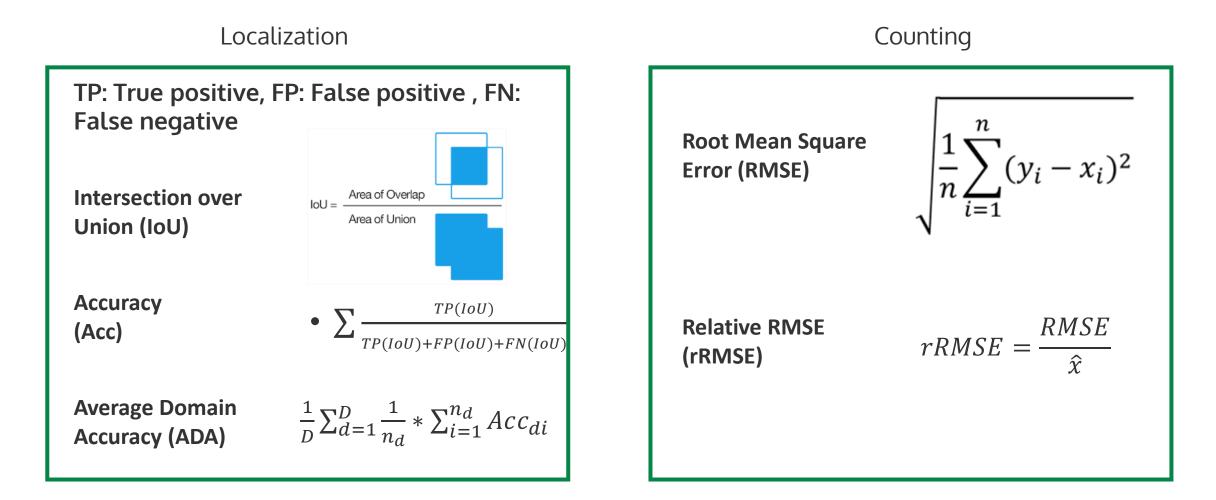
Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun

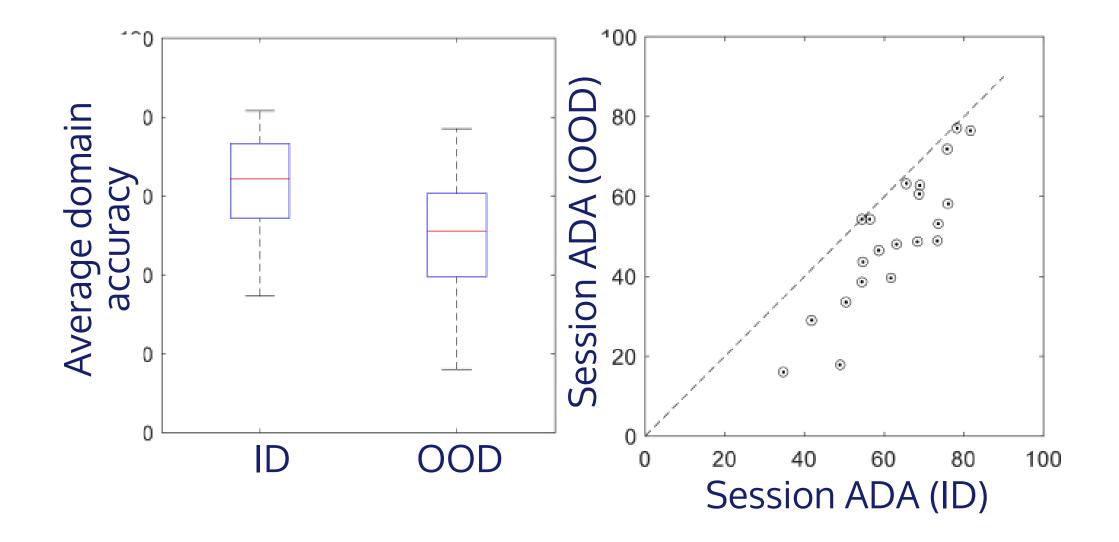
$$egin{aligned} \mathcal{L}(p,u,t^u,v) &= \mathcal{L}_{ ext{cls}}(p,u) + 1[u \geq 1]\mathcal{L}_{ ext{box}}(t^u,v) \ \mathcal{L}_{ ext{cls}}(p,u) &= -\log p_u \ \mathcal{L}_{ ext{box}}(t^u,v) &= \sum_{i \in \{x,y,w,h\}} L_1^{ ext{smooth}}(t^u_i-v_i) \end{aligned}$$



Design metrics to evaluate robustness



OOD largely underperform against ID model





How to solve the domain shift problem ?





Global Wheat Challenges

2020 kaggle















Growing science for life

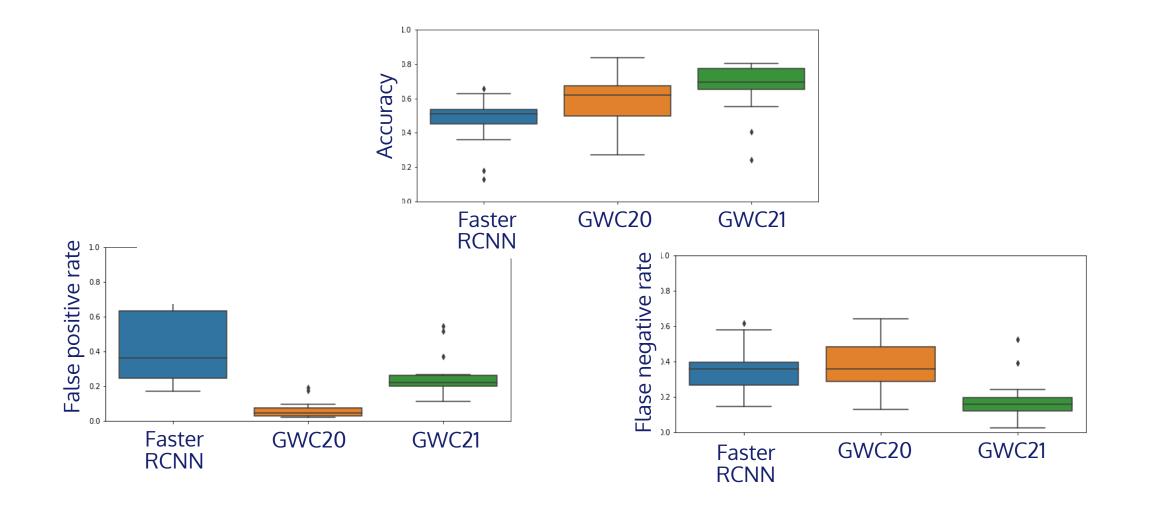
Nutrien - a Founding Partner

Common characteristics of winning solutions

	Rank	Solution name	Data preparation	Data Augmentation	Architecture	Ensemble training strategy
Baseline		Madec	No	No	Faster-RCNN	No
	1	DungNB	No	Mixup ; Custom mosaic	EfficientDet; FasterRCNN	Random subsampling
GWC_2020	2	OverFeat	Jigsaw	Mixup, Cutmix	Efficentdet	Random subsampling
	3	Javu	No	Mixup	YoloV3	No
	1	RandomTeamNa me	No	Mosaic	Yolov5	Domain subsampling
GWC_2021	2	David_jeon	Model is applied on 1600 px images	Mosaic; CutMix	Yolov5	No
	3	SMART	Network to correct image	CutMix	Yolov4	Yes



Winning solutions in 2020 and 2021 overperforms our baseline



How to solve the domain shift problem ?



Explicitly train for robustness is not working yet

WILDS: A Benchmark of in-the-Wild Distribution Shifts

		nain generalizati	Subpopulation shift	Domain generalization + subpopulation shift						
Dataset	iWildCam	Camelyon17	RxRx1	OGB-MolPCBA	GlobalWheat	CivilComments	FMoW	PovertyMap	Amazon	Py150
Input (x)	camera trap photo	o tissue slide	cell image	molecular graph	wheat image	online comment	satellite image	satellite image	product review	code
Prediction (y)	animal species	tumor	perturbed gene	bioassays v	wheat head bbo	x toxicity	land use	asset wealth	sentiment	autocomplete
Domain (d)	camera	hospital	batch	scaffold	location, time	demographic	time, region	country, rural-urb	oan user	git repository
# domains	323	5	51	120,084	47	16	16 x 5	23 x 2	2,586	8,421
# examples	203,029	455,954	125,510	437,929	6,515	448,000	523,846	19,669	539,502	150,000
Train example						What do Black and LGBT people have to do with bicycle licensing?			Overall a solid package that has a good quality of construction for the price.	<pre>import numpy as np norm=np</pre>
Test example						As a Christian, I will not be patronizing any of those businesses.			I *loved* my French press, it's so perfect and came with all this fun stuff!	<pre>import subprocess as sp p=sp.Popen() stdout=p</pre>
Adapted from	Beery et al. 2020	Bandi et al. 2018	Taylor et al. 2019	Hu et al. 2020	David et al. 2021	Borkan et al. 2019	Christie et al. 2018	Yeh et al. 2020	Ni et al. 2019	Raychev et al. 2016

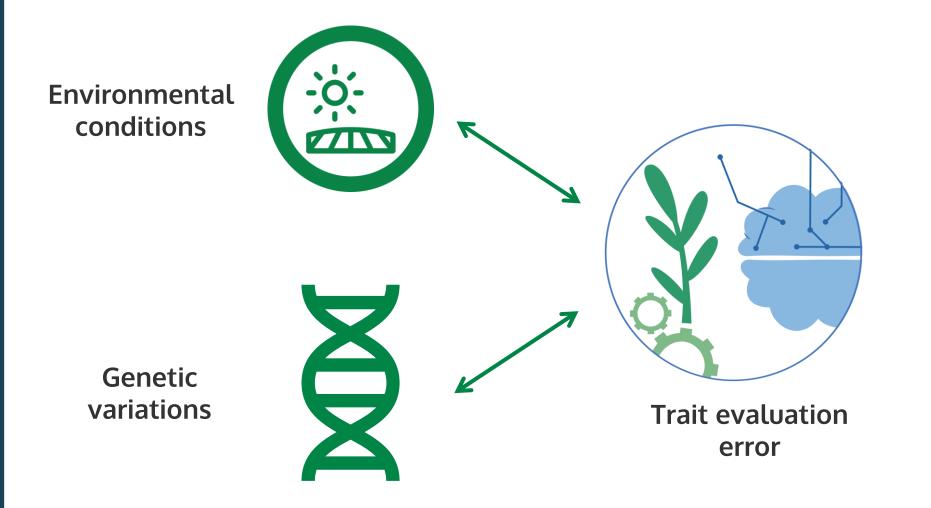
hiphen

How to solve the domain shift problem ?





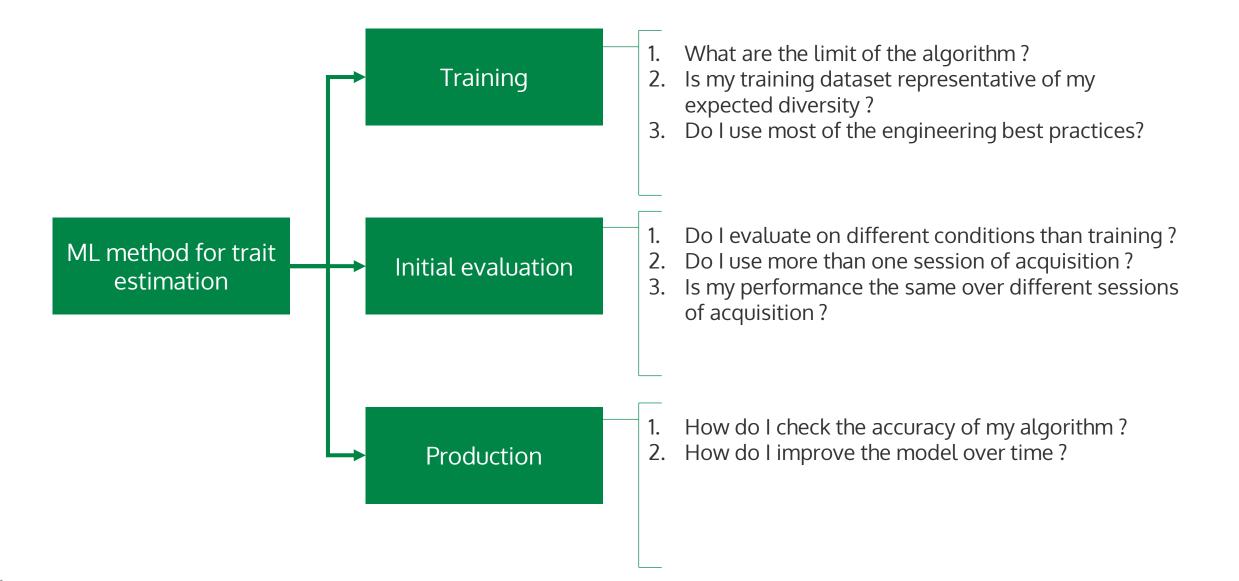
Trait evaluation over large range of conditions is difficult



The higher the expected diversity, the more difficult it is to measure a trait accurately.

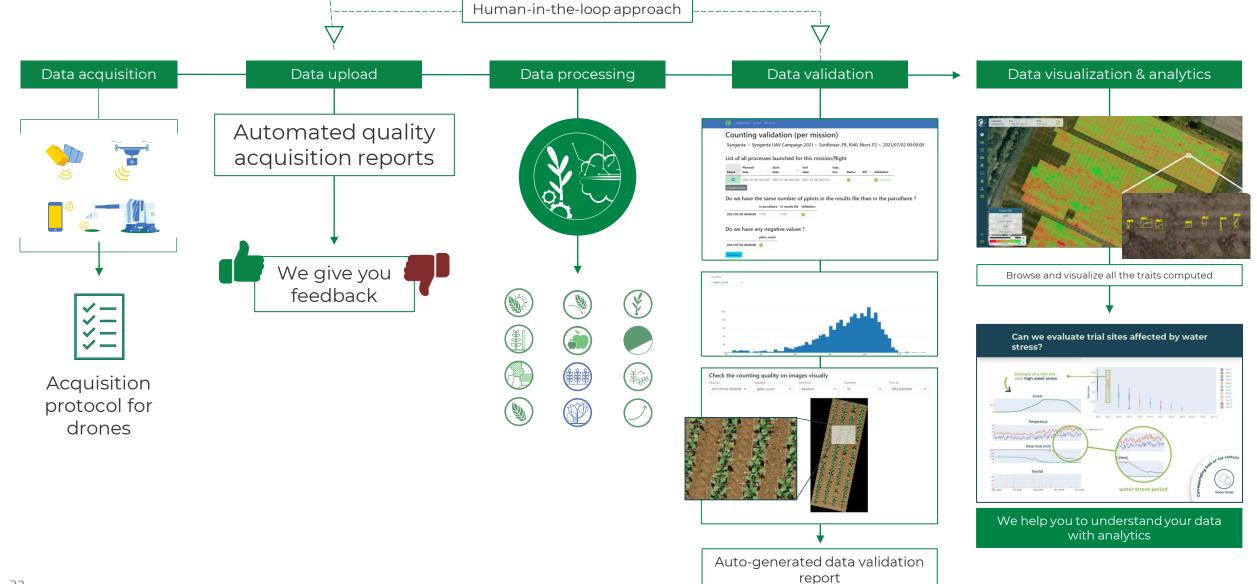
i hiphen

ML for phenotyping is more than a script





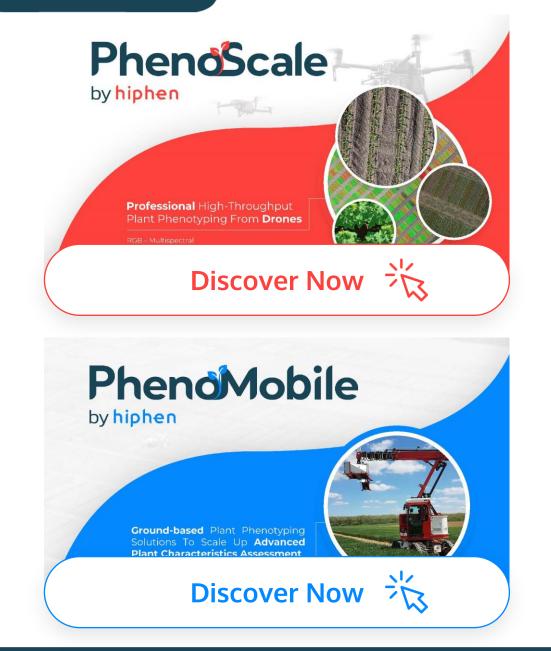
Hiphen's Human in the loop approach

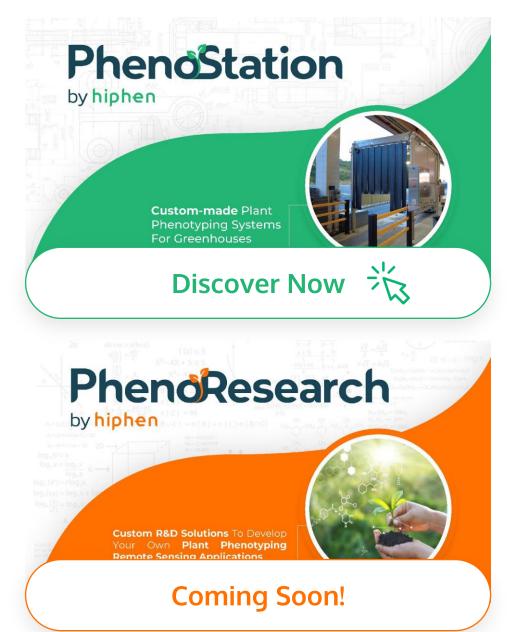




You want to discover more?

Discover more about our solutions with the catalogs

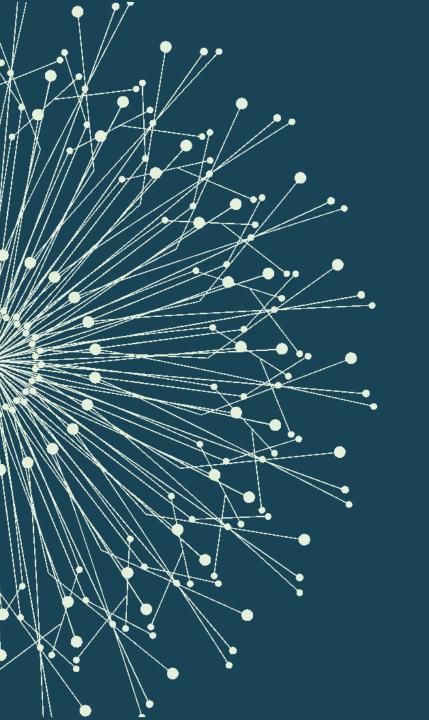






Welcome to Avignon!





Hiphen

120 rue Jean Dausset Agroparc – Bâtiment technicité 84140 Avignon France

+33 (0)7 84 14 31 63

hiphen-plant.com | contact@hiphen-plant.com

